PROCEEDINGS B

rspb.royalsocietypublishing.org

Research

Cite this article: Arbilly M, Lotem A. 2017 Constructive anthropomorphism: a functional evolutionary approach to the study of human-like cognitive mechanisms in animals. *Proc. R. Soc. B* **284**: 20171616. http://dx.doi.org/10.1098/rspb.2017.1616

Received: 19 July 2017
Accepted: 28 September 2017

Subject Category:
Behaviour

Subject Areas:
behaviour, cognition, evolution

Keywords:
cognitive evolution, emotions, language evolution, metacognition, episodic memory, empathy

Author for correspondence:
Michal Arbilly
e-mail: michalarbilly@gmail.com

Special Feature: Humans as a model for understanding biological fundamentals.

Electronic supplementary material is available online at https://dx.doi.org/10.6084/m9.figshare.c.3899368.

THE ROYAL SOCIETY
PUBLISHING

Constructive anthropomorphism: a functional evolutionary approach to the study of human-like cognitive mechanisms in animals

Michal Arbilly[1] and Arnon Lotem[2]

[1]Department of Biology, Emory University, Atlanta, GA 30022, USA
[2]Department of Zoology, Faculty of Life Sciences, Tel-Aviv University, Tel-Aviv 68878, Israel

MA, 0000-0002-3227-882X

Anthropomorphism, the attribution of human cognitive processes and emotional states to animals, is commonly viewed as non-scientific and potentially misleading. This is mainly because apparent similarity to humans can usually be explained by alternative, simpler mechanisms in animals, and because there is no explanatory power in analogies to human phenomena when these phenomena are not well understood. Yet, because it is also difficult to preclude real similarity and continuity in the evolution of humans' and animals' cognitive abilities, it may not be productive to completely ignore our understanding of human behaviour when thinking about animals. Here we propose that in applying a functional approach to the evolution of cognitive mechanisms, human cognition may be used to broaden our theoretical thinking and to generate testable hypotheses. Our goal is not to 'elevate' animals, but rather to find the minimal set of mechanistic principles that may explain 'advanced' cognitive abilities in humans, and consider under what conditions these mechanisms were likely to enhance fitness and to evolve in animals. We illustrate this approach, from relatively simple emotional states, to more advanced mechanisms, involved in planning and decision-making, episodic memory, metacognition, theory of mind, and consciousness.

1. Introduction

Anthropomorphism, the attribution of human cognitive processes or emotional states to animals, is frequently portrayed as bad science and students of animal behaviour are repeatedly warned not to fall into the trap of viewing their animal subjects as little humans [1]. There are several good reasons for this critical view and the issue has been discussed thoroughly in the past. In a behavioural take on Ockham's Razor, an argument known as Morgan's Canon has stated that, a behaviour should not be attributed to some complex psychological process if it can be explained by a simple one [2]. It has also been argued that one cannot conclude that because an animal behaves in a seemingly human way, its behaviour can be explained in the same way [3], and that an analogy is not an explanation [4].

However, while anthropomorphic tendencies are widely condemned, the question of how similar is animal cognition to human cognition has never ceased to challenge the scientific community (e.g. [5–8]). This question is especially relevant for understanding the evolution of human cognition from relatively simpler cognitive processes, and for the use of animal models in the study of human brain and behaviour. There is increasing evidence for human–animal similarities in mechanisms underlying a wide range of cognitive phenomena, from hormones or brain regions that are involved in pair-bonding and social attachment [9,10], to emotional states related to fear and aggression [11], and to spatial, episodic, and episodic-like, memory [12,13]. Nevertheless, even a high degree of mechanistic similarity cannot justify anthropomorphism. It is always possible that despite a similar mechanistic platform, added layers of sophistication in the human brain make the human condition critically different. For example, even if the same neurobiological processes were implicated in the social and mating behaviour of humans and prairie voles [9], it would not be possible to conclude that prairie voles are capable of 'loving' each other as humans do.

Much of the difficulty in comparing human and animal cognition stems from our incomplete understanding of human cognitive traits. It should be clear, for example, that in order to test whether prairie voles 'love' each other, we should first define in some mechanistic terms what is 'love' in humans; however, that is not a simple task (see review in [14]). On the other hand, our inability to define, to measure, or to observe human-equivalent processes in animals, cannot preclude their existence. In other words, while anthropomorphism is wrong, it is equally wrong to assume that whatever we cannot observe or measure does not exist.

Here, we consider a 'middle way'. We believe that the natural tendency of using our human experiences when thinking about animals (i.e. the tendency to anthropomorphize) can actually be harnessed productively to generate hypotheses regarding cognitive mechanisms and their evolution (see also [1,15,16]). We suggest a particular approach that we shall call 'constructive anthropomorphism' and take the opportunity of this special volume to present it.

Similar to the title of this volume (humans as a model for understanding biological fundamentals), our approach is to use humans as a model. The advantage of the human model is that it forces us to consider complex cognitive abilities that are normally not attributed to animals, explain them using simple biological principles, and then, to carefully examine their possible application to animals. Note that our goal is not to 'elevate' animals' cognitive abilities to those of humans. Using the human model may actually result in sharpening the differences, not necessarily in highlighting similarities [5].

Previous attempts to compare human and animal cognition have resulted in an emphasis on the use of rigorous experimental protocols, needed to avoid unsubstantiated anthropomorphic claims (e.g. [17–19]); however, this emphasis may also result in a conservative approach that cannot tolerate unproven, yet theoretically plausible ideas. Our approach is different in being mostly theoretical, and is thus focused on exploring the likelihood of various possibilities. Our goal is to find the minimal set of mechanistic principles that may explain certain 'advanced' cognitive abilities in humans, and to consider under what conditions such mechanisms were likely (or unlikely) to enhance fitness and thus evolve (or not) in other animals. This theoretical exercise will become clearer as we walk through the different examples below and show how it could be helpful in generating novel explanations and testable predictions.

We are, of course, not the first to use a functional evolutionary approach in the study of cognitive mechanisms ([20–22], and recent work by McNamara, Houston, and co-workers e.g. [23–26]). However, while previous work emphasized the adaptive value of cognitive traits regardless of whether they are based on similar or different mechanisms in animals and humans, here this potential mechanistic similarity is the main focus of the paper. We use a functional evolutionary approach to examine under what conditions the same mechanisms observed in humans were likely to evolve in animals.

In what follows, we describe our approach using several examples. We begin with emotional states, and gradually introduce more complex phenomena, such as those required for decision-making and planning, episodic memory, metacognition, mentalization and consciousness. Given the wide scope of each of these challenging subjects, our treatment here is inevitably brief and will not do them full justice. Discussion will, therefore, be limited to specific aspects and should be taken mainly as a proof of concept. As the computational and energetic costs of mechanisms discussed are unknown, we focus on potential benefits, thus setting the minimal requirement for their evolution.

Finally, relying on similar theoretical approaches, we assume that cognitive mechanisms (including highly advanced ones in humans), can be broken down into associative learning principles that can construct complex representations of past experiences in the brain [17,27–30], most probably in the form of a network [31–34]. Accordingly, advanced cognitive mechanisms are not viewed as alternatives to associative learning but rather as mechanisms that evolved from, and are based on, associative principles. This working hypothesis will allow us to propose concrete mechanistic explanations to advanced cognitive traits in humans, and to consider their possible evolution in animals.

2. Emotional states and their representations in memory

Combining different perspectives on emotional systems [23,35–40], we suggest viewing emotions very broadly (and at least for the present discussion) as 'state reporting systems'. That is, an emotional system is a system that identifies a pre-specified state (based on some pre-specified signals) and reports it to other systems of the body, which then execute a set of pre-specified actions. Consider first the very basic state of 'hunger'; it is a good starting point, as the occurrence of hunger in animals is not under dispute. In a state of hunger, some lack of nutrients is identified and signalled to other systems in the body. Those systems respond by executing a set of physiological actions (metabolic, hormonal, etc.), as well as by behavioural actions, such as foraging, enhanced aggression and suppression of conflicting activities (such as mating and breeding), and in social animals, the state of hunger may also be signalled to other individuals (e.g. [41]). Finally, the state of hunger may also affect the behaviour of other systems, for example, it can improve the learning of food-related cues [42–44].

The example of hunger makes it easy to see that having a 'state reporting system' is adaptive. Each state requires a coordinated set of actions that is different from those required by other states (see [23,24,37] for a similar approach). Yet, while we may agree that most animals have a state of hunger, can we attribute the emotional state of hunger in humans to other animals? Would it be correct to assume that other animals feel or experience hunger as humans do? And if not, what makes it different?

According to our functional evolutionary approach, we should first ask what it means to 'feel' and how a mechanism that 'feels' hunger may be adaptive for animals. This question is important because in theory, a hunger system can also work by conditional rules or stimulus-response switches without involving the 'feeling' of hunger (see for examples models of state-dependent valuation [45] or of adaptive mood states

[23]). One way of defining the 'feeling of hunger' is to propose that the collection of neuronal activities that occur simultaneously during a state of hunger is somehow experienced and represented in memory. This representation can be viewed as the representation of how it feels to be hungry. Note that this is a minimalistic mechanism. It does not require self-awareness or consciousness and it bypasses the question of whether the animal's 'self' feels the hunger, or rather something in its body subconsciously feels it. However, it makes the critical assumption that the concept of 'feeling' requires that the sensory experience is somehow represented in memory. This is a necessary first step that helps define two critical questions: (i) is it conceivable that animals can construct a representation of the state of hunger in their memory? (ii) What is the adaptive value of having such a representation?

Considering current understanding of learning mechanisms, the answer to the first question is likely to be positive. If animals can construct a representation of complex visual or acoustic experiences (e.g. [46,47]), they should also be able to represent at least some of the neuronal activities experienced during hunger. We can even speculate that such a representation emerges almost automatically as a result of neuronal co-activation that strengthens the connections between the involved neuronal units. But would it be adaptive to keep this representation? It is not at all clear, and perhaps not desirable, that any such co-activation would be represented in memory. It has been suggested that memory and learning parameters are shaped by natural selection based on their ability to construct and preserve adaptive representations while allowing non-adaptive ones to decay [34,48,49]. A potential advantage of having a representation of the state of hunger, is that it allows associating new information with the state of hunger, which may be useful for making context-appropriate decisions when being in the state of hunger again (see [50,51] for a similar approach to mood as a context for learning and recall in humans). More specifically, when the state of hunger activates the representation of hunger in memory, this representation activates past experiences that were associated with this state in the past, including a range of representations of actions or cues that are useful for finding food. This activation makes context-relevant information much more accessible and easy to recall.

In theory, this advantage of state-dependent recall may well be achieved without a representation of the state of hunger in memory. For example, the relevant information may be associated with the representation of food, which can then be activated by the state of hunger without hunger *per se* being represented. Further computational work may be necessary to clarify the conditions under which such alternative mechanisms can be as good as having a representation of hunger. However, given that an ephemeral representation of the experience of hunger is likely to emerge from neuronal co-activation during hunger, and given that for the same reason, foraging related information is also likely to be associated with this ephemeral representation, then a tendency to maintain these connected representations in memory would become adaptive: it will provide a useful hub in the network, offering immediate access to relevant information that otherwise may take longer to find.

Importantly, while this line of thought is mostly theoretical, it also generates testable predictions. If animals represent the state of hunger in their memory, as well as the state of satiation, they should also be able to learn to associate different signals with these different states. Thus, following the experimental paradigm developed for mood-dependent learning and recall in humans [50,51], if one is successful in training an animal to prefer red over blue when hungry, and blue over red when satiated, we may conclude that the animal must have had some representation of these two states; otherwise, these contrasting preferences could not have formed and the animal may always prefer red over blue. State-dependent learning of contrasting preferences may thus be taken as evidence for a representation of the two states in memory (see electronic supplementary material, Note 1 for the difference between state-dependent learning and state-dependent valuation in this context).

Finally, assuming that we accept the notion that representing the state of hunger in memory is both feasible and adaptive, and that we are even successful in providing experimental evidence for the existence of such representations in animals—does it mean that animals experience hunger as humans do? Our answer would be that the experience is probably different, but it is different as a result of the different range of associations connected with the state of hunger. In humans, the state of hunger may be associated with a much wider range of representations, and may therefore provide a richer experience than in animals. For example, in humans the state of hunger may also be associated with some representations of 'self' and 'time' that are possibly not well developed in animals (see below), and those would allow humans to represent concepts such as 'I was also hungry yesterday', which goes well beyond the basic experience of feeling hungry. However, the difference between animals and humans is in the added layers of associations to the emotional state of hunger, not in having or not having this state represented in memory.

The reasoning outlined in our discussion of hunger can now be applied to other emotional states. In the electronic supplementary material, Appendix A, we explain in some detail how it can be applied to fear, as well as to human-like emotional states such as 'jealousy' and 'being in love'.

3. Decision-making and planning: human language as a model

One of the best examples of how human cognition research can enrich our understanding of animal cognition comes from recent work on human language. While it is generally agreed that language is unique to humans [52,53], it is also believed that the main cognitive mechanisms that are needed to support language are not unique to humans [32,54,55]. Some studies seek the roots of these advanced abilities in specialized mechanisms, such as those required for vocal learning in songbirds [47,56], but recent work suggests that these abilities may not be specialized for language development and are needed for learning structure in time and space [48], and for planning sequences of motor actions [57]. Recent computational studies have demonstrated, for example, that the necessary components for a computer program that exhibits linguistic abilities [33] can evolve in the context of animal foraging in structured environments [32], and can also be used to capture elements of animal innovation and creativity [58]. Admittedly, our 'constructive anthropomorphism' approach was largely inspired by this line of research in language evolution. This is a clear case where scientists try to find the minimal set of mechanistic principles that may explain an advanced ability in humans, and consider under what conditions such mechanisms were likely to enhance fitness and to evolve in animals.

Dealing with the evolution of language, one must consider sophisticated mechanisms for data segmentation (as in segmenting a sentence into words) [59,60], for constructing hierarchical representation of complex statistical dependencies [61], and for complex decision-making and planning that are needed to construct sentences from sequences of words [33]. Such mechanisms are no doubt far more sophisticated than the learning rules normally applied in the fields of human and animal decision-making [45,62–66]. However, realistic tasks of learning and decision-making may require all these abilities, because tasks in nature are also not defined and simplified as they are in the laboratory. Many realistic tasks of animal foraging, or hunting, require animals to make sense of the world around them and solve complex problems of perception and categorization, problems that are very similar in their statistical nature to those of language learning. Importantly, recent studies in human decision-making have demonstrated that perplexing phenomena that could not be explained by traditional decision-making models, are in fact expected to emerge in learning models that consider sequential dependencies between actions and observations; that is, learning of statistical patterns that implies that what is expected to follow a particular event depends critically on what preceded this event (e.g. that the probability of finding food after 'C' may be high in the case of AC, but low in the case of BC [67–69]). Such learning of sequential dependencies is precisely what language acquisition models require (because sequences of syllables give words and sentences their meaning) and what decision-making models tended to ignore. Thus, using modelling approaches developed for language learning may help improve our understanding of animal decision-making and planning.

Finally, discussions of human decision-making and planning may frequently involve terms such as goals, desires, causality and intentionality, that are difficult to evaluate in the context of animal behaviour (e.g. see [70] for in-depth discussion). Our approach does not provide easy solutions for such problems but may offer a way to think about them (see electronic supplementary material, Appendix B).

4. The evolutionary roots of episodic memory

Episodic memory, the ability to episodically recall unique past experience and to have mental representation of events in time [71,72] was once thought to be limited to humans (e.g. [73]). However, extensive work, mostly by Clayton and co-workers (e.g. [74–78]), has demonstrated that some animals can certainly have episodic-like memory. Studying food-caching behaviour in scrub jays, Clayton *et al*. were able to show that individuals can remember 'what', 'when' and 'where' they cached [75]. The fact that caching behaviour could reveal the existence of episodic-like memory in some animals suggests that episodic-like memory may be common, but difficult to detect in animals that do not cache. Considering this possibility, Clayton *et al*. [79] took a functional evolutionary approach (of the kind we endorse in this paper) and proposed several cases in which having an episodic-like memory may be adaptive and therefore likely to have evolved. They suggested, for example, the need to keep track of who did what and to whom in primate societies, the need of brood parasites to simultaneously monitor the breeding chronology of various potential hosts, and the need of polygynous males to keep track of the reproductive status and behaviour of their different females. We strongly agree with this approach (see also [80]), but would like to take it a step further. As explained below, we suggest that the basic computational principles of episodic-like memory are actually necessary under a much wider range of circumstances, which means that at least some forms of episodic memory have evolved quite early in behavioural evolution.

To see why episodic memory may actually be essential to memory systems, and therefore needed by most animals, it is useful to return to the well-recognized distinction between semantic and episodic memory [72]. This distinction implies that semantic and episodic memories represent alternative modes of data storage that serve different needs. The need for semantic memory is quite clear. Most associative networks that represent the statistical relationship (or more precisely, the transitional probabilities) between sequential events and items in nature may be viewed as semantic networks (including simple learning models that represent the probability of receiving a reward following a signal or an action). Such networks are necessary for predicting future events and planning future actions. However, to construct a useful semantic network, repeated observations are normally averaged over time (or aggregated in some other way) and then represented in memory as a generalized graph (see electronic supplementary material, Note 2). While this process improves statistical accuracy, it inevitably erases the original observations, which means that real historical episodes and their chronological order are not preserved. For most learning tasks this is not a problem and in fact most learning models are designed this way (e.g. [62,64]). However, there is a range of decisions and actions in an animal's life, for which precise historical data are necessary. Consider, for example, a bird that feeds on spiders that find shelter under leaves or little rocks. Based on repeated observations, the bird can construct a semantic representation, indicating that the probability of finding a spider underneath a certain type of leaf is approximately 0.4, and underneath a certain type of rock is approximately 0.2, etc. This can clearly guide foraging behaviour. However, when a particular spider found by the bird starts running away from one shelter to another, it is critical for the bird to remember where the spider was last seen. This information cannot be provided by the semantic network. Thus, a memory trace of recent events should be stored separately, without being immediately integrated into the semantic network.

The need for such recent episodic-like memory can be quite general, as animals must keep track of recent events in order to respond to them. Remembering one's recent actions is also valuable, as these actions, or their outcomes, may be informative for subsequent decision-making (as in systematic search, for example). From a computational point of view, these trace memories are already episodic. They represent a particular sequence of events in a specific point in time; this representation is unique and must be separated from the general representation of semantic knowledge. Humans may be much better than other animals in representing longer and more complex sequences of historical events and those may also be associated with a richer range of associations and concepts, some of which may be unique to humans. Certainly, only humans can recall a particular conversation they have had, a story they have read, or use language to describe them. Yet, the essence of episodic memory may be best characterized by the very basic computational problem of separating specific historical data from general semantic knowledge, a problem that must have been solved quite early in behavioural evolution.

5. Metacognition

Metacognition is the capacity to monitor and control one's cognitive processes [81], or, more intuitively, as one's ability to know what she knows or to think what she thinks [82,83]. Similar to episodic memory, metacognition has been viewed as an advanced cognitive ability that is limited to humans, but recent evidence suggests that it can be found in animals, and may be rooted in relatively basic computational processes. It is obviously difficult to study metacognition in animals, but using perception, memory and food-concealment paradigms, recent studies suggest that animals do develop some sense of certainty or uncertainty regarding their level of knowledge (e.g. [84,85]). The interpretations of these studies are nevertheless debated (e.g. [81,86,87]).

One view is that animal metacognition, like human metacognition, requires high-level cognitive processes that are deliberate and decisional, and possibly involve self-reflection and consciousness [81]. The mechanistic nature of these 'high-level' processes is yet to be specified and is currently unclear (see [86]). Alternative approaches are that animal metacognition can be explained by associative mechanisms [88], or similarly may be viewed as a discrimination mechanism that is tuned into internal signals of memory strength that provide discriminative cues [86,89]. According to this view, animals' sense of certainty or uncertainty of their knowledge does not require self-reflection or consciousness, but can develop by gradually associating internal signals, normally correlated with memory strength (or memory accuracy), with successful outcomes. Thus, the sense of certainty emerges by the extent to which such internal signals predict successful outcomes. Importantly, studies in humans suggest that subjects cannot assess their memory strength or accuracy directly, but instead use internal signals such as the experience of 'ease of processing', and 'response time' [87], or the fluency of action [90], that are normally correlated with memory strength. In this light, metacognition in both humans and animals is much more associative than self-reflective, or in other words, one doesn't really know what he knows, but rather learns to map the relationship between memory-related cues and the outcomes of their actions (see [87] for a review).

This associative view of metacognition can help us see not only how animal metacognition may evolve, but also that the conditions for its evolution were likely quite common. For most animals, a tendency to associate internal memory-related cues with successful outcomes should be both feasible and adaptive. It should be feasible because it only requires some tuning of domain-general associative learning mechanisms, directing learners to be attentive to memory-related cues. It is likely to be adaptive because it improves decision-making; it helps predict whether an action is likely to succeed, and should therefore be executed, or likely to fail and should therefore be avoided. The conditions for the evolution of such metacognitive mechanisms require that decision-making will indeed be improved (which depends on how the learned correlation between cues and outcomes reduces uncertainty), but do not seem to require a great deal of 'high-level' cognitive sophistication.

Moreover, following our approach of identifying the minimal set of mechanistic principles that may explain an advanced cognitive ability, and considering under what conditions it was likely to first evolve in animals, we suggest that the mechanism described is not necessarily specific to metacognition. Animals needed such a mechanism for assessing what they can do, not only for assessing what they know. In other words, we suggest that animals use the same mechanism for assessing knowledge and for assessing abilities. Knowledge is simply more cryptic and was traditionally discussed in the context of metacognition, but the assessment of knowledge may not be different from the assessment of physical abilities.

To see the similarity, consider for example a dog who is about to decide whether it should run and fetch a stick thrown by its owner over a fence separating two yards. To make this decision, the dog needs to assess whether it really knows that the stick is behind the fence, and it also needs to assess its ability to jump over the fence. While only the first assessment is commonly viewed as metacognitive, both require a very similar mechanism. The dog may be relatively certain that the stick is behind the fence if it quickly and easily remembered that he just saw the stick flying in this direction, and in the past, such ease and speed of recalling where an object was last seen were associated with finding this object. As for the fence, the dog may be relatively certain that it can jump over the fence if it can quickly and easily remember that barriers of such height were associated with successful jumps in the past. If the dog had never jumped that high before, or jumped to such heights very rarely, it will find it difficult to recall relevant past memories, which will make it less certain about its ability. Thus, the assessment of abilities, like the assessment of knowledge, requires access to context-relevant information that was associated in the past with successful performances. Because the information about physical abilities is represented in memory, the cognitive process of assessing the strength or accuracy of such memories involves the same problem faced by animals that need to assess their knowledge. The only difference is that memories about abilities are related to self-actions while memories about the outside world are normally viewed as knowledge. In fact, when ability is improved through experience, like in the case of one's ability to climb trees, for example, the assessment of whether he 'can climb' or 'knows how to climb' is completely mixed. The roots of metacognition may thus lie in the mechanisms of motor control that evolved as early as the time dragonflies needed to intercept their prey [91].

An open question that still remains is to what extent animals 'feel' their level of uncertainty or rather respond to memory-related cues automatically. We have already considered a similar question when discussing whether animals 'feel' their emotional states (see §2; electronic supplementary material, Appendix A). We believe that this question can be addressed similarly. Assuming that the level of certainty is a state that can be characterized by a combination of neuronal activities, it should be possible to represent it in memory, and it will be adaptive to do so if it can help facilitate state-dependent recall (see §2). Interestingly, an advantage of state-dependent recall is already implied in the associative account of metacognition that was just discussed (e.g. [87]). The memory-related cues that are suggested to be used in metacognition, such as the 'ease of processing' or 'response time', are preserved in memory, represent various states of uncertainty, and these states are associated with the outcomes of similar actions taken under the same states in the past, which means that being in a given state allows one to recall its expected outcomes. Thus, according to our minimalistic definition in §2, if animals can remember their state of uncertainty they can also 'feel' it. Yet, as discussed for emotional states, the representation of such states in humans may be associated with a much richer set of memories and concepts than in animals.

6. Empathy, theory of mind, consciousness and self-awareness

Clearly, many questions cannot be adequately addressed within the limited scope of this paper. These include the most challenging questions of whether animals feel empathy, are capable of developing a theory of mind, and the extent to which animals are conscious or have some sense of self-awareness. While any attempt to address these questions so briefly cannot be convincing, we suggest that the same approach taken thus far in this paper may be instructive in dealing with these complex issues. Namely, we should seek the simplest mechanisms that can explain some basic forms of such cognitive abilities and consider under what conditions they were likely to evolve in animals. We briefly sketch a few examples.

An essential component of empathy and theory of mind is the ability to attribute mental states to others [92,93]. This ability appears non-trivial because one cannot experience the mental states of others (their pain, their fears, their intentions, etc.), which makes it difficult to see how this ability can develop through associative learning. However, from a computational point of view, attributing mental states to others may be achieved through the same generalization mechanisms that allow attributing any previously unseen traits to other individuals. This process is in fact very basic to learning. For example, a bird may expect that a novel type of grasshopper is edible if it is sufficiently similar to previously eaten grasshoppers, and it may also expect it to be capable of flying if other grasshoppers flew away from this bird in the past. The same generalization process allows a child to attribute the ability of riding a bicycle to another child even if she never saw this particular child riding a bicycle before. Statistically speaking, this generalization would be correct. The transition from this form of generalization to the attribution of mental states, such as pain, may be possible as long as the observing child has acquired sufficient experience to classify herself as a child (i.e. as belonging to the same general category of the observed individual). At this point, she can generalize her own ability to feel pain to another child, as long as she views this child as sufficiently similar to herself and the context to be sufficiently similar to a context where she experienced pain in the past (e.g. when falling off a bicycle and starting to cry). From a computational point of view, the generalization from observable to non-observable and from self to others may not require much more than any generalization, but like any generalization, it requires the accumulation of sufficient information. A detailed account of how theory of mind may develop through generalization in associative networks can be found in §3.1 of [94], and is in line with recent theories that view mind reading as simulations [95,96]. Because the ability to generalize and to simulate depends on the accumulation of sufficient relevant information, it is quite possible that animals' ability to generalize in the mental domain is constrained by their limited social attention and by the lack of language and cultural transmission (see [97] for related discussion). However, being based on generalization processes, some level of attributing mental states to others may be quite feasible for animals, and may be adaptive for predicting the behaviour of other group members.

Consciousness and self-awareness may similarly evolve from simpler mechanisms, for example, those that allow animals to build a model of their own body—an ability that has been explicitly developed in robots [98] and is akin to some level of self-awareness. Further development of the concept of 'self' may result from the need to separate information about 'self' versus 'non-self', and to separate information associated with different individuals, where 'self' is just one of them. While the cognitive mechanisms supporting consciousness and self-awareness are not yet clear, it has been suggested that their computational ingredients are inherent to any representational system (e.g. [99] for a review). Viewing in this light, it is not impossible that animals are capable of some level of consciousness and self-awareness. The question of why animals need these abilities, and thus how likely it is that such abilities have evolved in animals (given that it is perhaps feasible), remains open and certainly deserves more in-depth exploration than we can afford here.

7. Summary and conclusion

In this paper, we have suggested that human cognition may be used to broaden our theoretical thinking about animal cognition and about cognitive evolution. We proposed that this can be done by identifying a minimal set of mechanistic principles that may explain advanced cognitive abilities in humans, and by considering under what conditions such mechanisms were likely to enhance fitness and thus to evolve in animals. To demonstrate this approach, we applied it to a set of well-known human cognitive abilities. We started with emotional states and suggested that humans' capacity of feeling emotions may be rooted in the adaptive value of representing such states in memory, which can facilitate state-dependent learning and recall of relevant information. We continued by showing how recent work on language learning in humans may shed new light on the evolution of complex learning mechanisms in animals, introducing concepts of statistical learning and associative networks to problems of decision-making and planning faced by animals. Similarly, examining episodic memory and metacognition, once considered unique to humans, helped identify their core mechanisms and possible evolutionary background in animals: the need to separate episodic and semantic memories quite early in behavioural evolution and the availability of associative models for metacognition suggest that some forms of episodic memory and metacognition may be common in animals. Moreover, as we stressed above, basic metacognitive mechanisms for assessing knowledge were already needed for assessing physical abilities, making them almost ubiquitous among animals. Finally, we sketched possible directions for applying our approach to address open questions regarding empathy, mind reading, consciousness and self-awareness. While further research is clearly needed, we hope that theoretical approaches of the kind presented here may be useful in specifying potential mechanisms and in improving our understanding of cognitive evolution.

Data accessibility. This article has no additional data.

Authors' contributions. A.L. and M.A. co-wrote the paper.

Competing interests. We have no competing interests.

Funding. A.L. was supported by Israel Science Foundation grant 871/15.

Acknowledgements. We thank two anonymous reviewers for thoughtful comments that improved the paper, Sarah Brosnan for encouragement and patience, and Joe Halpern, Oren Kolodny and Shimon Edelman for productive discussions.

References

1. Shettleworth SJ. 2010 *Cognition, evolution and behavior*, 2nd edn. New York, NY: Oxford University Press.
2. Morgan CL. 1894 *An introduction to comparative psychology*. London, UK: Walter Scott.
3. Heyes C. 2008 Beast machines? Questions of animal consciousness. In *Frontiers of consciousness: Chichele Lectures* (eds M Davies, L Weiskrantz), pp. 259–274.
4. Wynne CDL. 2004 Consciousness should be ascribed to animals only with extreme caution. *Nature* **428**, 2004. (doi:10.1038/428606a)
5. Shettleworth SJ. 2012 Modularity, comparative cognition and human uniqueness. *Phil. Trans. R. Soc. B* **367**, 2794–2802. (doi:10.1098/rstb.2012.0211)
6. Meketa I. 2014 A critique of the principle of cognitive simplicity in comparative cognition. *Biol. Phil.* **29**, 731–745. (doi:10.1007/s10539-014-9429-z)
7. Burkart JM, Schubiger MN, van Schaik CP. 2016 The evolution of general intelligence. *Behav. Brain Sci.* **40**, 1–65. (doi:10.1017/S0140525X16000959)
8. Dewey C. 2017 Anthropomorphism and anthropectomy as friendly competitors. *Phil. Psychol.* **5089**, 1–22. (doi:10.1080/09515089.2017.1334116)
9. Mcgraw LA, Young LJ. 2010 The prairie vole: an emerging model organism for understanding the social brain. *Trends Neurosci.* **33**, 103. (doi:10.1016/j.tins.2009.11.006)
10. Maroun M, Wagner S. 2016 Oxytocin and memory of emotional stimuli: some dance to remember, some dance to forget. *Biol. Psychiatry* **79**, 203–212. (doi:10.1016/j.biopsych.2015.07.016)
11. McCall C, Singer T. 2012 The animal and human neuroendocrinology of social cognition, motivation and behavior. *Nat. Neurosci.* **15**, 681–688. (doi:10.1038/nn.3084)
12. Burgess N, Maguire EA, O'Keefe J. 2002 The human hippocampus and spatial and episodic memory. *Neuron* **35**, 625–641. (doi:10.1016/S0896-6273(02)00830-9)
13. Dere E, Kart-Teke E, Huston JP, De Souza Silva MA. 2006 The case for episodic memory in animals. *Neurosci. Biobehav. Rev.* **30**, 1206–1224. (doi:10.1016/j.neubiorev.2006.09.005)
14. Fredrickson BL. 2016 Love: positivity resonance as a fresh, evidence-based perspective on an age-old topic. In *Handbook of emotions* (eds L Feldman, BM Lewis, JM Haviland-Jones), pp. 847–858. New York, NY: The Guilford Press.
15. De Waal FBM. 1999 Anthropomorphism and anthropodenial: consistency in our thinking about humans and other animals. *Phil. Top.* **27**, 255–280. (doi:10.2307/43154308)
16. Sober E. 2012 Anthropomorphism, parsimony, and common. *Mind Lang.* **27**, 229–238. (doi:10.1111/j.1468-0017.2012.01442.x)
17. Heyes C. 2012 Simple minds: a qualified defence of associative learning. *Phil. Trans. R. Soc. B* **367**, 2695–2703. (doi:10.1098/rstb.2012.0217)
18. Vasconcelos M, Hollis K, Nowbahari E, Kacelnik A. 2012 Pro-sociality without empathy. *Biol. Lett.* **8**, 910–912. (doi:10.1098/rsbl.2012.0554)
19. Scarf D, Smith C, Stuart M. 2014 A spoon full of studies helps the comparison go down: a comparative analysis of Tulving's spoon test. *Front. Psychol.* **5**, 1–6. (doi:10.3389/fpsyg.2014.00893)
20. Kamil AC. 1988 *A synthetic approach to the study of animal intelligence*. In *Nebraska Symp. Motiv. 1987, vol. 35 Comp. Perspect. Mod. Psychol.*, pp. 257–308.
21. Shettleworth SJ. 1993 Where is the comparison in comparative cognition? Alternative research programs. *Psychol. Sci.* **4**, 179–184. (doi:10.1111/j.1467-9280.1993.tb00484.x)
22. Real L. 1991 Animal choice behavior and the evolution of cognitive architecture. *Science* **253**, 980–986. (doi:10.1126/science.1887231)
23. Trimmer PC, Paul E, Mendl M, McNamara J, Houston AI. 2013 On the evolution and optimality of mood states. *Behav. Sci.* **3**, 501–521. (doi:10.3390/bs3030501)
24. McNamara JM, Houston AI. 2009 Integrating function and mechanism. *Trends Ecol. Evol.* **24**, 670–675. (doi:10.1016/j.tree.2009.05.011)
25. McNamara JM, Fawcett TW, Houston AI. 2013 An adaptive response to uncertainty generates positive and negative contrast effects. *Science* **340**, 1084–1086. (doi:10.1126/science.1230599)
26. Fawcett TW, McNamara JM, Houston AI. 2012 When is it adaptive to be patient? A general framework for evaluating delayed rewards. *Behav. Process.* **89**, 128–136. (doi:10.1016/j.beproc.2011.08.015)
27. Lotem A, Halpern JY, Edelman S, Kolodny O. 2017 The evolution of cognitive mechanisms in response to cultural innovations. *Proc. Natl Acad. Sci. USA* **114**, 7775–7781. (doi:10.1073/pnas.1620742114)
28. Edelman S. 2015 The minority report: some common assumptions to reconsider in the modelling of the brain and behaviour. *J. Exp. Theor. Artif. Intell.* **3079**, 751–776. (doi:10.1080/0952813X.2015.1042534)
29. Cook R, Bird G, Catmur C, Press C, Heyes C. 2014 Mirror neurons: from origin to function. *Behav. Brain Sci.* **37**, 177–192. (doi:10.1017/s0140525x13000903)
30. Mnih V *et al.* 2015 Human-level control through deep reinforcement learning. *Nature* **518**, 529–533. (doi:10.1038/nature14236)
31. Faber T, Joerges J, Menzel R. 1999 Associative learning modifies neural representations of odors in the insect brain. *Nat. Neurosci.* **2**, 74–78. (doi:10.1038/4576)
32. Kolodny O, Edelman S, Lotem A. 2015 Evolution of protolinguistic abilities as a by-product of learning to forage in structured environments. *Proc. R. Soc. B* **282**, 20150353. (doi:10.1098/rspb.2015.0353)
33. Kolodny O, Lotem A, Edelman S. 2015 Learning a generative probabilistic grammar of experience: a process-level model of language acquisition. *Cogn. Sci.* **39**, 227–267. (doi:10.1111/cogs.12140)
34. Lotem A, Halpern JY. 2012 Coevolution of learning and data-acquisition mechanisms: a model for cognitive evolution. *Phil. Trans. R. Soc. B* **367**, 2686–2694. (doi:10.1098/rstb.2012.0213)
35. Damasio A. 2001 Fundamental feelings. *Nature* **413**, 781. (doi:10.1038/35101669)
36. Adolphs R. 2015 How can we study emotion? Towards a functional concept of emotion states. *Jap. J. Anim. Psychol.* **22**, 11–22. (doi:10.2502/janip.65.1.3)
37. Bach DR, Dayan P. 2017 Opinion: algorithms for survival: a comparative perspective on emotions. *Nat. Rev. Neurosci.* **18**, 311–319. (doi:10.1038/nrn.2017.35)
38. Keltner D, Gross JJ. 1999 Functional accounts of emotions. *Cogn. Emot.* **13**, 467–480. (doi:10.1080/026999399379140)
39. Etkin A, Buechel C, Gross JJ. 2015 The neural bases of emotion regulation. *Nat. Rev. Neurosci.* **16**, 693–700. (doi:10.1038/nrn4044)
40. Schachter S, Singer JE. 1962 Cognitive, social, and physiological determinants of emotional state. *Psychol. Rev.* **69**, 379–399. (doi:10.1037/h0046234)
41. Kilner R. 1997 Mouth colour is a reliable signal of need in begging canary nestlings. *Proc. R. Soc. B* **264**, 963–968. (doi:10.1098/rspb.1997.0133)
42. Marsh B, Schuck-Paim C, Kacelnik A. 2004 Energetic state during learning affects foraging choices in starlings. *Behav. Ecol.* **15**, 396–399. (doi:10.1093/beheco/arh034)
43. Pompilio L, Kacelnik A, Behmer ST. 2006 State-dependent learned valuation drives choice in an invertebrate. *Science* **311**, 1613–1615. (doi:10.1126/science.1122469)
44. Aw JM, Holbrook RI, Burt de Perera T, Kacelnik A. 2009 State-dependent valuation learning in fish: banded tetras prefer stimuli associated with greater past deprivation. *Behav. Process.* **81**, 333–336. (doi:10.1016/j.beproc.2008.09.002)
45. McNamara JM, Trimmer PC, Houston AI. 2012 The ecological rationality of state-dependent valuation. *Psychol. Rev.* **119**, 114–119. (doi:10.1037/a0025958)
46. Tsoar A, Nathan R, Bartan Y, Vyssotski A, Dell G, Ulanovsky N. 2011 Large-scale navigational map in a mammal. *Proc. Natl Acad. Sci. USA* **108**, E718–E724. (doi:10.1073/pnas.1107365108)
47. Lipkind D *et al.* 2013 Stepwise acquisition of vocal combinatorial capacity in songbirds and human infants. *Nature* **498**, 104–108. (doi:10.1038/nature12173.Stepwise)
48. Goldstein MH, Waterfall HR, Lotem A, Halpern JY, Schwade JA, Onnis L, Edelman S. 2010 General cognitive principles for learning structure in time

and space. *Trends Cogn. Sci.* **14**, 249–258. (doi:10.1016/j.tics.2010.02.004)

49. Richards BA, Frankland PW. 2017 The persistence and transience of memory. *Neuron* **94**, 1071–1084. (doi:10.1016/j.neuron.2017.04.037)
50. Bower GH, Monteiro KP, Gilligan SG. 1978 Emotional mood as a context for learning and recall. *J. Verbal Learn. Verbal Behav.* **17**, 573–585. (doi:10.1016/S0022-5371(78)90348-1)
51. Bower GH. 1981 Mood and memory. *Am. Psychol.* **36**, 129–148. (doi:10.1016/0005-7967(87)90052-0)
52. Hauser MD, Chomsky N, Fitch WT. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
53. Christiansen MH, Kirby S. 2003 Language evolution: the hardest question in science? In *Language evolution* (eds MH Christiansen, S Kirby), pp. 1–15. Oxford, UK: Oxford University Press.
54. Christiansen MH, Chater N. 2008 Language as shaped by the brain. *Behav. Brain Sci.* **31**, 458–489. (doi:10.1017/S0140525X08004998)
55. Zuberbühler K. 2003 Referential signalling in non-human primates: cognitive presursors and limitations for the evolution of language. *Adv. Study Behav.* **33**, 265–307. (doi:10.1016/S0065-3454(03)33006-2)
56. Scharff C, Petri J. 2011 Evo-devo, deep homology and FoxP2: implications for the evolution of speech and language. *Phil. Trans. R. Soc. B* **366**, 2124–2140. (doi:10.1098/rstb.2011.0001)
57. Byrne RW. 1999 Imitation without intentionality. Using string parsing to copy the organization of behaviour. *Anim. Cogn.* **2**, 63–72. (doi:10.1007/s100710050025)
58. Kolodny O, Edelman S, Lotem A. 2015 Evolved to adapt: a computational approach to animal innovation and creativity. *Curr. Zool.* **61**, 350–367. (doi:10.1093/czoolo/61.2.350)
59. Saffran JR, Aslin RN, Newport EL. 1996 Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928. (doi:10.1126/science.274.5294.1926)
60. Brent MR. 1999 Speech segmentationand word discovery: a computational perspective. *Trends Cogn. Sci.* **3**, 294–301. (doi:10.1016/S1364-6613(99) 01350-9)
61. Solan Z, Horn D, Ruppin E, Edelman S. 2005 Unsupervised learning of natural languages. *Proc. Natl Acad. Sci. USA* **102**, 11 629–11 634. (doi:10.1073/pnas.0409746102)
62. Arbilly M, Motro U, Feldman MW, Lotem A. 2010 Co-evolution of learning complexity and social foraging strategies. *J. Theor. Biol.* **267**, 573–581. (doi:10.1016/j.jtbi.2010.09.026)
63. Katsnelson E, Motro U, Feldman MW, Lotem A. 2012 Evolution of learned strategy choice in a frequency-dependent game. *Proc. R. Soc. B* **279**, 1176–1184. (doi:10.1098/rspb.2011.1734)
64. Trimmer PC, McNamara JM, Houston AI, Marshall JAR. 2012 Does natural selection favour the Rescorla–Wagner rule? *J. Theor. Biol.* **302**, 39–52. (doi:10.1016/j.jtbi.2012.02.014)
65. Hamblin S, Giraldeau L-A. 2009 Finding the evolutionarily stable learning rule for frequency-dependent foraging. *Anim. Behav.* **78**, 1343–1350. (doi:10.1016/j.anbehav.2009.09.001)
66. Lange A, Dukas R. 2009 Bayesian approximations and extensions: optimal decisions for small brains and possibly big ones too. *J. Theor. Biol.* **259**, 503–516. (doi:10.1016/j.jtbi.2009.03.020)
67. Plonsky O, Teodorescu K, Erev I. 2015 Reliance on small samples, the wavy recency effect, and similarity-based learning. *Psychol. Rev.* **122**, 621–647. (doi:10.1037/a0039413)
68. Hochman G, Erev I. 2013 The partial-reinforcement extinction effect and the contingent-sampling hypothesis. *Psychon. Bull. Rev.* **20**, 1336–1342. (doi:10.3758/s13423-013-0432-1)
69. Humphreys LG. 1939 The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *J. Exp. Psychol.* **25**, 141–158. (doi:10.1037/h0058138)
70. Heyes C, Dickinson A. 1990 The intentionality of animal action. *Mind Lang.* **5**, 87–104. (doi:10.1111/j.1468-0017.1990.tb00154.x)
71. Tulving E. 2002 Episodic memory: from mind to brain. *Annu. Rev. Psychol.* **53**, 1–25. (doi:10.1146/annurev.psych.53.100901.135114)
72. Tulving E. 1972 Episodic and semantic memory. *Organ. Mem.* **1**, 381–403. (doi:10.1017/S0140525X00047257)
73. Suddendorf T, Corballis MC. 1997 Mental time travel and the evolution of the human mind. *Genet. Soc. Gen. Psychol. Monogr*. **123**, 133–167.
74. Clayton NS, Dickinson A. 1999 Scrub jays (*Aphelocoma coerulescens*) remember the relative time of caching as well as the location and content of their caches. *J. Comp. Psychol.* **113**, 403–416. (doi:10.1037/0735-7036.113.4.403)
75. Clayton NS, Dickinson A. 1998 Episodic-like memory during cache recovery by scrub jays. *Nature* **395**, 272–274. (doi:10.1038/26216)
76. Dally JM, Emery NJ, Clayton NS. 2006 Food-caching western scrub-jays keep track of who was watching when. *Science* **312**, 1662–1665. (doi:10.1126/science.1126539)
77. Clayton NS, Yu KS, Dickinson A. 2001 Scrub jays (*Aphelocoma coerulescens*) form integrated memories of the multiple features of caching episodes. *J. Exp. Psychol. Anim. Behav. Process.* **27**, 17–29. (doi:10.1037/0097-7403.27.1.17)
78. Clayton NS, Yu KS, Dickinson A. 2003 Interacting cache memories: evidence for flexible memory use by western scrub-jays (*Aphelocoma californica*). *J. Exp. Psychol.* **29**, 14–22. (doi:10.1037/0097-7403.29.1.14)
79. Clayton NS, Griffiths PD, Emery NJ, Dickinson A. 2001 Elements of episodic-like memory in animals. *Phil. Trans. R. Soc. B* **356**, 1483–1491. (doi:10.1098/rstb.2001.0947)
80. Allen TA, Fortin NJ. 2013 The evolution of episodic memory. *Proc. Natl Acad. Sci. USA* **110**, 10 379–10 386. (doi:10.1073/pnas.1301199110)
81. Smith JD, Couchman JJ, Beran MJ. 2014 Animal metacognition: a tale of two comparative psychologies. *J. Comp. Psychol.* **128**, 115–131. (doi:10.1037/a0033105)
82. Heyes C. 2016 Who knows? Metacognitive social learning strategies. *Trends Cogn. Sci.* **20**, 204–213. (doi:10.1016/j.tics.2015.12.007)
83. Shea N, Boldt A, Bang D, Yeung N, Heyes C, Frith CD. 2014 Supra-personal cognitive control and metacognition. *Trends Cogn. Sci.* **18**, 186–193. (doi:10.1016/j.tics.2014.01.006)
84. Hampton RR. 2001 Rhesus monkeys know when they remember. *Proc. Natl Acad. Sci. USA* **98**, 5359–5362. (doi:10.1073/pnas.071600998)
85. Perry CJ, Barron AB. 2013 Honey bees selectively avoid difficult choices. *Proc. Natl Acad. Sci. USA* **110**, 19 155–19 159. (doi:10.1073/pnas.1314571110)
86. Basile BM, Hampton RR. 2014 Metacognition as discrimination: commentary on Smith *et al*. (2014). *J. Comp. Psychol.* **128**, 135–137. (doi:10.1037/a0034412)
87. Kornell N. 2014 Where is the 'meta' in animal metacognition? *J. Comp. Psychol.* **128**, 143–149. (doi:10.1037/a0033444)
88. Le Pelley ME. 2012 Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *J. Exp. Psychol. Learn. Mem. Cogn.* **38**, 686–708. (doi:10.1037/a0026478)
89. Hampton RR. 2005 Can rhesus monkeys discriminate between remembering and forgetting? In *Miss. link cogn. orig. self-reflective conscious* (eds HS Terrace, J Metcalf), pp. 272–295. Oxford, UK: Oxford University Press.
90. Chambon V, Haggard P. 2012 Sense of control depends on fluency of action selection, not motor performance. *Cognition* **125**, 441–451. (doi:10.1016/j.cognition.2012.07.011)
91. Mischiati M, Lin H-T, Herold P, Imler E, Olberg R, Leonardo A. 2014 Internal models direct dragonfly interception steering. *Nature* **517**, 333–338. (doi:10.1038/nature14045)
92. Premack D, Woodruff G. 1978 Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* **4**, 515–526. (doi:10.1016/j.celrep.2011.1011.1001.7)
93. Davis MH. 1983 Measuring individual differences in empathy: evidence for a multidimensional approach. *J. Pers. Soc. Psychol.* **44**, 113–126. (doi:10.1037/0022-3514.44.1.113)
94. Lotem A, Halpern JY. 2008 A data-acquisition model for learning and cognitive development and its implications for autism. *Cornell Comput. Inf. Sci. Tech. Rep.* http://hdl.handle.net/1813/10178.
95. Ramnani N, Miall RC. 2004 A system in the human brain for predicting the actions of others. *Nat. Neurosci.* **7**, 85–90. (doi:10.1038/nn1168)
96. Goldman AI. 2006 Simulating minds: the philosophy, psychology, and neuroscience of mindreading. Oxford, UK: Oxford University Press.
97. Heyes C, Frith CD. 2014 The cultural evolution of mind reading. *Science* **344**, 1243091. (doi:10.1126/science.1243091)
98. Bongard J, Zykov V, Lipson H. 2006 Resilient machines through continuous self-modeling. *Science* **314**, 1118–1121. (doi:10.1126/science.1133687)
99. Edelman S. 2008 *Computing the mind: how the mind really works*. Oxford, UK: Oxford University Press.