nature reviews psychology

https://doi.org/10.1038/s44159-024-00283-3

Perspective

Check for updates

Uniquely human intelligence arose from expanded information capacity

Jessica F. Cantlon [1,2] & Steven T. Piantadosi [3,4]

Abstract

Most theories of how human cognition is unique propose specific representational capacities or biases, often thought to arise through evolutionary change. In this Perspective, we argue that the evidence that supports these domain-specific theories is confounded by general information-processing differences. We argue that human uniqueness arises through genetic quantitative increases in the global capacity to process information and share it among systems such as memory, attention and learning. This change explains regularities across numerous subdomains of cognition, behavioural comparisons between species and phenomena in child development. This strict evolutionary continuity theory of human intelligence is consistent with comparative evidence about neural evolution and computational constraints of memory on the ability to represent rules, patterns and abstract generalizations. We show how these differences in the degree of information processing capacity yield differences in kind for human cognition relative to other animals.

Sections

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA, USA. [2]Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA, USA. [3]Department of Psychology, University of California, Berkeley, CA, USA. [4]Helen Wills Neuroscience Institute, University of California, Berkeley, CA, USA. e-mail: jcantlon@andrew.cmu.edu; stp@berkeley.edu

Introduction

One of the deepest mysteries about cognition is how human thought differs from that of humanity's close primate relatives. This question targets the core of human nature itself, simultaneously pointing backwards to the evolution of natural intelligence and forwards to the likely mechanisms needed to make artificial intelligence.

The theories that have gained the most traction in the study of human uniqueness tend to be 'silver bullet' hypotheses that posit a single, specific adaptation or 'instinct' —such as social reasoning or hierarchical syntax —as the central catalyst of human-like thought[1–8]. Such domain-specific theories have largely dominated the field's hypothesis space. These theories tend to propose domain-specific or modular adaptations that are innate in humans and arose uniquely in human evolution.

Alternatives to a domain-specific evolutionary story have not reached a consensus on a general cognitive mechanism or biological basis for what is unique about humans[9–19]. However, several newer theories of human uniqueness suggest a shift towards more general learning adaptations[12,13,19–22]. These theories vary in how much innate structure they assume, from very little to a lot[23]. One class of theories that sometimes emphasizes the general cognitive origins of human intelligence is the cultural intelligence hypothesis[13,19,24–33]. The cultural intelligence hypothesis suggests that humans stand out among animal species due to their unique social learning ability and the accumulation of cultural knowledge it affords. Some variations of this hypothesis emphasize the role of purportedly innate, domain-specific mechanisms such as joint attention, theory of mind, language and imitation, whereas others emphasize the importance of global increases in general cognition such as the innovation rate[26] and flexibility[13]. Even cultural intelligence theories that posit innate adaptations to human social learning agree that domain-general learning and memory do much of the heavy lifting in the assimilation of new, unique human knowledge[12,31,33].

However, all these theories are inadequate for explaining in detail how humans excel at the diversity of tasks they do, and therefore fail to explain how unique capabilities emerge in humans. We argue that the differences between human and non-human primates are not due to one specialized cognitive adaptation or even a bunch of them. Instead, human uniqueness largely results from a global adaptation for increased information processing capacity, which alters human cognition profoundly and qualitatively. Information processing capacity refers to the amount of information[34,35] per unit time that can be stored and transmitted between cognitive mechanisms or subsystems.

In this Perspective, we propose that global, genetic differences in learning and memory are sufficient to account for uniquely human capacities across domains. We begin by detailing silver-bullet theories, cultural intelligence theories and our information capacity theory. We then review evidence for the information capacity theory by examining three major predictions: continuity of ability across species; differences in capacity limitations across species; and qualitative changes in ability enabled by quantitative changes in capacity. We discuss information capacity across human and non-human species and explore capacity limitations in learning models and mathematical analyses. We conclude that information capacity is a key determinant of human uniqueness because it determines which rules and representations can be learned by species across a range of domains.

Functional specializations within species preclude a unitary definition of intelligence across species (Box 1). Differences in intelligence between species might not be driven by the same parameters as differences within a species —we make no argument about differences within a species in the current Perspective. The definition of human intelligence specifically is also a matter of debate, but the term 'intelligence' is often used to refer to abilities that psychologists have identified as distinctive in humans —capacities such as language, recursion, theory of mind, massive memory, logical and relational reasoning, complex tool use and rich conceptual systems.

Theories of human uniqueness

Domain-specific theories have dominated the theory space around human uniqueness for decades and are still influential in more expansive modern theories of human cognitive origins[1,18]. The domain-specific focus on human origins has generated many different silver-bullet proposals. Silver-bullet theories of human cognition are often derived from sociobiological[36,37] or adaptationist theories of human behaviour[1,38,39] but they also stem from counter-behaviourist theories about human cognition[40] and from even earlier theories of biological preparedness that helped to spawn the cognitive revolution[23,41]. Some evolutionary frameworks propose that human behaviour is largely influenced by genetically specialized cognitive functions[1,18,42,43], whereas others propose much less innate structure[23,31]. Many of the hypothetical human-specific adaptations that various researchers propose are in the social cognitive domain, including cheater detection[2], theory of mind[44,45], imitation[46] and joint attention[47], whereas others are abstract computations such as symbolic geometry[3,48,49], recursion[50,51], relational reasoning[52] and language grammar[39,53–55].

Silver-bullet theories face multiple logical challenges in explaining the wide range of behaviours that set humans apart from other primates. One is that the key behaviours that set humans apart from other primates —behaviours such as language, mathematics, logical reasoning, complex social structures, tools and technology —generally must be acquired. Although each behaviour undoubtedly relies on some innate machinery, the specific representations that adults possess in any of these domains seem to be fairly unconstrained and critically unknown to the youngest human learners. A second problem is that the distinctively human domains are remarkably diverse in the representations and processes they require: a silver bullet for language is unlikely to be a silver bullet for mathematics or logic, much less for social reasoning and tool use. Researchers in these domains have tended to highlight their own area as the critical special one, while failing to account for uniquely human abilities in the others. Finally, each theory of human uniqueness posits that a different cognitive representation was the silver bullet, but they cannot all be the unique key piece. This observation has led some theorists to conclude that perhaps there was no single adaptation that led to human uniqueness but, rather, a patchwork of multiple specific adaptations[18]. However, a theory that post hoc posits evolutionary change in each domain lacks predictive power and parsimony because it has an unprincipled and unlimited attribution of unique adaptations for unique behaviours. More importantly, a patchwork adaptation theory contradicts the constituent domain-specific theories that form the patchwork, each of which maintains that humans are superior to non-human primates in one domain but ordinary in other domains[56].

Although some silver-bullet studies claim to disprove more general explanations of differences between humans and non-human primates, the pattern that emerges across studies is that humans differ in general ways from non-human primates, across many domains and cognitive processes. Human advantages extend beyond any one process or ability into basic object memory, semantic memory, set and

Box 1

Biological basis of information capacity differences

Addressing unique cognitive processes requires considering biological implementation. Unlike alternative theories, the theory that increased information capacity caused human intelligence has a plausible biological mechanism underlying human cognitive abilities and satisfies key criteria for a cause for intelligent behaviour in biological theory (such as Tinbergen's questions[274]). Species differences in information capacity are innate and rooted in their embryonic biology.

Functional specialization is a fact of species evolution[275]. Species differences in sensorimotor cortices firmly establish functional specializations that readily differentiate, for example, a bat from a possum from a primate. However, qualitative differences between humans and apes are much less profound[276,277]. Instead, there are well-known adaptive advantages to generalized increases in neural tissue for cognitive plasticity that drive unique cognitive abilities in humans.

Mammal and bird species with larger brains relative to their body size are more successful at surviving in novel environments than species with smaller brains[30]. This pattern suggests that increased brain size provides a fitness benefit in vertebrates[26], and implicates a global implementation of human advantages in cognition. Brain size in humans is expanded across most cortical regions relative to other primates[278,279]. Even the prefrontal cortex, which was once thought to be qualitatively distinct in humans, seems to have expanded largely quantitatively relative to chimpanzees[277,276].

The number of cortical neurons is related to information processing capacity, and can predict general cognitive abilities across species, such as the ability to control prepotent responses or make precise quantitative judgements[179,183,265,280]. The number of neurons is genetically determined, set before birth for each species by the size of their neural progenitor pool and duration of cortical neurogenesis[264,265,281–283], and is therefore a likely substrate for adaptation[284]. Humans have an extended period of cortical neurogenesis during prenatal development compared with other primates[269,285]. Extended cortical neurogenesis probably increases neural density in human cortex generally, although it might be more likely to increase frontal density than posterior density because prolongations are more efficiently implemented there within the developmental schedule[281]. Prior theories argued that human prefrontal cortex underwent special expansion relative to other brain regions for functionally specific reasons (for instance, cognitive control[14]). But prefrontal expansion seems, instead, a continuation of a new allometric trend, a yoked scaling between brain parts, that began in a more distant ancestor shared with other apes[281], and might be a consequence of general selection for neural density under the ease of embryonic rostral–caudal neurogenesis rather than specific functional selection for a frontal cognitive process.

Although there is growing genetic evidence in favour of species-specific circuits for general learning and motivation[281], there is no parallel evidence for circuits supporting purported silver bullets of human uniqueness, including language, theory of mind, cheater detection or relational reasoning. In fact, genetic evidence for these domain-specific theories, such as predicted differences between identical versus fraternal twins or predicted task-specific correlations with genes, does not bear out in the data[13]. More broadly, there are no theories for how new genetic cognitive modules such as these would arise, nor how they would have arisen in humans alone. Finally, existing theories of human uniqueness cannot explain why corvids perform so similarly to non-human primates on many tasks[286]. By contrast, the information capacity theory easily explains this as the result of similar neural densities between corvids and primates that afford similar flexibility and learning rates[287,288].

Evolutionary theories of human intelligence that take the chimpanzee mind as a starting point and factor the effects of scaling-up on cognition are grounded in genetic physiology. The information capacity theory proposes that increased neural and glial density, determined by the size of the neural progenitor pool at birth, qualitatively improves cognition. A plausible pathway to human uniqueness is one in which genetic changes in the human brain increase neural density quantitatively, which leads to exponential increases in information processing capacity, and results in conceptual learning leaps during child development that are unseen in other primates. This theory provides a more straightforward account of cognitive evolution by natural selection and predicts the unmistakable biological continuity between primate and human brains.

sequence memory, memory duration, cognitive control and simple learning rates —differences in these general cognitive functions are not predicted by domain-specific theories.

Theories that attribute human uniqueness to domain-general adaptations, such as the cultural intelligence hypothesis, tend to focus on evolutionarily new general-purpose neural functions and structures[57], such as 'quantum leaps' in adaptive specializations[58], cognitive control in the prefrontal cortex[59,60], adaptations for abstraction[52] and other evolutionary discontinuities[18]. These theories propose an evolutionary discontinuity between humans and non-human primates. There is general agreement among cultural intelligence theories that humans are unique in their social learning abilities, which enhance human cognition through the accumulation of knowledge in culture[18,19]. Some cultural theories effectively propose silver bullets by emphasizing innate, domain-specific mechanisms such as joint attention, theory of mind, language and imitation as key catalysts for human uniqueness[18,28,31–33,46]. Other theories emphasize expansions in general cognitive mechanisms such as learning, memory and executive function as the evolutionary foundation of human innovation, complex communication and flexibility[13,17,25,26,61–63]. These latter theories show that global increases in general cognition are adaptive and hypothesize that general cognitive processes ratchet up unique cognitive abilities that define humans through adaptations and environmental opportunities for social learning. However, none of these prior theories explain how general capacities benefit uniquely human cognitive abilities.

By contrast, we propose that a uniquely human genetic change in global information capacity, implemented computationally as an increased bit rate and biologically as increased cortical density, is powerful enough to yield all of the qualitative differences between human and non-human cognition. Thus, our idea differs from other domain-general theories because it is a strict continuity theory of the evolution of human uniqueness. Our theory is synergistic with versions of the cultural intelligence hypothesis that propose evolutionary expansions in learning and memory as a key source of human uniqueness[13,29,30]. It builds on constructivist theories of conceptual change in childhood[64–66] and complements facets of the cultural intelligence hypothesis that emphasize the uniqueness of learning inputs during human development[12,19,31]. Our hypothesis expands those prior ideas by offering a cognitive explanation for how so many qualitatively new types of cognition could emerge in humans through simple, quantitative changes in their global information capacity over evolution.

We focus on how human cognition is unique at the cognitive algorithmic level. However, information capacity constrains cognitive systems because neural energy is required to represent information and transform it computationally[67,68]. The key for human-like cognition is that all neural systems must work together —improvements in social reasoning or tool use will not help much unless information in these systems can be conveyed to systems for learning, memory and inference. In turn, learned and inferred representations guide systems for perception and interface tightly with memory architecture. Increased global information capacity would not only be advantageous but perhaps also essential for any expansion of cognitive faculties in human evolution, ensuring that enhancements in one domain could be effectively integrated and utilized across the cognitive system. This mutualistic, integrative view directly engages the evolutionary reality of human uniqueness in global neural density, the behavioural predictiveness of such neural measures across species, and the likely ease with which general information capacity parameters are changed genetically in evolution (Box 1).

Our theory yields three concrete predictions. First, it predicts some degree of success for all species even in domains that are argued to reflect unique human ability. To address this prediction, we review the comparative cognitive and child development literature, demonstrating a continuity of success across humans and non-human animals. Second, the theory predicts quantifiable differences between humans and other species on basic information processing measures. To this end, we review literature documenting predictable and systematic performance gaps between species across most cognitive tasks, including simple memory and learning paradigms. Last, our account predicts that small changes in information capacity could yield big, qualitative changes in behaviour. We present mathematical analysis, machine learning models and cognitive models for which capacity constraints have profound consequences.

Continuity of ability across species

The first prediction of the information capacity theory is that non-human species will exhibit some degree of success on cognitive tests argued to capture the key to human uniqueness. Relational reasoning, social reasoning, causal reasoning and tools, and symbolic thought and language have been the focus of major silver-bullet theories (Table 1). We discuss each of these domains in turn and show how evidence in each supports our first prediction of partial success by non-human primates.

Social reasoning

Multiple silver-bullet theories posit social reasoning as the key to human uniqueness, claiming that one or a few social computations that humans engage in are genetic adaptations[8,33,42,46,56,69–71]. One long-standing hypothesis is that human intelligence was sparked by 'theory of mind' abilities[56,72,73], or the capacity to predict behaviour based on knowledge that another being has a mind. Some theorists even argue that theory of mind is an innate module[44,45]. One problem with the notion of theory of mind as an innate human adaptation is that its emergence in childhood depends on experience and its learning trajectories are gradual and vary widely[8]. General cognitive capacities such as working memory and relational reasoning increase with age and, along with the variability in the frequency of social learning input, result in gradual and variable age-related development of theory of mind in humans.

Non-human animals can reason about the mental states of others, including goals, views and actions. The difference in social reasoning between humans and other species lies in the breadth and depth of how it is deployed across contexts and complex tasks[74–76]. For instance, similar to young human children, non-human animals possess shallow knowledge about agents that is linked to concrete features of the world rather than internal states such as deception, ignorance or surprise[77].

Non-human primates have demonstrated all cognitive constructs thought to be essential to human social intelligence and uniquely human: theory of mind[78–81], imitation[82,83], gaze following[84], joint attention[85,86], teaching[87,88], helping[89,90] and cooperation[91]. These findings demonstrate that humans and non-human primates share substantial social cognitive continuity.

The difference between humans and non-human primates is often one of degree[8,84,92,93]. For example, a study of imitation found that some chimpanzees and capuchin monkeys could socially learn one-stage and two-stage puzzle box behaviours but that only human children (aged 3–4 years) imitated and modelled the longer and more complex three-stage solutions to the puzzle box task[94]. The degree of complexity also influences social reasoning in chimpanzees during teaching behaviours[89]. These findings indicate that the amount of information processing demand causes species differences in social learning.

Although most theorists think that human social cognition is rooted in discontinuous cognitive 'traits'[95], another way to see theory of mind is as a continuum from basic action prediction to complex belief attribution[96], or from shallow reasoning based on representing others' knowledge to abstract reasoning based on inferred beliefs[97,98]. Children and non-human primates are often characterized as able to reason about rational action or behaviour and adult humans about mental states[94,98,99]. Differences in complexity and abstractness between human and non-human social reasoning imply a bottleneck that is quantitative rather than trait-based or qualitative[100].

Many cultural intelligence theories emphasize social learning specifically as a key source of human and non-human differences in social cognition[13,19–21]. For instance, biological preparedness accounts focus on biases in shared mechanisms as the hypothetical locus of genetic differences between humans and other species[99]. These theories emphasize innate biases such as prosociality in mechanisms such as attention and motivation. There are clear differences in successes and failures on social learning tasks between human children and non-human primates[19]. But the innate biases have a comparatively small role in those differences because their function is primarily to alter the frequency of encountering different learning inputs —which means that the bulk of human uniqueness arises from learning. Importantly, non-human primates cannot acquire abilities at the level of a human

Table 1 | Domains of human uniqueness

Domain	Proposed human uniqueness	Evidence for continuity across species	Proposed influence of information capacity
Social reasoning	Theory of mind	Human development of theory of mind depends on age[8,245,246], relational reasoning capacity[247,248] and experience	Working memory, relational reasoning capacity and the frequencies of learning inputs increase quantitatively with age and determine theory of mind ability in humans
		Non-human animals can reason about simple, limited mental states of others such as gaze–action relations[88–93,99,101,102]	Information capacity limits non-human animals to low-complexity mind–behaviour relations versus more complex ones
		Young children and non-human primates' reasoning tends to be limited to shallow, concrete and perceptual features of agents[78–93,95,96]	Shallow processing requires fewer layers of information processing than deep processing and is more likely when information capacity is limited
	Social learning and imitation	Non-human animals can socially learn single items, associations and short sequences with short durations but not more demanding sequences[91,94,97,103]	Quantitative capacity differences limit the number of items and the duration over which sequences are held in memory — learning capacity will be worse or absent for longer sequences
		Non-humans can socially learn some tasks that children aged 3–5 years learn[94,100] but differ from human children quantitatively, even in solo learning rates for those tasks[103,104]	Information capacity in non-human primates is insufficient for human-like learning; social learning demands more capacity than solo learning because it is passive and occurs over narrow bouts and durations
Relational reasoning	Relations between relations	Reasoning about relations between relations (quaternary relations) develops gradually in humans, in step with general cognitive capacities[105,233,246,248]	Humans need sufficient information processing capacity to manage the demands of quaternary relations, which results in gradual development of relational reasoning
		Apes and young children succeed at relational reasoning for unary, binary and ternary relations but similarly fail tasks requiring reasoning about quaternary relations[103–105,107]	Young human children and apes both have lower information capacities than older children and adults, which results in similar reasoning performance
		Young children and non-human primates succeed at quaternary relational tasks when they are given symbols for the relation[69,104,108,110,111]	Symbols benefit relational reasoning in a general way, by increasing information capacity for demanding relational problems in any species
		Corvids and monkeys succeed at simple relations-between-relations tasks[113,114,117]	Species with high neural densities succeed at perceptual relational matching tasks because relational reasoning depends on general information capacity
Causal reasoning	Higher-order causal relations	Non-human primates and young children often reason associatively rather than relationally about causal problems[119,124,128,130]	Differences in information capacity result in variations in the depth and breadth of causal reasoning abilities between species
	Complex tool use	Some non-human primates use tools, but their routines typically involve binary relations[118,120,124,127]	Information capacity constraints limit the depth and complexity of tool-use routines in non-human primates
		Humans have more complex routines[122,126], often involving multiple objects and nested subassembly routines, than other species[122–125,127,128,130]	Differences in information capacity lead to quantitative disparities in acquired tool complexity and diversity between species
		Non-human species that create tools often have limited repertoires of one to a few types of tools[121–131]	The capacity to hold in mind a wide range of actions and objects allows humans to develop more extensive tool repertoires
Symbolic thought and language	Symbolic representations	Non-human animals can learn symbols that represent objects and verbs[165,167,168], and tokens to represent relations[111,112,115,171], but the depth and number of such representations is smaller than in humans[169,173,174]	The synergy between language and information capacity is self-reinforcing, and humans' innately higher capacity yields faster acquisition of symbols and concepts compared with non-human animals
		Symbol-trained chimpanzees show human-like benefits in reasoning about relations between relations[111,112,114,115]	There is a general synergy between symbols and information capacity, not specific to humans
	Recursion	Human recursion has limits, as evidenced by the struggle to understand sentences with multiple levels of recursive embedding[133–136]	Even humans need sufficient information processing capacity to manage the demands of hierarchical embedding
		Some success for non-human species on hierarchical tasks, specifically for high-capacity species such as monkeys and corvids, suggests a general origin for recursion in humans[157,158]	Non-human species can learn hierarchical patterns but do so more slowly and shallowly than humans; differences between humans and non-human animals in the capacity to synthesize large amounts of information yield differences in hierarchical thought

child even when their environment offers higher frequencies of direct social input. This difference means that there is something critical in the learning component of social learning that differs between humans and non-humans beyond innate biases in sociality. Learning rates are constrained by information capacity, and because social learning occurs by observation it requires the ability to learn passively and on someone else's schedule and pace, which makes social learning more demanding than individual learning. Although learning rates for individual versus social learning conditions are rarely compared between species, we suggest that the time course of social learning required to

attain human-like characteristics is outside the learning capacity for non-human primates on many tasks. Thus, information capacity could have a causal role in determining the successes and failures of social learning across species.

Relational reasoning

Other silver-bullet theories suggest that human uniqueness rests on forms of non-social reasoning such as the proposed 'super-module' of relational reasoning[52,101]. Relational reasoning requires deep logical representations —minimally, it requires comparing two relations across four entities, a quaternary representation. For instance, relational reasoning is involved in comparing one relation with another, such as knowing that the relationship between the concept of a dog as it relates to a doghouse and a bird as it relates to a birdhouse is the same.

Many species form common class and category concepts but often fail at tasks requiring reasoning about relations between relations[102,103]. Both non-human primates and children younger than 5 years of age often fail at these tasks[69,104]. Some researchers have argued that they can succeed only if given a symbol to represent the relation (such as the word 'same' for children)[104]. Consistent with this account, children younger than 4 years of age succeed at relational tasks only when given relational labels, whereas children older than 4 years of age spontaneously succeed —presumably because older children have acquired the necessary relational labels[69,105,106].

If there is a continuity of ability across human children and non-human primates, then providing similar symbols to non-human primates should enable them to succeed at more complex relational reasoning. Indeed, chimpanzees provided with symbols (language or shape tokens) succeed at quaternary relational tasks[104,107–109]. The pattern that only 'apes with symbols' succeed at relational reasoning tasks suggests that tokens of some sort are the key to comparing relations. However, there are exceptions to this generalization[110–112]. For instance, corvids (birds with similar neural densities to monkeys) can succeed at relational matching tasks (distinguishing same and different entities) involving shape, colour or size[113]. However, apes and bonobos succeed more readily than corvids on the same task, and comparably with a 3-year-old child[114]. These latter findings indicate that relational reasoning is not uniquely human or strictly language-dependent —it can be learned by non-humans and is influenced by general capacities such as a species' learning rate.

Overall, a strict version of the hypothesis that the presence of relational reasoning in humans distinguishes them from other species is probably not correct. Instead, humans might have an easier time processing relations, in part, due to the power of language to package concepts into lighter-weight mental representations, or chunks. Another important component is likely to be humans' greater general information processing capacity, which can also contribute to effective management of the informational demands of quaternary comparison. Humans' higher capacity for relational reasoning probably has broad applications across domains, including supporting sophisticated concepts in social and physical settings.

Causal reasoning and tools

Capacities in causal reasoning have also been posited as the silver bullet of human uniqueness. Humans' causal reasoning is linked to exceptional exploitation of the environment, such as tool use and novel foraging behaviours such as armed hunting and farming[18]. Although some non-human primate species can use tools[115], identify useful tools[116] and create tools[114], humans have the ability to create a much higher number of novel tools[18,117].

Non-human species that create tools —such as rooks, ravens, crows, capuchins and chimpanzees —often have limited repertoires of one to a few types of tools, whereas humans have larger repertoires by the time they are 5 years old[113,118–121]. Chimpanzees have the largest tool repertoires among non-human species and can use around 20 different tools[122]. Human children have large tool repertoires, which they initially learn from observing experienced teachers, but by age 7–8 years they go further and begin creating tools[121].

Non-human primates have limited insight, breadth and generalization of even simple causal relations[119,123,124], which probably constrains their tool use. In the tube-trap task, in which a target object must be removed from a tube by using a probe while avoiding trap obstacles, non-human animals show a lack of insight and minimal causal generalization when presented with novel but similar tubes. Rooks and chimpanzees show near-transfer to new similar tube traps after learning to solve the original tube, but not far-transfer to visually distinct tubes[124,125]. Minor tweaks to tube-trap apparatuses can help non-human primate performance and non-humans, particularly apes and corvids, solve these simple tool-use tasks similarly to children who are younger than 5 years old. Specifically, young children and non-human primates can make some simple causal inferences about traps but often revert to narrow associative reasoning such as side biases that do not generalize to novel tasks[126]. However, as children approach 4–5 years of age, they quickly acquire broad causal knowledge of tube-trap tasks and generalize to rotated and reconfigured tubes[124]. Older children and adults not only learn action–object associations during tool use but make broad causal inferences that generalize to other tasks. The presence of some successful learning and generalization in non-human animals implies that causal reasoning is not entirely absent[126]. Rather, differences in learning and generalization between humans and non-human animals could be caused by the amount of information required to solve a physical problem. The types of causal concepts and tool-use routines that non-human primates and other animals learn are probably shorter, simpler and less enduring than those of humans.

Tool use in non-human species often involves a single object impacting another object but rarely requires multiple objects organized into an action hierarchy. The integration of multiple objects and actions into a behaviour is rare in non-human species[11]. Although some species use objects in binary combinations, such as hitting a nut with a stone, non-human tool use lacks nested subassembly routines, which are observed frequently in human behaviour[11]. For example, field data show that non-human animals do not frequently implement hierarchical action routines where one tool is used to make another tool[11]. These patterns suggest that the limitations that non-human animals face in causal reasoning are quantitative ones in the breadth and depth of action abstraction, and the number of nested levels of subroutines. These quantitative limitations are plausibly explained by differences in information capacity.

Symbolic thought and language

Many theories of human uniqueness centre on language, but the specific evolutionary changes responsible for this ability are debated. Some theorists have suggested that recursion is the defining element of uniquely human language[11,50,51,127]. Although the term was not defined in these proposals, the example given by these authors was sentential embedding, the ability to take a sentence S (such as 'It is raining') and put it inside another sentence, such as 'Mary thinks that S' ('Mary thinks

that it is raining')[51]. This capacity presumably provides the ability to build complex, hierarchical expressions that are predicated on other complex expressions. Despite the importance of recursion, humans face a limit of two levels of recursive embedding in certain constructions, struggling to understand sentences such as 'The senator the chef the mouse saw attacked laughed'[128,129]. Additionally, not all human languages use sentential embedding[130–132], and it has been argued that relatively little daily language use involves recursive capacity[133,134]. These discrepancies challenge the notion that recursion broadly is the keystone of human-like thought.

A revised proposal is that the relevant sense of recursion for human uniqueness is the capacity to unboundedly put two arbitrary elements together into a new unit[135–138]. For example, the sentence 'Wendy ate marshmallows' is recursive in this sense because it has three words. An operation that can only put two words together at a time must therefore operate multiple times on these words, including on its own output. For example, most theories assume the verb and object are combined into a new unit, {ate, marshmallows}, and then that unit is combined recursively with the subject to form {Wendy, {ate, marshmallows}}.

This idea critically relies on the assumption that even simple sentences involve underlying binary structures. Some theorists have argued that binary structure building is the simplest computational operation[55], the key to human thought and language[53], and the result of a single mutation leading to a rapid evolutionary change[139]. This proposed capacity is distinct from non-human animals' ability to generate sequences of behaviour as sequences need not have the required hierarchical organization[140]. However, this proposal has been critiqued logically and in evolutionary models[141,142]. For instance, there are properties of language that cannot be explained by minimal change to enable recursive structure building, including features that are useful in communication, even when only partially present[6]. Communicative properties of human language tie it to animal communication systems in that they suggest continuity in the evolution of language, driven by the pressures of usage, rather than genetically determined discontinuity[143,144]. Such communicative explanations in linguistics have found empirical support across linguistic subdomains[140,145–148], often showing how information-theoretic or usage-based constraints shape the form of linguistic systems[149].

More generally, the existence of innate grammatical constraints — recursive structure building or others —is widely contested[150–153]. It has been argued that, essentially, none of the key properties of language previously suggested to be universal across human groups actually are[154].

Instead of language-specific adaptations, the human capacity for representing and synthesizing large amounts of information might yield more complex, hierarchical patterns of thought. Humans tend to attribute tree-like structures to data[155,156]: people across ages and cultures generalize ambiguous data hierarchically. In one study, Indigenous Amazonian adults, American adults, American children and non-human primates (macaques) were taught sequences of symbols[157]. All humans tended to generalize to novel combinations of the learned symbols hierarchically rather than using an ordinal strategy, even though either strategy was consistent with the training data. Macaques only used a hierarchical strategy with additional training, suggesting that hierarchical generalization was not out of reach for them but was a weaker bias. Children's use of a hierarchical strategy in this task was predicted by their working memory capacity, suggesting that the development of hierarchical reasoning is gradual and limited by information capacity. Corvids, a family of birds with similar neural densities to primates, perform at least as well as macaques on the same task, further highlighting continuity between humans and the rest of the animal kingdom in the capacity to generalize hierarchically[158]. These data that show some success for non-human species on hierarchical tasks, and specifically for species with high information processing capacity, implicate a more general origin story for recursion in humans that does not depend on specific adaptations for language.

A general problem for language-based theories of human uniqueness is that they are at odds with the continuity between humans and non-human animals in many aspects of language learning[159]. Non-human animals can learn hundreds of words and symbols, generate strings and comprehend basic syntax[160–164]. Baboons show expectations about some aspects of semantics in word learning tasks[165] and exhibit some capacity for compositionality[166]. A trained bonobo (Kanzi) could comprehend sentences and differentiate them, but had limited capacity with coordinate noun phrases and a shorter mean utterance length than a 2-year-old human child[167]. Although these differences are sometimes interpreted as evidence of a unique human genetic adaptation for grammar, they could also reflect quantitative differences in abilities that snowballed into qualitative ones. For instance, insufficient capacity to deal with parts of a phrase can impair one's access to the hierarchical structure of sentences, regardless of a grammar-specific adaptation.

General computations related to language and symbolic structures have also been proposed as uniquely human. Humans show fluent learning of symbolic mental algorithms and procedures[3,48], including advantages over non-human primates in domains such as geometry[49]. However, it is unclear what class of algorithms other species are capable of acquiring and whether their limitations are due to memory or underlying algorithmic competence (Box 1). Indeed, the question of uniquely human language abilities might not even make sense if underlying memory capacities differ between species.

Informational limits are a known constraint on language learning and might be critical in how children acquire language[168]. However, non-human primates face more drastic information limits. For instance, non-human primates' inability to learn human-like language has been attributed to their difficulties learning and remembering even lexical items[167]. Similarly, humans are thought to uniquely possess the ability to learn massive numbers of arbitrary symbols; human knowledge of tens of thousands of arbitrary word meanings contrasts greatly with the few dozen calls known to non-human primates[169]. Many accounts of human language emphasize the role of memorized structures or chunks (rather than a small set of syntactic rules) in determining linguistic competence[170,171]. Human learners acquire, on average, 2,000 bits of information per day about word meanings, from birth to adulthood[172] —a monumental feat of memory with probably no parallel in animal cognition.

Information capacity constrains the ability to learn language but language also enhances information capacity once learned. The cognitive changes associated with symbolic representation are believed to be profound. Compositionality enables new conceptual structures and information transfer between domains[173] and grammar provides an "endless compacting of information limited only by human memory"[174] akin to hierarchical organization for memory or concepts[175]. Words provide a handle on complex meanings[176]. For instance, the word 'hundred' does not need be broken down into 'ten tens'[177] and the word 'aunt' need not be broken down into its component defining relations to access these meanings. This synergistic relationship between symbols

and information capacity is not unique to humans. For example, as described above, chimpanzees trained with symbols for 'same' and 'different' learned to reason about relations between relations more easily than those without symbol training —showing clear benefits of symbols for complex rule learning even in non-human primates[114–116].

We conclude that human uniqueness in symbolic thought and language lies not in an exclusive adaptation for grammar but, rather, in humans' extraordinary capacity to manage vast quantities of information. Non-human animals have demonstrated their ability to learn words, construct sequences and grasp basic syntax to a certain extent, challenging the notion of exclusive grammar adaptations in humans. What truly sets humans apart is their unparalleled information capacity and their ability to transcend information limits in language acquisition. Unlike any other species, humans possess remarkable learning and retention of an extensive array of symbols and rules. Once acquired, language serves as a powerful tool for expanding information capacity. This synergy between symbols and information is observed in symbol-trained chimpanzees, suggesting that successfully acquired symbols benefit reasoning in other animals too, but the outcome of this synergy is amplified in humans because they store more symbols, including symbols upon symbols in recursive structures.

Species differences in capacity

The second major prediction of our information capacity theory is that human and non-human primate species differ in their general information capacity. We predict that species will differ on basic, domain-general cognitive tasks involving general learning and memory.

All domains that are purported to define human uniqueness — including complex relational concepts such as social relations, mind–behaviour relations, analogies, rule induction and grammar —are probably constrained by more fundamental domain-general limitations on information capacity. Human adults are also subject to limitations on information capacity (Box 2). Cognitive information processing capacities such as learning rate, attention, memory and control vary greatly across species[10,14,29,62,178–184]. Such differences feed into every other cognitive process, with the consequence that performance on any task requiring learning, attention, memory or cognitive control will differ across species because of those underlying constraints. It is therefore hard to posit fundamental representational differences in any individual task without accounting for global informational factors. Moreover, such differences feed into learning, which can amplify the differences: animals use informational capacities adaptively to learn new things and therefore some species might end up with fundamentally different representations because of their information capacity rather than domain-specific factors. Four key domain-general capacities are particularly relevant to cross-species comparisons: simple learning rates, simple memory capacity, cognitive control, and sequence learning and memory. We review these four capacities and their impacts on purported human-unique domains here.

Simple learning rates

Some species learn novel associations, generalizations, rules and strategies more quickly than others[61,185,186]. For instance, oddity learning, in which animals must tip over the odd object between three wooden objects on a board to find a food reward, has shown species differences. In one typical experiment, chimpanzees and monkeys learned to decipher which object was the odd one above chance (60%) in 1,152 trials, whereas cats and raccoons performed at chance for 4,800 trials[187]. Chimpanzees reached 90% accuracy at 2,208 trials, whereas monkeys took 3,508 trials. Human children aged 3–5 years learned to 90% accuracy in one fifth of the trials (mean = 203 trials)

Box 2

Capacity limits in adulthood

Capacity limits are a key constraint in human adult cognition, apparent in all posited silver-bullet domains. In many cases, adult human abilities bump up against processing limitations, which suggests that species with different limitations should not show behaviour that is qualitatively human-like. For instance, adults exhibit upper limits on the number of meaningful chunks of information they can hold in mind at one time, which interacts with the resource demands of the task[289]. Effects of capacity limits on adult cognition are observed in relational, causal and social reasoning and language. For example, adults have highly limited abilities to understand recursive embedding[136,137]. Furthermore, adults' analogical reasoning shifts from abstract and relational to more concrete and object-based when information processing is taxed, such as with increased working memory load[290]. Finally, human social interactions become more challenging and less cohesive if there are more than four people involved in the discourse, a phenomenon known as 'the dinner party problem'[291]. Thus, adult human cognition is limited by its information capacity, and there are certain rules and patterns that are too long or complex for humans to comprehend.

'Rational' accounts of cognition attempt to explain behaviour in terms of what would be expected from an 'ideal' learner who has a perfect memory[292,293]. However, models that posit that adults are efficient at solving problems have obvious difficulty approximating behaviour when adults' behaviour is suboptimal. A growing area of computational modelling seeks to understand so-called resource rational models, which formalize optimal or efficient use of finite mental resources such as memory or attention[294,295]. Resource rational models posit that people are efficient relative to their resource constraints, but their resource constraints prevent them from achieving normatively perfect behaviour. This type of approach can be seen in theories of human perception[296], decision-making[297,298], encoding of subjective value[299], generalization[300], processing of centre-embedding linguistic structure[301] and number cognition[302]. Across these domains, models with limited resources provide a better account of human behaviour than those that assume unlimited or unrestricted resources. Limited capacity is therefore a promising approach to understanding cognition that highlights ways in which finite informational capacity shapes how humans solve problems.

and those aged 6 years in one twentieth of the trials (mean = 96 trials) compared with chimpanzees (mean = 2,208 trials)[188]. With instruction, human children learn even faster than they do by trial and error[189]. Thus, humans learn abstract rules faster than primates from early childhood, due partly to an inherently higher learning rate and partly to instruction, enabling more abstract rule learning in humans compared with other primates. Abstract rules with arbitrary stimuli such as wooden blocks are unlikely to arise from domain-specific adaptations and silver-bullet representations because the target of learning is arbitrary stimuli and the rule applies to features of those stimuli generally. Instead, species differences on these rule-learning tasks indicate varying domain-general capacities.

Learning rates vary even within narrower clades of species, such as among primates[10,190]. In basic learning set tasks[185,187,191], animals must choose correctly between two arbitrary stimuli (such as a cylinder and a cube) to get a food reward. Once they do so, two new objects (such as a sphere and a cone) are presented and they must learn which is rewarded. Attention set-shifting involves switching between stimuli on the basis of one dimension (such as shape) and then another (such as colour) within the same task. Animals learn the overarching task rule rather than just an association between object or feature and reward, as demonstrated by the fact that they learn faster with each new contingency[10]. A meta-analysis of learning set performance with primates revealed that apes exhibit faster learning than Old World monkeys, and Old World monkeys perform better than prosimians[10]. Humans, including young children, acquire learning set rules in a fraction of the trials it takes apes and monkeys to learn[192,193]. As with oddity learning, species differences in this task are attributable to differences in underlying information capacity. All species we described can learn the basic associative task and the task sets with arbitrary stimuli but they consistently differ in the rates at which they acquire and generalize those contingencies, implicating general learning differences rather than differences in domain-specific adaptations.

Reversal learning is similar to learning set tasks but requires the exact opposite choice of what was just learned. During reversals, the previously unrewarded object becomes the rewarded one and vice versa[61,194,195]. Animals who learn broad task strategies are more successful on the task than those who rely on associations, as they learn to represent the task strategically with a rule (such as 'reverse' or 'win–stay, lose–shift') and can adapt when an item that yielded a reward for many trials suddenly becomes unrewarded. Some species (including apes) show fast rates of acquisition on both initial learning set and reversals, but others (including lemurs) show fast learning rates for sets but not reversals. Thus, some species easily learn simple associations but do not learn abstract rules as easily.

Reversal learning shows qualitative differences in task performance between species. Species with lower learning capacities are unable to transfer their learning. After learning one association ('choose A, not B') they must build up a new association de novo ('choose B, not A'), so each reversal takes them a long time to learn[186]. There was substantial variability among 79 primate species who were trained to identical criteria on reversal learning[186]. Some primates such as apes learned a global strategy or rule whereas other species such as prosimians learned a series of associations. Some animals lack the flexibility to readily derive an abstract rule, perhaps in part because they perseverate or adhere to known successful responses.

These results suggest that general learning capacity yields qualitative differences in cognition because different species solve identical tasks in different ways. Systematic performance gaps emerge between species even across simple, domain-general tasks—which supports the second prediction of our theory.

Simple memory capacity

Humans have a large memory capacity, estimated at a billion bits[196], with language alone requiring 12.5 million bits[177]. With training, human memory capacity can be substantially improved[197,198]. Humans can learn and remember more than 2,500 new, unique visual items in a simple 5-h experiment and are thought to be capable of maintaining more than 200,000 unique visual items in memory at a time[199]. Using a comparable paradigm, pigeons required 700 sessions to learn 1,000 images and baboons required 3–5 years to learn 3,500–5,000 images[200]. The pigeons and baboons showed set size effects, suggesting similar underlying mechanisms to human memory[201]. Baboons' ability to learn four times more stimuli than pigeons suggests genetic differences in simple memory capacity between primates and pigeons. Humans are estimated to have 50 times the memory capacity of baboons[202] so those genetic differences in simple memory capacity between pigeons and non-human primates plausibly extend to differences between humans and non-human primates as well. Such memory differences are certain to cause profound species differences in cognitive domains such as relational reasoning, social cognition, complex action and tool use, and complex rule learning that require representing many items such as symbols, words and concepts.

Cognitive control

Flexibility during problem-solving is a criterion for intelligent behaviour[29,183]. Some species are more likely to 'get stuck' during problem-solving than others. There are a few general mechanisms relevant to flexibility, which might be termed control, attention, inhibition, self-regulation or executive function. These mechanisms have a role in tasks that require remembering the past, avoiding errors and consciousness, all task behaviours that are proposed as markers of human uniqueness attributed to unique neural evolution of human prefrontal cortex[14]. However, the role of these general cognitive mechanisms in behavioural flexibility is enhanced not just in species with larger prefrontal regions but also in species with larger overall brains and relatively small prefrontal cortices.

Two self-regulation tasks include the A-not-B task, in which a prepotent response (A) is built up and then a new response (B) is prompted and researchers measure the time needed to adopt the new response[203], and the perseveration test, in which a naturally prepotent response is unsuccessful and researchers measure how long it takes subjects to change course[202,204]. Species vary in their ability to control prepotent responses in these tasks, with larger-brained animals such as apes, elephants and dolphins exhibiting the highest degree of control, and marmosets and rats showing the weakest control[183]. Overall neural densities, which are genetically determined, seem to influence an animal's capacity for self-regulation. Humans rank high in control relative to other primates, starting around age 2–3 years[205], which suggests that human advantages in cognitive control develop early. The early development of advantages in self-regulation suggests a genetic and general cause for human advantages in cognitive control. This ability is essential for conceptual change and complex learning—for example, substantial evidence shows that these general cognitive capacities are critical precursors to relational reasoning in human children[206]. Together, these findings implicate a general and evolutionary basis for species differences in acquiring simple rules, hinging on their ability to rapidly integrate new information into new action.

Box 3

Information capacity constrains rule learning

A computational model can be used to evaluate how capacity influences what structures a learner can acquire. Domain-general pattern learning models are able to find concise algorithmic descriptions of discrete data sequences[303]. For example, when provided with input such as 'abbabbabb', such a model would learn an iterative or recursive representation that generalizes to longer sequences following this pattern, such as 'abbabbabbabbabb ...'. When given input such as 'aaabbbccc' the model would generalize to both shorter ('aabbcc') and longer ('aaaabbbbcccc') strings that also follow this $a^n b^n c^n$ pattern. This program learning model can acquire key structures in natural language, including different kinds of recursion and hierarchy by constructing representations of grammars out of a handful of algorithmic pieces, just as programmers build complex programs out of a few dozen built-in operations[304].

Very similar models can be found in, for example, sequence memory[305] and other cognitive domains[48]. All such models work by finding programs that provide a concise description of the data. For example, if learners saw the sequence 'abbabbabb', they might construct a program F defined as F() : = repeat(pair('a', pair('b', 'b')), 3). Here, F first pairs together 'bb' from its component parts (using pair('b','b')), and then pairs that structure with an 'a' to yield 'abb' (using pair('a', pair('b', 'b'))). The 'repeat' operation then repeats the entire 'abb' structure three times. Thus, this program F is one way to describe the regularities in the string 'abbabbabb'. The task of the learner is to find program F by searching over ways of composing the parts to capture the data (or an approximation to it). Such models work essentially similar to scientists who observe data and try to craft formal theories to explain the data they see. The models are typically biased to find short or concise programs to capture the data they observe, following 'minimum description length' accounts of statistical inference[306] as well as theories of artificial intelligence that argue that an intelligent agent should try to find concise programs to explain observations[307,308].

The complexity of pattern a learning model can find in a sequence is a function of the length of the data its memory system can handle. We used a version of a sequential rule-learning model[308] that includes several operations including repetition (such as 'repeat'), alternation, reversals, list-building functions (such as 'pair') as well as recursive and logical operations[309] to demonstrate this point. This model also enables learners to generate arbitrarily long sequences, meaning that one could hypothesize that the observed data were part of an infinite longer sequence. We looked at the length of every possible data sequence (such as 'abbabbabb') versus the length of the shortest computing program (rule) that would produce it. The key is that if there is a program that is shorter than the data sequence, this means that there is a pattern present that a learner could detect. Generally, for short sequences there is no program shorter than the data, meaning that even an ideal learner would memorize these sequences and not find any pattern, nor be able to generalize a pattern to longer sequences. However, once the learning model had enough memory to use more than four or five items, it became possible to detect patterns in the input and create a shorter generating program, just as in the ABAB example. Thus, idealized learners with memory of fewer than four or five items would not be able to discover much structure from the world, even in principle. In this way, limits on information capacity obscure competence.

Sequence learning and memory

The capacity to hold multiple items in mind at the same time is needed to solve complex problems. For example, to recognize that a pattern such as 'ABBA' also applies to 'CDDC' and 'EFFE', one must hold the four elements of 'ABBA' in mind, extract a rule and apply it to four new elements (Box 3). Species vary in their ability to hold multiple items in mind[207]. Many songbirds demonstrate exquisite sequential memory, although it is species-specific[205]. Humans excel at learning arbitrary sequences of colours, tones or shapes[208]. A meta-analysis of 108 experiments across 14 species[209] found that humans required far fewer trials to reach a given level of performance on sequencing tasks than other species. For example, rats required thousands of trials to discriminate between two 3-item sequences above 50%; pigeons required 300 trials to reach 95% accuracy, and then showed almost no improvement for 2,000 trials; and humans reached nearly 100% accuracy in 10 trials. On a sequence discrimination task, humans took 30 trials to reach 90% accuracy whereas macaques took 400 trials[210].

Human adults might use language encoding as a tool to enhance their capacity on sequencing tasks with verbalizable materials[211]. However, non-verbal tasks such as the Corsi tapping task also show capacity differences between humans and non-human primates[211]. Baboons performed significantly worse than humans on this task, being only consistently above chance with three-item and four-item sequences, whereas humans were above chance up to six items. Humans also showed evidence of using a sequencing strategy involving encoding relative distances and proximities among the sequence items, which demands more resources initially than rote memory of locations but could feed back into better memory in the long run. This strategy difference suggests a more limited capacity in non-human animals to represent sequential relations compared with humans[66,162,212,213].

The amount of time an individual can hold something in working memory (duration) is independent of the number of items an individual can hold in working memory[214]. The duration of working memory varies between species. On a one-item delayed matching task with one distractor, memory duration is similar across non-human primates and, potentially, some birds, whereas insect memory is poorer[213]. The serial probe recognition task[215] tests memory duration per item position, without requiring animals to remember the sequence order. Pigeons, monkeys and humans all perform this task with high accuracy and show the signatures of primacy and recency in their memory performance[216]. However, pigeon memory decays approximately three times faster than monkey memory, and monkey memory decays three times faster than human memory.

Chimpanzees showed impressive memory capacity on a touch-screen task in which numerals between '1' and '9' are briefly flashed on a screen and subjects have to press the locations in a trained order[217].

Chimpanzees practised this task over a 1-year period, 4 times per week for 50 trials per session. After training, they were given a novel testing session in which the items were flashed at sub-second speeds and their performance was compared with that of humans. With six items, chimpanzees were less accurate than humans and achieved 65% accuracy, compared with 80% for humans during the two testing sessions, and with five items, chimpanzees' performance was generally within the range of human performance. However, non-human primates typically perform much worse than humans on memory tasks. For example, chimpanzees and macaques struggle to integrate information in the memory game (in which players flip over pairs of cards to find matching pairs), tracking items over fewer than 4 cards[218] compared with humans tracking items over 12 cards. Thus, whereas chimpanzees can perform similarly to humans on some memory tasks, typically humans excel in this area compared with other species[213].

Other domains requiring sequence memory are tool use and language. Non-human animals cannot produce complex, hierarchical sequences of behaviour in a range of domains from communication to tool use[11,219] but they can integrate a limited number of action–object relations into their tool-using routines. Most non-human animal tool use involves a single object impacting another object[11]. However, human tool use often requires hierarchical embedding: using one object to make another object and then impacting a third object (and deeper levels of subroutines)[11]. Hierarchical embedding is the same limitation observed in ape language wherein apes did not produce multi-word utterances and struggled to comprehend multi-item hierarchical syntax[166]. General differences in hierarchical reasoning for action and language might in this way distinguish humans from non-human primates[11,220]. This general constraint on non-human primate action and communication is a quantitative one —the number of levels of embedding in communication and action.

Differences in item capacity and temporal decay in sequence memory, as seen between humans and non-human primates, can substantially impact a species' ability to extract various types of rules from new information, even when there are no other differences in representation. Intricate rules, such as analogies, grammars and hierarchies, require the observation of patterns across multiple items that unfold over time. When one's capacity to faithfully represent multiple items over an extended duration is small, it is impossible to extract the kinds of complex patterns or rules that only emerge in longer sequences and sets. Consequently, memory capacity for sets and sequences is a critical bottleneck for representing relations and rules across domains, including social cognition, tool use and communication.

Qualitative impacts from quantitative change

The third prediction of the information capacity theory is that quantitative change in information capacity can underlie qualitative leaps in ability. In this section we discuss these impacts across human development and in machine computation.

Developmental change in representing relations

Theories of uniquely human cognitive processes are stymied by the lack of evidence of new neural functions or conceptual primitives that are both innate and unique to humans[221]. However, this state is not an issue for theories of uniquely human information capacity. Indeed, the same developmental primitives are present in humans as in other primate species: associative and statistical learning[216,222], ordinality and iteration[220,223], object representations[224], spatiotemporal intuitions[130], quantitative and numerical reasoning[225], categorization and generalization[226], gaze following[227] and speech segmentation[228].

It has been argued that an ape can do anything a human child can do up until around the age of 3 years[179]. At around 3 years of age, something changes in the human child that affords deeper abstract concepts and mental operations. According to our theory, this developmental change is due to an enhanced information processing capacity, an account that has a counterpart in constructivist theories of human development dating back to the 1960s. At that time, it was theorized that intelligence in children emerges by way of a generalized combinatorial system: developing general resources for mentally writing down combinatorial rules that are abstract and generalizable[229].

Expanding on these initial ideas, many researchers have described how conceptual change in childhood is fuelled by increases in general capacities[48,68,70,230–232]. For instance, the speeds at which many cognitive processes take place, such as mental addition, mental rotation, memory search and simple motor skills, follow a consistent and predictable exponential pattern of change throughout development[233]. This observation indicates that there is a general constraint on global cognitive processing and reflects properties of the information capacity of children.

Major leaps in human conceptual development require integrating four items into a logical rule or relation. There is a mathematical reason for this quantitative constraint on logical rule learning (Box 3). Complex patterns and rules only exist across four or more items organized into dimensions, subgroups and hierarchies (not across one, two or three items). Logical rules requiring four items are sometimes called quaternary relations[70]. Analogies[70], embedded and conjoint conditional rules[234], recursive syntax[235], conjunctive syntax with two subjects, a verb, and an object[167], centre-embedded hierarchies[162,236] and the successor function in counting[237] can only be extracted across sets or sequences of at least four entities. For example, an analogy such as 'dog is to wolf as cat is to cougar' requires representing dog, wolf, cat and cougar in an A:B::C:D rule. Furthermore, a centre-embedded hierarchy is only distinct from iteration when it has at least four items (for instance in an A^nB^n rule[162]).

The types of rules and patterns that children can explicitly represent at different ages vary quantitatively: unary relations at age 1 year, binary relations at age 2 years, ternary relations at age 5 years and quaternary relations at age 11 years[70]. Unary means the child is representing one relation, such as 'The colour of the sky is blue'. The number of relations scales with the depth of meaning from concrete to operational —unary relations compare object and feature similarity whereas quaternary relations compare operations such as the similarity of 'similarities' between groups. For example, in category similarity judgements, children's knowledge transitions from the unary object-based sameness (for instance, a dog is more similar to a wolf than a parrot) to quaternary relational-sameness (for instance, a dog is similar to a wolf in the same way a cat is similar to a cougar)[70]. We suggest that between the ages of 3 and 5 years children become capable of representing rules and relations across four items because they develop greater (functional) information processing capacity, enabling complex rules and operations.

Non-human primates typically lack the ability to represent quaternary relations. Even with years of language and symbol training, non-human primates struggle with the combinatorial functions of language learning[50,167,233]. Thus, they remain stuck at the level of a human child aged 2–3 years when it comes to generating and comprehending a sentence or relational phrase. Multiple studies have shown that even

with more than 4 years of experience, language-trained apes have an approximate mean utterance length of around 1.6 words, which is similar to 2-year-old children[167,238]. By contrast, by 3 years of age human children have mean utterance lengths of three to four words[239].

One longitudinal project directly compared language comprehension in a 2-year-old child and a 2-year-old bonobo who had similar language experiences. The child and bonobo were trained from 3 months of age to produce words and strings with a lexigram keyboard and to comprehend spoken language[167]. Over a testing period of 9 months, the child initially produced utterance lengths averaging 1.91 morphemes and rose to an average of 3.19, whereas the bonobo's utterances began at 1.15 morphemes and remained stuck at that level throughout the testing period. Language comprehension was tested by presenting an array of real objects and asking the subject via novel spoken sentences to carry out actions with those objects (for example, 'Put the ball on the pine needles'). Comprehension of language was far easier than production for the bonobo, who could carry out tasks from utterances that were longer than those he produced. This pattern mirrors child language development[167] and highlights the importance of memory constraints in interpreting language production. The bonobo also used word order to interpret the meaning of the sentences and could understand syntactic reversals (such as 'Put some water in the coke' versus 'Put some coke in the water'). The specific limitation of the bonobo compared with human children was with long utterances and conjunctive constructions that burden memory capacity, such as 'Give the peas and the sweet potatoes to Kelly'. These findings suggest that the key constraint on the bonobo's language learning was the capacity for multiple simultaneous representations.

Developmental change in domain-general capacity

Also consistent with the notion that quantitative changes yield conceptual change, developmental changes in domain-general capacities are related to children's conceptual changes[68]. For example, the emergence of theory of mind accompanies conceptual change in early childhood. Theory of mind requires integrating multiple relations between minds and behaviours of oneself and others[240]. Infants have knowledge about the goal-directed and intentional acts of people and themselves but struggle to integrate that information across multiple agents and events[159]. Tasks such as the false belief task that measure theory of mind ability show poor performance until around the age of 5 years, after which children typically succeed[241]. Children's individual differences on this task are highly correlated with their executive function abilities[234,242]. Furthermore, relational capacity explains 80% of age-related variance in the performance of children aged 3–5 years on false belief tasks[243]. That is, children's performance on tasks such as transitive inference, which require high relational capacity (but not theory of mind), predicts their performance on tasks that do require theory of mind. The capacity to integrate multiple relations develops gradually in children and once present, theory of mind can develop. Although the emergence of theory of mind might look similar to a qualitative conceptual change, it is underlain by quantitative changes in information processing capacity.

Over development, expansions in general information capacities increase the number of rules and relations that children can represent, compare or embed, which increases the types of rules that can be represented. Qualitative conceptual change requires semantic knowledge and experience but is fundamentally fuelled by changes in information processing capacity. Thus, the phylogenetic and ontogenetic causes of intelligence have a common coupling to information capacity.

Information capacity in computation

Capacity differences, even minor or continuous changes, can profoundly impact computational performance. For instance, Turing machines, a standard abstract model of computation, are formalized mathematically as possessing an infinite memory. However, early results in complexity theory showed that if the computer has a tight bound on memory available (as a function of its input size), then the class of computations it can execute is strictly limited. If the amount of memory available to process an input of size n is less than $\log(\log(n))$, then the computer is provably only capable of recognizing regular languages[244] —languages that can be processed with a finite number of memory states[245]. Such capacity-limited computers are therefore incapable of processing the kinds of hierarchies such as context-free languages that are considered to approximate human language, much less more complex context-sensitive features or transformations[246]. This result seems to establish a 'quantum jump' in memory requirements for regular versus non-regular languages: a small quantitative increase in memory can lead to qualitatively different computational ability.

These formal results provide a clear demarcation between regular and non-regular languages that is governed by the amount of available memory: algorithms or species that use comparatively little memory cannot be generating or recognizing anything other than the simplest kinds of string patterns. Although these results are more than 60 years old, they are strongly connected to contemporary cognitive science. Several studies have examined what kinds of string patterns animals are capable of processing[157,164,247–250]; other work has attempted to rigorously characterize the level of computational complexity in human languages[251–256]. The primary distinction examined in animal work is whether non-human species are capable of recognizing or generating strings from languages other than regular languages, but the results are mixed and difficult to interpret, in part, because the sets of strings studied can often be processed with other heuristics.

More general findings about the linkage between memory and computational ability can be found in the space hierarchy theorem[257], which proves that computers that are given more memory capacity (as a function of their input size) are capable of solving a larger number of problems. This conclusion can be contrasted with, for example, the fact that many modifications of Turing machines do not increase their computational capability. For example, altering a standard Turing machine to give it two tapes, or even two-dimensional tapes, does not fundamentally change the rules it can use or the problems it can solve. But allocating it more internal memory space, as a function of input size, does.

When a learner's memory is only a few items, it will not be possible for them to learn classes of patterns that only become apparent after several items. For example, neither the sequence AB nor ABA leads one to see the repetition pattern in ABAB, meaning that this repetition pattern would be inaccessible for learners with memories of fewer than four items. This idea can be formulated in the context of structured learning models, for which limited memory of the input data would prevent acquisition of patterns, including hierarchical structure. In this case, the best a learner could do is try to memorize the data. The behavioural consequence of this memory limitation would be limited rule-like generalizations in relational reasoning, imitation, tool use, language learning and any domain that requires representations of multiple actions, agents and entities.

Although this impact is debated, capacity constraints have also been argued to critically shape machine learning capability in modern neural networks. Machine learning performance depends on

parameters such as the amount of training data, the number of training steps and the number and arrangement of nodes in the network[258,259]. Some models only show above-floor performance on tasks involving human-like intelligence when they are able to use billions of parameters. Indeed, although some tasks scale smoothly with model size, at other times changes in capacity are profound and discontinuous as the parameters or capacity of the models is increased[260]. Such neural networks show systematic scaling patterns, often power laws, between performance and both the amount of data and the neural network size or capacity[261–263]. The highest performing deep learning models can memorize massive amounts of data, fitting even random labels[264,265], and it is likely that this ability for memorization is critical for extracting the higher-order regularities necessary for high performance in domains such as language usage and image classification. Indeed, modern high-capacity systems have been argued to show signs of general intelligence, succeeding on a diverse array of tasks such as mathematics, tool use, theory of mind and programming[266], although the ability of these systems to truly reason in a human-like way is contested[267].

Part of the success of modern neural networks has been driven by the realization that increasing computational power can outperform tailored (silver-bullet) representations[268]. For domains including speech recognition, chess and computer vision, experts once believed that tailored representations were the solution, but these efforts stalled and, arguably, even inhibited future progress[269]. These problems were eventually solved using simpler techniques with models that built in less knowledge of each specific domain, combined with higher computing power and more general capacity[269].

Together, these results generally show that the informational limits faced by a system are an important determinant of the range of computations it can perform. Moreover, the success of the best learning models does not seem to be due to carefully constructed silver-bullet representations but, rather, the discovery of how to make learning scale to large data sets and numbers of parameters, suggesting that general scaling of informational capacity might have been a route to the evolution of human-like intelligence.

Conclusion

We propose that global, genetic differences in learning and memory are sufficient to account for uniquely human reasoning across domains, as an alternative to theories that require qualitative, domain-specific evolutionary changes specifically in human cognition to explain uniquely human behaviours. This theory of human uniqueness makes three concrete predictions. First, the theory predicts that non-human primates will show some degree of success on tasks previously purported to draw on domain-specific specializations in humans. Second, the theory predicts continuous, quantifiable advantages for humans relative to non-human species on basic tasks, even tasks that only require representations of a few items, actions or features. Last, the theory predicts that subtle increases in information capacity yield qualitative changes in behaviour in humans, non-human animals and computational systems more broadly.

No existing theories have been able to adequately disentangle information capacity from domain-specific functions, nor explain the concrete implications these capacity limitations have on the ability to learn and represent knowledge. Theoreticians of domain-specific theories have been too eager to posit that specific mechanisms are unique to humans, leaving all the other differences in the information processing machinery of cognition unaddressed and confounding. We suspect that global information capacity tends to take a backseat in theories of human uniqueness because researchers underestimate its potential to yield qualitative changes in cognition between species. However, general capacity constraints have profound consequences and information capacity determines which species can learn the hierarchical, abstract patterns and generalizations and which species, simply, cannot. We showed how information capacity gradually expands over human development and how it relates to qualitative changes in human cognition. Expansions in information capacity enable rule-based, compressed representations of complex phenomena such as sequences, sets and relations that are abstract and generalizable (Box 3). The ability to think symbolically probably enriches this benefit of information capacity even further. Independently of the neural or cognitive instantiation, changes to capacity have a surprising qualitative effect on the abilities of computational systems: differences in degree yield differences in kind.

We detailed how basic differences in information capacity between humans and other animals can lead to qualitatively unique human cognition, both developmentally and evolutionarily, offering new research possibilities. New directions include testing how variations in information capacity relate to complex rule learning and concept formation across species, how the synergy between information capacity and language expands concept and rule learning during human development, and the role of sequence and set memory in rule extraction across different species and age groups. One key prediction is that the capacity for complex rule learning (whether social, physical or abstract) is inherently and mathematically linked to the capacity for representing multiple items simultaneously —a hypothesis that can be tested in animals, developing children and machines. A key experimental direction will be to test causal evidence for the role of information capacity on the tasks hypothesized to show human uniqueness. However, measuring information capacity directly poses a challenge (Box 4), due to the complexity of the target behaviours and the limited tasks for which information can be formally characterized at present. Information measures vary across inputs, abstractions and processes. For instance, channel capacity for high-level vision of objects might differ between species, even if the capacity for low-level vision is the same because the information processing demands of high-level versus low-level abstractions probably differ. Information processing demands also can vary across content types, such as between social interactions and object use. Simulations of information capacity under different processing conditions are critical for developing experimental predictions.

'How' human cognition is unique is probably intertwined with 'why' human cognition is unique. Several authors have proposed accounts of human evolution that provide a compelling answer to 'why' new cognitive abilities might have emerged specifically in humans. Perhaps primitive human environments contained unique survival pressures to cooperate and learn socially[31] or humans uniquely benefited from cumulative culture, which altered the human environment so dramatically that it prompted new genetic adaptations[12,270]. The problem with prior 'why' accounts such as these is that they do not explain 'how', cognitively, human minds and brains changed[271]. In light of the behavioural, neurobiological, computational and developmental evidence we reviewed, we propose that unique expansions in global, generic information capacity are the most plausible genetically based cognitive adaptation to ratcheting environmental pressures on learning, memory, attention, semantics and logical rule use arising from ever more demanding human culture.

Box 4

Measuring information capacity

There are several methods that have been developed to measure information processing, and each targets information specific to a given behavioural task or setting.

Speed and bit rate

One approach to measuring information capacity is to use tasks in which participants are asked to respond as quickly and accurately as possible to a stimulus. Hick's law quantifies the 'rate of information gain' in a simple choice paradigm[310]; memory tasks such as digit span and working memory measure item, sequence and channel capacity for information[213,308]. Posner cueing tasks, flanker tasks and Stroop tasks also have been used to measure bit rates in humans with simple rules and stimuli[311]. One study showed that the speeds at which numerous diverse cognitive processes occur, including mental addition, mental rotation, memory search and simple motor skills, exhibit a consistent and predictable exponential pattern of change over development[239] — a global cognitive processing constraint reflecting children's information capacity. Additionally, some work has provided estimates of information capacity in humans for long-term memory and linguistic storage[180,202,205], often by quantifying what coding capacity would be required to achieve the observed performance.

Predictability

Machine learning models are often evaluated by computing the log probability that they assign to the observed data, which is an information measure (often surprisal or cross-entropy). Following the same logic, one could measure the effective predictability of, for example, sequential stimuli for any species and task, and use that to quantify information processing ability. For instance, in one study people were asked to predict upcoming letters in text and their accuracy was used to deduce how much information they had about upcoming linguistic material[312]. In a context in which an individual predicts upcoming sequential material that follows a novel pattern — for example, simple formal language rules — accuracy could be used to estimate how much information about the pattern is internalized. For example, from an individual's accuracy or pattern of errors in predicting the next character in the sequence 'abbacabbacabbac ...', one can compute how much they have learned about the sequence (see Reber[309] for people's learning of similar string patterns and Saffran et al.[313] for work on early language learning). Empirically measuring predictive ability in this setting is potentially powerful because above-chance accuracy means that some information must be present, which can be examined as a function of developmental age, amount of training exposure or species.

Learning rate and lapse rate

The learning rate and the lapse rate are another pair of measures for estimating information capacity[314]. The learning rate quantifies change in accuracy per unit time, and the lapse rate is the asymptote of the learning curve. Both of these metrics vary across species and can be used to query global information processing across tasks. Lapse rates are especially useful for measuring general motivation during a given task, which is a common confound for comparing cognition between populations. Similarly, species might make different speed–accuracy trade-offs in any given task[315], with different strategic choices potentially confounding conclusions about overall ability.

Recoding

Estimating and comparing information capacity between groups is difficult when recoding occurs[316]. Recoding is the process by which information is mentally compressed during task performance. Humans are very good at recoding information using chunking, rules, heuristics or verbalization, thereby making space for more information. Non-human primates also have some ability to recode information[171,217]. The ways that humans and non-human primates recode and compress information are only beginning to be understood, but it is critical to understand these phenomena in order to measure information capacity. For example, symbolic recoding could facilitate human learning during passive or social learning tasks in ways that exceed the capabilities of non-human primates.

The uniquely human adaptation for information processing capacity provides an opportunity to represent multiple complex behavioural alternatives, to enact flexibility and innovation[272] and to learn quickly, and yields a substrate for 'cognitive gadgets', wherein humans acquire unique capacities by building on cultural innovations[13]. One cognitive gadget, writing, provides important clues to the pressures humans faced for information capacity. Over cultural evolution, the human species was so pressured for increased information capacity that they invented writing, a revolutionary leap forward in the development of our species that enables information capacity to be externalized[273], frees up internal processing and affords the development of more complex concepts. In other words, writing enabled humans to think more abstractly and logically by increasing information capacity. Today, humans have gone to even greater lengths: the Internet, computers and smartphones are testaments to the substantial pressure humans currently face—and probably faced in the past—to increase information capacity.

Darwin claimed that humans differed from other primates in degree rather than kind, which has been dismissed by some as a mistaken claim[52]. But if there is one lesson from formal computer science, it is that differences in degree yield differences in kind. Evolution doubled the information processing capacity of humans, which set in motion a cycle of advances between logic innovation and developmental acquisition that snowballed over millions of years of human cultural evolution. Other species could never learn human-like rules and structures, internalize human-like logic and relations or exhibit the same complex behaviour as humans when their cognitive systems face severe information constraints. Unique information capacity magnifies human cognition, qualitatively alters its representations and processes, and is essential for understanding the evolution and development of human intelligence.

Published online: xx xx xxxx

References

1. Burkart, J. M., Schubiger, M. N. & van Schaik, C. P. The evolution of general intelligence. *Behav. Brain Sci*. **40**, e195 (2017).
2. Cosmides, L. & Tooby, J. in *The New Cognitive Neurosciences* 2nd edn (Gazzaniga, M. S.) 1259–1270 (MIT Press, 2000).
3. Dehaene, S., Al Roumi, F., Lakretz, Y., Planton, S. & Sablé-Meyer Symbols and mental programs: a hypothesis about human singularity. *Trends Cogn. Sci*. **26**, 751–766 (2022).
4. Pinker, S. *The Blank Slate: The Modern Denial of Human Nature* (Penguin, 2003).
5. Pinker, S. in *Understanding Moral Sentiments* (eds Putnam, H. et al.) 59–80 (Routledge, 2017).
6. Pinker, S. & Jackendoff, R. The faculty of language: what's special about it? *Cognition* **95**, 201–236 (2005).
7. Povinelli, D. J., Bering, J. M. & Giambrone, S. Toward a science of other minds: escaping the argument by analogy. *Cogn. Sci*. **24**, 509–541 (2000).
8. Wellman, H. M. Theory of mind: the state of the art. *Eur. J. Dev. Psychol*. **15**, 728–755 (2018).
9. Deacon, T. W. *The Symbolic Species: The Co-Evolution of Language and the Brain* No. 202 (W. W. Norton, 1998).
10. Deaner, R. O., Van Schaik, C. P. & Johnson, V. Do some taxa have better domain-general cognition than others? A meta-analysis of nonhuman primate studies. *Evol. Psychol*. **4**, https://doi.org/10.1177/147470490600400114 (2006).
11. Greenfield, P. M. Language, tools and brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behav. Brain Sci*. **14**, 531–551 (1991).
12. Henrich, J. *The Secret of Our Success* (Princeton Univ. Press, 2015).
13. Heyes, C. *Cognitive Gadgets: The Cultural Evolution of Thinking* (Harvard Univ. Press, 2018).
14. Passingham, R. E. & Wise, S. P. *The Neurobiology of the Prefrontal Cortex: Anatomy, Evolution, and the origin of Insight* (Oxford Univ. Press, 2012).
15. Premack, D. Is language the key to human intelligence? *Science* **303**, 318–320 (2004).
16. Premack, D. Why humans are unique: three theories. *Perspect. Psychol. Sci*. **5**, 22–32 (2010).
17. Reader, S. M., Hager, Y. & Laland, K. N. The evolution of primate general and cultural intelligence. *Philos. Trans. R. Soc. Lond. B Biol. Sci*. **366**, 1017–1027 (2011).
18. Laland, K. & Seed, A. Understanding human cognitive uniqueness. *Annu. Rev. Psychol*. **72**, 689–716 (2021).
19. Tomasello, M. The ultra-social animal. *Eur. J. Soc. Psychol*. **44**, 187–194 (2014).
20. Laland, K. N. *Darwin's Unfinished Symphony: How Culture Made the Human Mind* (Princeton Univ. Press, 2017).
21. Richerson, P. J. & Boyd, R. *Not by Genes Alone: How Culture Transformed Human Evolution* (Univ. of Chicago Press, 2018).
22. Sterelny, K. *The Evolved Apprentice* (MIT Press, 2012).
23. Turner, C. R. & Walmsley, L. D. Preparedness in cultural learning. *Synthese* **199**, 81–100 (2021).
24. Laland, K. N. & Brown, G. R. *Sense and Nonsense: Evolutionary Perspectives on Human Behaviour* (Oxford Univ. Press, 2011).
25. Laland, K. N. & Reader, S. M. in *Innovation in Cultural Systems: Contributions from Evolutionary Anthropology* (eds O'Brien, M. J. & Shennan, S. J.) 37–51 (2010).
26. Lefebvre, L., Whittle, P., Lascaris, E. & Finkelstein, A. Feeding innovations and forebrain size in birds. *Anim. Behav*. **53**, 549–560 (1997).
27. MacLean, E. L. Unraveling the evolution of uniquely human cognition. *Proc. Natl Acad. Sci. USA* **113**, 6348–6354 (2016).
28. Moll, H. & Tomasello, M. Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philos Trans R. Soc. Lond. B Biol. Sci*. **362**, 639–648 (2007).
29. Reader, S. M. & Laland, K. N. Social intelligence, innovation, and enhanced brain size in primates. *Proc. Natl Acad. Sci. USA* **99**, 4436–4441 (2002).
30. Sol, D., Bacher, S., Reader, S. M. & Lefebvre, L. Brain size predicts the success of mammal species introduced into novel environments. *Am. Nat*. **172**, S63–S71 (2008).
31. Tomasello, M. *A Natural History of Human Thinking* (Harvard Univ. Press, 2018).
32. Tomasello, M., Kruger, A. C. & Ratner, H. H. Cultural learning. *Behav. Brain Sci*. **16**, 495–511 (1993).
33. Van Schaik, C. P. & Burkart, J. M. Social learning and evolution: the cultural intelligence hypothesis. *Philos. Trans. R. Soc. Lond. B, Biol. Sci*. **366**, 1008–1016 (2011).
34. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J*. **27**, 379–423 (1948).
35. Stone, J. V. *Information Theory: A Tutorial Introduction* (Sebtel, 2015).
36. Wilson, E. O. *Sociobiology: The New Synthesis* (Harvard Univ. Press, 2000).
37. Lumsden, C. J. & Wilson, E. O. Précis of genes, mind, and culture. *Behav. Brain Sci*. **5**, 1–7 (1982).
38. Barkow, J. H., Cosmides, L. & Tooby, J. (eds) *The Adapted Mind: Evolutionary Psychology and the Generation of Culture* (Oxford Univ. Press, 1995).
39. Pinker, S. & Bloom, P. Natural language and natural selection. *Behav. Brain Sci*. **13**, 707–727 (1990).
40. Chomsky, N. Persistent topics in linguistic theory. *Diogenes* **13**, 13–20 (1965).
41. Garcia, J., Kimeldorf, D. & Koelling, R. Conditioned aversion to saccharin resulting from exposure to gamma radiation. *Science* **122**, 157–158 (1955).
42. Ermer, E., Cosmides, L. & Tooby, J. in *The Evolution of Mind: Fundamental Questions and Controversies* (eds Gangestad, S. W. & Simpson, J. A.) 153–160 (2007).
43. Sperber, D. in *Language, Brain and Cognitive Development: Essays in Honor of Jacques Mehler* Vol. 7 (ed. Dupoux, E.) 47–57 (MIT Press, 2001).
44. Scholl, B. J. & Leslie, A. M. Modularity, development and 'theory of mind'. *Mind Lang*. **14**, 131–153 (1999).
45. Leslie, A. M., Friedman, O. & German, T. P. Core mechanisms in 'theory of mind'. *Trends Cogn. Sci*. **8**, 528–533 (2004).
46. Meltzoff, A. N. & Williamson, R. A. in *Imitation: Social, Cognitive, and Theoretical Perspectives* (ed. Zelazo, P. D.) 651–682 (Oxford Academic, 2013).
47. Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci*. **28**, 675–735 (2005).
48. Rule, J. S., Tenenbaum, J. B. & Piantadosi, S. T. The child as hacker. *Trends Cogn. Sci*. **24**, 900–915 (2020).
49. Sablé-Meyer et al. Sensitivity to geometric shape regularity in humans and baboons: a putative signature of human singularity. *Proc. Natl Acad. Sci. USA* **118**, e2023123118 (2021).
50. Premack, D. The codes of man and beasts. *Behav. Brain Sci*. **6**, 125–137 (1983).
51. Hauser, M. D., Chomsky, N. & Fitch, W. T. The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579 (2002).
52. Penn, D. C., Holyoak, K. J. & Povinelli, D. J. Darwin's mistake: explaining the discontinuity between human and nonhuman minds. *Behav. Brain Sci*. **31**, 109–130 (2008).
53. Chomsky, N. *The Minimalist Program* (MIT Press, 1995).
54. Fisher, S. E. & Marcus, G. F. The eloquent ape: genes, brains and the evolution of language. *Nat. Rev. Genet*. **7**, 9–20 (2006).
55. Berwick, R. C. & Chomsky, N. *Why Only Us: Language and Evolution* (MIT Press, 2016).
56. Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B. & Tomasello, M. Humans have evolved specialized skills of social cognition: the cultural intelligence hypothesis. *Science* **317**, 1360–1366 (2007).
57. Hodos, W. in *The Neurosciences: Second Study Program* (ed. Schmidt, F. O.) 26–39 (Rockefeller Univ. Press, 1970).
58. Jerison, H. J. Animal intelligence as encephalization. *Philos. Trans. R. Soc. B* **308**, 21–35 (1985).
59. Passingham, R. E. Visual discrimination learning after selective prefrontal ablations in monkeys (*Macaca mulatta*). *Neuropsychologia* **10**, 27-39 (1972).
60. Passingham, R. *Understanding the Prefrontal Cortex: Selective Advantage, Connectivity, and Neural Operations* (Oxford Univ. Press, 2021).
61. Bolhuis, J. J., Brown, G. R., Richardson, R. C. & Laland, K. N. Darwin in mind: new opportunities for evolutionary psychology. *PLoS Biol*. **9**, e1001109 (2011).
62. Christiansen, M. H. & Chater, N. *Creating Language: Integrating Evolution, Acquisition, and Processing* (MIT Press, 2016).
63. Sol, D. Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biol. Lett*. **5**, 130–133 (2009).
64. Carey, S., Zaitchik, D. & Bascandziev, I. Theories of development: in dialog with Jean Piaget. *Dev. Rev*. **38**, 36–54 (2015).
65. Christie, S. & Gentner, D. Language helps children succeed on a classic analogy task. *Cogn. Sci*. **38**, 383–397 (2014).
66. Halford, G. S., Wilson, W. H. & Phillips, S. Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behav. Brain Sci*. **21**, 803–831 (1998).
67. Stone, J. V. *Principles of Neural Information Theory* (Sebtel, 2018).
68. Sterling, P. & Laughlin, S. *Principles of Neural Design* (MIT Press, 2015).
69. Humphrey, N. K. *The Social Function of Intellect* (Cambridge Univ. Press, 1976).
70. Byrne, R. W. & Whiten, A. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans* (Oxford Univ. Press, 1988).
71. Harris, P. L. in *Handbook of Child Psychology: Cognition, Perception, and Language* (eds D. Kuhn, R. S. Siegler, W. Damon & R. M. Lerner) 811–858 (Wiley, 2006).
72. Boyd, J. H. Have we found the Holy Grail? Theory of mind as a unifying construct. *J. Relig. Health* **47**, 366–385 (2008).
73. Byrne, R. W. & Bates, L. A. Sociality, evolution and cognition. *Curr. Biol*. **17**, R714–R723 (2007).
74. Horschler, D. J., MacLean, E. L. & Santos, L. R. Do non-human primates really represent others' beliefs? *Trends Cogn. Sci*. **24**, 594–605 (2020).
75. Lyons, D. E. & Santos, L. R. Ecology, domain specificity, and the origins of theory of mind: is competition the catalyst? *Philos. Compass* **1**, 481–492 (2006).
76. Lewis, L. & Krupenye, C. Theory of mind in nonhuman primates. *Behav. Brain Rev*. **21**, 101–114 (2021).
77. Heyes, C. Animal mindreading: what's the problem? *Psychon. Bull. Rev*. **22**, 313–327 (2015).
78. Hare, B., Call, J. & Tomasello, M. Do chimpanzees know what conspecifics know? *Anim. Behav*. **61**, 139–151 (2001).
79. Hare, B., Call, J. & Tomasello, M. Chimpanzees deceive a human competitor by hiding. *Cognition* **101**, 495–514 (2006).
80. Kano, F. & Call, J. Cross-species variation in gaze following and conspecific preference among great apes, human infants and adults. *Anim. Behav*. **91**, 137–150 (2014).
81. MacLean, E. L. & Hare, B. Bonobos and chimpanzees infer the target of another's attention. *Anim. Behav*. **83**, 345–353 (2012).
82. Horner, V., Whiten, A., Flynn, E. & De Waal, F. B. Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proc. Natl Acad. Sci. USA* **103**, 13878–13883 (2006).
83. van de Waal, E., Claidière, N. & Whiten, A. Wild vervet monkeys copy alternative methods for opening an artificial fruit. *Anim. Cogn*. **18**, 617–627 (2015).

84. Tomasello, M., Call, J. & Hare, B. Five primate species follow the visual gaze of conspecifics. *Anim. Behav.* **55**, 1063–1069 (1998).
85. Okamoto-Barth, S. & Tomonaga, M. in *Cognitive Development in Chimpanzees* 155–171 (Springer Tokyo, 2006).
86. Carpenter, M. & Call, J. in *Agency and Joint Attention* (ed. Metcalfe, J. & Terrace, H. S.) 49–61 (2013).
87. Musgrave, S. et al. Teaching varies with task complexity in wild chimpanzees. *Proc. Natl Acad. Sci. USA* **117**, 969–976 (2020).
88. Whiten, A. Social learning: peering deeper into ape culture. *Curr. Biol.* **29**, R845–R847 (2019).
89. Drayton, L. A. & Santos, L. R. Capuchins' (*Cebus apella*) sensitivity to others' goal-directed actions in a helping context. *Anim. Cogn.* **17**, 689–700 (2014).
90. Warneken, F., Chen, F. & Tomasello, M. Cooperative activities in young children and chimpanzees. *Child Dev.* **77**, 640–663 (2006).
91. Melis, A. P., Hare, B. & Tomasello, M. Engineering cooperation in chimpanzees: tolerance constraints on cooperation. *Anim. Behav.* **72**, 275–286 (2006).
92. Bard, K. A., Bakeman, R., Boysen, S. T. & Leavens, D. A. Emotional engagements predict and enhance social cognition in young chimpanzees. *Dev. Sci.* **17**, 682–696 (2014).
93. Gergely, G. & Csibra, G. Teleological reasoning in infancy: the naıve theory of rational action. *Trends Cogn. Sci.* **7**, 287–292 (2003).
94. Penn, D. C. & Povinelli, D. On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 731–744 (2007).
95. Hamann, K., Warneken, F., Greenberg, J. R. & Tomasello, M. Collaboration encourages equal sharing in children but not in chimpanzees. *Nature* **476**, 328–331 (2011).
96. Heyes, C. False belief in infancy: a fresh look. *Dev. Sci.* **17**, 647–659 (2014).
97. Phillips, J. et al. Knowledge before belief. *Behav. Brain Sci.* **44**, e140 (2021).
98. Rakoczy, H. Foundations of theory of mind and its development in early childhood. *Nat. Rev. Psychol.* **1**, 223–235 (2022).
99. Heyes, C., Chater, N. & Dwyer, D. M. Sinking in: the peripheral Baldwinisation of human cognition. *Trends Cogn. Sci.* **24**, 884–899 (2020).
100. Halford, G. S., Wilson, W. H. & Phillips, S. Relational knowledge: the foundation of higher cognition. *Trends Cogn. Sci.* **14**, 497–505 (2010).
101. Penn, D. C., Holyoak, K. J., Povinelli, D. J. & Vaesen, K. So, are we the massively lucky species? *Behav. Brain Sci.* **35**, 236 (2012).
102. Thompson, R. K. & Oden, D. L. Categorical perception and conceptual judgments by nonhuman primates: the paleological monkey and the analogical ape. *Cogn. Sci.* **24**, 363–396 (2000).
103. Wasserman, E. A., Castro, L. & Fagot, J. in *APA Handbook of Comparative Psychology* (eds. Call, J. et al.) 359–384 (American Psychological Association, 2017).
104. Gillan, D. J., Premack, D. & Woodruff, G. Reasoning in the chimpanzee: I. Analogical reasoning. *J. Exp. Psychol. Anim. B* **7**, 1 (1981).
105. Kotovsky, L. & Gentner, D. Comparison and categorization in the development of relational similarity. *Child Dev.* **67**, 2797–2822 (1996).
106. Rattermann, M. J. & Gentner, D. More evidence for a relational shift in the development of analogy: children's performance on a causal-mapping task. *Cogn. Dev.* **13**, 453–478 (1998).
107. Thompson, R. K., Oden, D. L. & Boysen, S. T. Language-naive chimpanzees (*Pan troglodytes*) judge relations between relations in a conceptual matching-to-sample task. *J. Exp. Psychol. Anim. B* **23**, 31 (1997).
108. Oden, D. L., Thompson, R. K. & Premack, D. in *The Analogical Mind: Perspectives from Cognitive Science* (eds Gentner, D. et al.) 471–491 (2001).
109. Flemming, T. M., Beran, M. J. & Washburn, D. A. Disconnect in concept learning by rhesus monkeys (*Macaca mulatta*): judgment of relations and relations-between-relations. *J. Exp. Psychol. Anim. B* **33**, 55 (2007).
110. Christie, S., Gentner, D., Call, J. & Haun, D. B. M. Sensitivity to relational similarity and object similarity in apes and children. *Curr. Biol.* **26**, 531–535 (2016).
111. Kennedy, E. H. & Fragaszy, D. M. Analogical reasoning in a capuchin monkey (*Cebus apella*). *J. Comp. Psychol.* **122**, 167 (2008).
112. Smirnova, A., Zorina, Z., Obozova, T. & Wasserman, E. Crows spontaneously exhibit analogical reasoning. *Curr. Biol.* **25**, 256–260 (2015).
113. Emery, N. J. & Clayton, N. S. Tool use and physical cognition in birds and mammals. *Curr. Opin. Neurobiol.* **19**, 27–33 (2009).
114. Fragaszy, D., Izar, P., Visalberghi, E., Ottoni, E. B. & Oliveira, M. Wild capuchin monkeys use anvils and stone pounding tools. *Am. J. Primatol.* **64**, 359–366 (2004).
115. Visalberghi, E., Sabbatini, G., Taylor, A. H. & Hunt, G. R. in *APA Handbook of Comparative Psychology: Perception, Learning, and Cognition* (Call, J., Burghardt, G. M., Pepperberg, I. M., Snowdon, C. T. & Zentall, T.) 673–701 (American Psychological Association, 2017).
116. Fujita, K., Kuroshima, H. & Asai, S. How do tufted capuchin monkeys (*Cebus apella*) understand causality involved in tool use? *J. Exp. Psychol. Anim. B* **29**, 233 (2003).
117. Beck, S. R., Apperly, I. A., Chappell, J., Guthrie, C. & Cutting, N. Making tools isn't child's play. *Cognition* **119**, 301–306 (2011).
118. Kabadayi, C. & Osvath, M. Ravens parallel great apes in flexible planning for tool-use and bartering. *Science* **357**, 202–204 (2017).
119. Gallot, Q. & Gruber, T. Spontaneous use and modification of a feather as a tool in a captive common raven. *Ethology* **125**, 755–758 (2019).
120. Ruiz, A. M. & Santos, L. R. in *Tool Use in Animals* (eds Sanz, C. M., Call, J. & Boesch, C.) 119–133 (Cambridge Univ. Press, 2013).
121. Povinelli, D. J. *Folk Physics for Apes* (Oxford Univ. Press, 2000).
122. Sanz, C. M. & Morgan, D. B. Chimpanzee tool technology in the Goualougo Triangle, Republic of Congo. *J. Hum. Evol.* **52**, 420–433 (2007).
123. Visalberghi, E. & Limongelli, L. Lack of comprehension of cause–effect relations in tool-using capuchin monkeys (*Cebus apella*). *J. Comp. Psychol.* **108**, 15–22 (1994).
124. Want, S. C. & Harris, P. L. Learning from other people's mistakes: causal understanding in learning to use a tool. *Child Dev.* **72**, 431–443 (2001).
125. Seed, A. M., Tebbich, S., Emery, N. J. & Clayton, N. S. Investigating physical cognition in rooks, *Corvus frugilegus*. *Curr. Biol.* **16**, 697–701 (2006).
126. Seed, A., Hanus, D. & Call, J. in *Tool Use and Causal Cognition* (eds McCormack, T. et al.) 89–110 (Oxford Scholarship Online, 2011).
127. Corballis, M. C. *The Recursive Mind* (Princeton Univ. Press, 2011).
128. Karlsson, F. Constraints on multiple center-embedding of clauses. *J. Linguist.* **43**, 365–392 (2007).
129. Gibson, E. & Thomas, J. Memory limitations and structural forgetting: the perception of complex ungrammatical sentences as grammatical. *Lang. Cogn. Proc.* **14**, 225 (1999).
130. Everett, D. Cultural constraints on grammar and cognition in Pirahã: another look at the design features of human language. *Curr. Anthropol.* **46**, 621–646 (2005).
131. Futrell, R., Stearns, L., Everett, D. L., Piantadosi, S. T. & Gibson, E. A corpus investigation of syntactic embedding in Pirahã. *PLoS ONE* **11**, e0145289 (2016).
132. Pullum, G. K. in *Current Controversies in Philosophy of Cognitive Science* (eds. Lerner, A. J. et al.) 29–45 (Routledge, 2020).
133. Frank, S. L., Bod, R. & Christiansen, M. H. How hierarchical is language use? *Proc. Biol. Sci.* **279**, 4522–4531 (2012).
134. Frank, S. L. & Bod, R. Insensitivity of the human sentence-processing system to hierarchical structure. *Psychol. Sci.* **22**, 829–834 (2011).
135. Nevins, A., Pesetsky, D. & Rodrigues, C. Evidence and argumentation: a reply to Everett (2009). *Language* **85**, 671–681 (2009).
136. Nevins, A., Pesetsky, D. & Rodrigues, C. Pirahã exceptionality: a reassessment. *Language* **85**, 355–404 (2009).
137. Everett, D. L. Cultural constraints on grammar in Pirahã: a reply to Nevins, Pesetsky, and Rodrigues. *semantics.uchicago.edu*, https://semantics.uchicago.edu/kennedy/classes/s07/myths/everett07.pdf (2007).
138. Everett, D. L. The shrinking Chomskyan corner: a final reply to Nevins, Pesetsky, and Rodrigues. *LingBuzz eprints* 994 (2013).
139. Bolhuis, J. J., Tattersall, I., Chomsky, N. & Berwick, R. C. How could language have evolved? *PLoS Biol.* **12**, e1001934 (2014).
140. Gibson, E. et al. How efficiency shapes human language. *Trends Cogn. Sci.* **23**, 389–407 (2019).
141. Martins, P. T. & Boeckx, C. Language evolution and complexity considerations: the no half-merge fallacy. *PLoS Biol.* **17**, e3000389 (2019).
142. De Boer, B., Thompson, B., Ravignani, A. & Boeckx, C. Evolutionary dynamics do not motivate a single-mutant theory of human language. *Sci. Rep.* **10**, 1–9 (2020).
143. Hockett, C. F. The origin of speech. *Sci. Am.* **203**, 88–97 (1960).
144. Tomasello, M. *Origins of Human Communication* (MIT Press, 2010).
145. Piantadosi, S. T., Tily, H. & Gibson, E. The communicative function of ambiguity in language. *Cognition* **122**, 280–291 (2012).
146. Kemp, C., Xu, Y. & Regier, T. Semantic typology and efficient communication. *Ann. Rev. Linguist.* **4**, 109–128 (2018).
147. Kemp, C. & Regier, T. Kinship categories across languages reflect general communicative principles. *Science* **336**, 1049–1054 (2012).
148. Mollica, F. et al. The forms and meanings of grammatical markers support efficient communication. *Proc. Natl Acad. Sci. USA* **118**, e2025993118 (2021).
149. Futrell, R. & Hahn, M. Information theory as a bridge between language function and language form. *Front. Commun.* **7**, https://doi.org/10.3389/fcomm.2022.657725 (2022).
150. Dąbrowska, E. What exactly is universal grammar, and has anyone seen it? *Front Psychol* **6**, 852 (2015).
151. Scholz, B. C. & Pullum, G. K. in *Contemporary Debates in Cognitive Science* (ed. Stainton, R.) 59–80 (2006).
152. Pullum, G. K. Theorizing about the syntax of human language: a radical alternative to generative formalisms. *Cadernos de Linguística* **1**, 1–33 (2020).
153. Tomasello, M. Language is not an instinct. *Cogn. Dev.* **10**, 131–156 (1995).
154. Evans, N. & Levinson, S. C. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–448 (2009).
155. Fitch, W. T. Toward a computational framework for cognitive biology: unifying approaches from cognitive neuroscience and comparative cognition. *Phys. Life Rev.* **11**, 329–364 (2014).
156. Conway, C. M. & Christiansen, M. H. Sequential learning in non-human primates. *Trends Cogn. Sci.* **5**, 539–546 (2001).
157. Ferrigno, S., Cheyette, S. J., Piantadosi, S. T. & Cantlon, J. F. Recursive sequence generation in monkeys, children, US adults, and native Amazonians. *Sci. Adv.* **6**, eaaz1002 (2020).
158. Liao, D. A., Brecht, K. F., Johnston, M. & Nieder, A. Recursive sequence generation in crows. *Sci. Adv.* **8**, eabq3356 (2022).
159. Corballis, M. C. Language evolution: a changing perspective. *Trends Cogn. Sci.* **21**, 229–236 (2017).
160. Kaminski, J., Call, J. & Fischer, J. Word learning in a domestic dog: evidence for "fast mapping". *Science* **304**, 1682–1683 (2004).

161. Orlov, T., Yakovlev, V., Hochstein, S. & Zohary, E. Macaque monkeys categorize images by their ordinal number. *Nature* **404**, 77–80 (2000).
162. Savage-Rumbaugh, E. S., Shanker, S. & Taylor, T. J. *Apes, Language, and the Human Mind* (Oxford Univ. Press, 1998).
163. Savage-Rumbaugh, E. S. et al. Language comprehension in ape and child. *Monogr. Soc. Res. Child.* **58**, 1-222i-252 (1993).
164. Terrace, H. S., Son, L. K. & Brannon, E. M. Serial expertise of rhesus macaques. *Psychol. Sci.* **14**, 66–73 (2003).
165. Chemla, E., Dautriche, I., Buccola, B. & Fagot, J. Constraints on the lexicons of human languages have cognitive roots present in baboons (*Papio papio*). *Proc. Natl Acad. Sci. USA* **116**, 14926–14930 (2019).
166. Dautriche, I., Buccola, B., Berthet, M., Fagot, J. & Chemla, E. Evidence for compositionality in baboons (*Papio papio*) through the test case of negation. *Sci. Rep.* **12**, 1–12 (2022).
167. Terrace, H. S. *Why Chimpanzees Can't Learn Language and only Humans Can* (Columbia Univ. Press, 2019).
168. Newport, E. L. Maturational constraints on language learning. *Cogn. Sci.* **14**, 11–28 (1990).
169. Hurford, J. R. Human uniqueness, learned symbols and recursive thought. *Eur. Rev.* **12**, 551–565 (2004).
170. Culicover, P. W. & Jackendoff, R. *Simpler Syntax* (Oxford Univ. Press, 2005).
171. Goldberg, A. E. *Construction Grammar: A Construction Grammar Approach to Argument Structure* (Univ. of Chicago Press, 1995).
172. Mollica, F. & Piantadosi, S. T. Humans store about 1.5 megabytes of information during language acquisition. *R. Soc. Open Sci.* **6**, 181393 (2019).
173. Spelke, E. S. in *Language in Mind: Advances in the Study of Language and Thought* (eds. Gentner, D. & Goldin-Meadow, S.) 277–311 (MIT Press, 2003).
174. Premack, D. Human and animal cognition: continuity and discontinuity. *Proc. Natl Acad. Sci. USA* **104**, 13861–13867 (2007).
175. Collins, A. M. & Quillian, M. R. Retrieval time from semantic memory. *J. Verb. Learn Verb. Behav.* **8**, 240–247 (1969).
176. Isbilen, E. S. & Christiansen, M. H. Chunk-based memory constraints on the cultural evolution of language. *Top Cogn. Sci.* **12**, 713–726 (2020).
177. Levinson, S. C. *Space in Language and Cognition: Explorations in Cognitive Diversity* Vol. 5 (Cambridge Univ. Press, 2003).
178. Bitterman, M. E. The comparative analysis of learning: are the laws of learning the same in all animals? *Science* **188**, 699–709 (1975).
179. Bryer, M. A. et al. The evolution of quantitative sensitivity. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **377**, 20200529 (2022).
180. Harlow, H. F. The formation of learning sets. *Psychol. Rev.* **56**, 51–65 (1949).
181. Rumbaugh, D. M. & Pate, J. L. in *Behavioral Evolution and Integrative Levels: The T. C. Schneirla Conferences Series* (eds Greenboch, G. & Tobach, E.) 221–241 (Psychology Press, 2014).
182. Byrne, R. W. Primate cognition: comparing problems and skills. *Am. J. Primatol.* **37**, 127–141 (1995).
183. MacLean, E. L., Hare, B., Nunn, C. L., Addessi, E., Amici, F., Anderson, R. C., ... & Zhao, Y. (2014). The evolution of self-control. Proceedings of the National Academy of Sciences, 111(20), E2140-E2148.
184. Shettleworth, S. J. Modularity, comparative cognition and human uniqueness. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **367**, 2794–2802 (2012).
185. Warren, J. M. Reversal learning and the formation of learning sets by cats and rhesus monkeys. *J. Comp. Physiol. Pscyh.* **61**, 421 (1966).
186. Strong, P. N. & Hedges, M. Comparative studies in simple oddity learning: I. Cats, raccoons, monkeys, and chimpanzees. *Psychon. Sci.* **5**, 13–14 (1966).
187. Fobes, J. L. & King, J. E. in *Primate Behavior* (eds. Fobes, J. L. & King, J. E.) 327–360 (Academic Press, 1982).
188. Strong, P. N. Comparative studies in oddity learning: III. Apparatus transfer in chimpanzees and children. *Psychon. Sci.* **7**, 43–43 (1967).
189. Overman, W., Bachevalier, J., Miller, M. & Moore, K. Children's performance on "animal tests" of oddity: implications for cognitive processes required for tests of oddity and delayed nonmatch to sample. *J. Exp. Child Psychol.* **62**, 223–242 (1996).
190. Shultz, S. & Dunbar, R. I. M. Species differences in executive function correlate with hippocampus volume and neocortex ratio across nonhuman primates. *J Comp. Psychol.* **124**, 252 (2010).
191. Miles, R. C. Discrimination-learning sets. *Behav. Nonhuman Primates* **1**, 51–95 (1965).
192. Pope, S. M. et al. Optional-switch cognitive flexibility in primates: chimpanzees' (*Pan troglodytes*) intermediate susceptibility to cognitive set. *J. Comp. Psychol.* **134**, 98 (2020).
193. Roberts, A. C., Robbins, T. W. & Everitt, B. J. The effects of intradimensional and extradimensional shifts on visual discrimination learning in humans and non-human primates. *Q. J. Exp. Psychol. B* **40**, 321–341 (1988).
194. Gossette, R. L. & Slonim, P. Spatial SDR performances across three species of New World monkeys. *Psychon. Sci.* **14**, 222–223 (1969).
195. Rumbaugh, D. M. in *Primate Behavior: Developments in Field and Laboratory Research* Vol. 1(ed. Rosenblum, L. A.) 1–70 (Academic Press, 1970).
196. Landauer, T. K. How much do people remember? Some estimates of the quantity of learned information in long-term memory. *Cogn. Sci.* **10**, 477–493 (1986).
197. De Beni, R. & Cornoldi, C. Effects of the mnemotechnique of loci in the memorization of concrete words. *Acta Psychologica* **60**, 11–24 (1985).
198. Legge, E. L., Madan, C. R., Ng, E. T. & Caplan, J. B. Building a memory palace in minutes: equivalent memory performance using virtual versus conventional environments with the method of loci. *Acta Psychol.* **141**, 380–390 (2012).
199. Brady, T. F., Konkle, T., Alvarez, G. A. & Oliva, A. Visual long-term memory has a massive storage capacity for object details. *Proc. Natl Acad. Sci. USA* **105**, 14325–14329 (2008).
200. Fagot, J. & Cook, R. G. Evidence for large long-term memory capacities in baboons and pigeons and its implications for learning and the evolution of cognition. *Proc. Natl Acad. Sci. USA* **103**, 17564–17567 (2006).
201. Gibson, B., Wasserman, E. & Luck, S. J. Qualitative similarities in the visual short-term memory of pigeons and people. *Psychon. Bull. Rev.* **18**, 979–984 (2011).
202. Hauser, M. D. Perseveration, inhibition and the prefrontal cortex: a new look. *Curr. Opin. Neurobiol.* **9**, 214–222 (1999).
203. Diamond, A. Developmental time course in human infants and infant monkeys, and the neural bases of higher cognitive functions. *Ann. N. Y. Acad. Sci.* **608**, 637–676 (1990).
204. Washburn, D. A. Stroop-like effects for monkeys and humans: processing speed or strength of association? *Psychol. Sci.* **5**, 375–379 (1994).
205. Catchpole, C. K. & Slater, P. J. *Bird Song: Biological Themes and Variations* (Cambridge Univ. Press, 2003).
206. Richland, L. E. & Burchinal, M. R. Early executive function predicts reasoning development. *Psychol. Sci.* **24**, 87–92 (2013).
207. Brady, R. J. & Hampton, R. R. Nonverbal working memory for novel images in rhesus monkeys. *Curr. Biol.* **28**, 3903–3910 (2018).
208. Terrace, H. S. The phylogeny and ontogeny of serial memory: list learning by pigeons and monkeys. *Psychol. Sci.* **4**, 162–169 (1993).
209. Ghirlanda, S., Lind, J. & Enquist, M. Memory for stimulus sequences: a divide between humans and other animals? *R. Soc. Open Sci.* **4**, 161011 (2017).
210. Izumi, A. Effect of temporal separation on tone-sequence discrimination in monkeys. *Hear. Res.* **175**, 75–81 (2003).
211. Fagot, J. & De Lillo, C. A comparative study of working memory: immediate serial spatial recall in baboons (*Papio papio*) and humans. *Neuropsychologia* **49**, 3870–3880 (2011).
212. Malassis, R., Dehaene, S. & Fagot, J. Baboons (*Papio papio*) process a context-free but not a context-sensitive grammar. *Sci. Rep.* **10**, 1–12 (2020).
213. Lind, J., Enquist, M. & Ghirlanda, S. Animal memory: a review of delayed matching-to-sample data. *Behav. Process.* **117**, 52–58 (2015).
214. Devkar, D. T., Wright, A. A. & Ma, W. J. The same type of visual working memory limitations in humans and monkeys. *J. Vis.* **15**, 13–13 (2015).
215. Wright, A. A., Santiago, H. C., Sands, S. R., Kendrick, D. R. & Cook, R. G. Memory processing of serial lists by pigeons, monkeys, and people. *Science* **229**, 287–289 (1985).
216. Aslin, R. N., Saffran, J. R. & Newport, E. L. in *The Emergence of Language* (ed. MacWhinney, B.) 359–380 (Psychology Press, 1999).
217. Inoue, S. & Matsuzawa, T. Working memory of numerals in chimpanzees. *Curr. Biol.* **17**, R1004–R1005 (2007).
218. Washburn, D. A., Gulledge, J. P., James, F. & Rumbaugh, D. M. A species difference in visuospatial working memory: does language link "what" with "where"? *Int. J. Comp. Psychol.* **20** https://doi.org/10.46867/ijcp.2007.20.01.02 (2007).
219. Greenfield, P. M. Language, tools, and brain revisited. *Behav. Brain Sci.* **21**, 159–163 (1998).
220. Terrace, H. S. The simultaneous chain: a new approach to serial learning. *Trends Cogn. Sci.* **9**, 202–210 (2005).
221. Spelke, E. S. & Kinzler, K. D. Core knowledge. *Dev. Sci.* **10**, 89–96 (2007).
222. Newport, E. L. & Aslin, R. N. in *Proc. 24th Annual Boston University Conf. Language Development* Vol. 1(eds Howell, S. C. et al.) 1–21 (Cascadilla Press, 2000).
223. Cantlon, J. F. & Brannon, E. M. Shared system for ordering small and large numbers in monkeys and humans. *Psychol. Sci.* **17**, 401–406 (2006).
224. Vallortigara, G. Core knowledge of object, number, and geometry: a comparative and neural approach. *Cogn. Neuropsychol.* **29**, 213–236 (2012).
225. Ferrigno, S., Hughes, K. D. & Cantlon, J. F. Precocious quantitative cognition in monkeys. *Psychon. Bull. Rev.* **23**, 141–147 (2016).
226. Smith, J. D. et al. Implicit and explicit categorization: a tale of four species. *Neurosci. Biobehav. Rev.* **36**, 2355–2369 (2012).
227. Tomasello, M., Hare, B. & Fogleman, T. The ontogeny of gaze following in chimpanzees, *Pan troglodytes*, and rhesus macaques, *Macaca mulatta*. *Anim. Behav.* **61**, 335–343 (2001).
228. Kuhl, P. K. & Miller, J. D. Speech perception by the chinchilla: identification functions for synthetic VOT stimuli. *J. Acoust. Soc. Am.* **63**, 905–917 (1978).
229. Piaget, J. & Inhelder, B. *The Psychology of the Child* (Basic Books, 1969).
230. Xu, F. & Kushnir, T. Infants are rational constructivist learners. *Curr. Direc. Psychol. Sci.* **22**, 28–32 (2013).
231. Kail, R. Speed of information processing: developmental change and links to intelligence. *J. School Psychol.* **38**, 51–61 (2000).
232. Zelazo, P. D. et al. The development of executive function in early childhood. *Monogr. Soc. Res. Child Dev.* **63**, i-151 (2003).
233. Terrace, H. S., Petitto, L. A., Sanders, R. J. & Bever, T. G. Can an ape create a sentence? *Science* **206**, 891–902 (1979).
234. Frye, D., Zelazo, P. D., Brooks, P. J. & Samuels, M. C. Inference and action in early causal reasoning. *Dev. Psychol.* **32**, 120 (1996).
235. Corballis, M. C. Recursion, language, and starlings. *Cogn. Sci.* **31**, 697–704 (2007).
236. Rey, A., Perruchet, P. & Fagot, J. Centre-embedded structures are a by-product of associative learning and working memory constraints: evidence from baboons (*Papio Papio*). *Cognition* **123**, 180–184 (2012).

237. Le Corre, M. & Carey, S. One, two, three, four, nothing more: how numerals are mapped onto core knowledge of number in the acquisition of the counting principles. *Cognition* **105**, 395–438 (2007).
238. Pepperberg, I. M. Animal language studies: what happened? *Psychon. Bull. Rev.* **24**, 181–185 (2017).
239. Rice, M. L. et al. Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *J. Speech Lang. Hear. Res.* **53**, 333–349 (2010).
240. Lillard, A. Developing a cultural theory of mind: the CIAO approach. *Curr. Dir. Psychol. Sci.* **8**, 57–61 (1999).
241. Sabbagh, M. A., Xu, F., Carlson, S. M., Moses, L. J. & Lee, K. The development of executive functioning and theory of mind: a comparison of Chinese and US preschoolers. *Psychol. Sci.* **17**, 74–81 (2006).
242. Devine, R. T. & Hughes, C. Relations between false belief understanding and executive function in early childhood: a meta-analysis. *Child Dev.* **85**, 1777–1794 (2014).
243. Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D. & Jones, T. Theory of mind and relational complexity. *Child Dev.* **74**, 1476–1499 (2003).
244. Stearns, R. E., Hartmanis, J. & Lewis, P. M. in *6th Annu. Symp. Switching Circuit Theory and Logical Design* 179–190 (IEEE, 1965).
245. Hopcroft, J. E. & Ullman, J. D. Some results on tape-bounded Turing machines. *JACM* **16**, 168–177 (1969).
246. Chomsky, N. *Syntactic Structures* (Mouton, 1957).
247. Gentner, T. Q., Fenn, K. M., Margoliash, D. & Nusbaum, H. C. Recursive syntactic pattern learning by songbirds. *Nature* **440**, 1204–1207 (2006).
248. Fitch, W. T. & Hauser, M. D. Computational constraints on syntactic processing in a nonhuman primate. *Science* **303**, 377–380 (2004).
249. Perruchet, P. & Rey, A. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychon. Bull. Rev.* **12**, 307–313 (2005).
250. Dedhe, A. M., Piantadosi, S. T. & Cantlon, J. F. Cognitive mechanisms underlying recursive pattern processing in human adults. *Cogn. Sci.* **47**, e13273 (2023).
251. Reich, P. A. The finiteness of natural language. *Language* **45**, 831–843 (1969).
252. Pullum, G. K. & Gazdar, G. Natural languages and context-free languages. *Linguist. Philos.* **4**, 471–504 (1982).
253. Jäger, G. & Rogers, J. Formal language theory: refining the Chomsky hierarchy. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **367**, 1956–1970 (2012).
254. Higginbotham, J. in *The Formal Complexity of Natural Language* (ed. Savitch, W. J.) 335–348 (Springer Netherlands, 1984).
255. Shieber, S. M. in *The Formal Complexity of Natural Language* (ed. Savitch, W. J.) 320–334 (Springer Netherlands, 1985).
256. Savitch, W. J., Bach, E., Marsh, W. E. & Safran-Naveh, G. (eds) *The Formal Complexity of Natural Language* Vol. 33 (Springer Science & Business Media, 2012).
257. Papadimitriou, C. H. *Computational Complexity* (Pearson, 1993).
258. Bommasani, R. et al. On the opportunities and risks of foundation models. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2108.07258 (2021).
259. Srivastava, A. et al. Beyond the imitation game: quantifying and extrapolating the capabilities of language models. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2206.04615 (2022).
260. Ganguli, D. et al. Predictability and surprise in large generative models. In *ACM Conf. on Fairness, Accountability, and Transparency* 1747–1764 (ACM, 2022).
261. Kaplan, J. et al. Scaling laws for neural language models. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2001.08361 (2020).
262. Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. Understanding deep learning requires rethinking generalization. Preprint at *arXiv* https://doi.org/10.48550/arXiv.1611.03530 (2017).
263. Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. Understanding deep learning (still) requires rethinking generalization. *Commun. ACM* **64**, 107–115 (2021).
264. Rakic, P. A small step for the cell, a giant leap for mankind: a hypothesis of neocortical expansion during evolution. *Trends Neurosci.* **18**, 383–388 (1995).
265. Rakic, P. Pre-and post-developmental neurogenesis in primates. *Clin. Neurosci. Res.* **2**, 29–39 (2002).
266. Bubeck, S. et al. Sparks of artificial general intelligence: early experiments with gpt-4. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2303.12712 (2023).
267. Mahowald, K. et al. Dissociating language and thought in large language models: a cognitive perspective. Preprint at *arXiv* https://doi.org/10.48550/arXiv.2301.06627 (2023).
268. Sutton, H. The bitter lesson. *Incomplete Ideas* http://www.incompleteideas.net/IncIdeas/BitterLesson.html (2019).
269. Marchetto, M. C. et al. Species specific maturation profiles of human, chimpanzee and bonobo neural cells. *eLife* **8**, e37527 (2019).
270. Laland, K. N., Kendal, J. R. & Brown, G. R. The niche construction perspective: implications for evolution and human behaviour. *J. Evol. Psychol.* **5**, 51–66 (2007).
271. Clarke, E. & Heyes, C. The swashbuckling anthropologist: Henrich on the secret of our success. *Biol. Philos.* **32**, 289–305 (2017).
272. Lefebvre, L., Reader, S. M. & Sol, D. Brains, innovations and evolution in birds and primates. *Brain Behav. Evol.* **63**, 233 (2004).
273. Ong, W. J. & Hartley, J. *Orality and Literacy* (Routledge, 2013).
274. Tinbergen, N. On aims and methods of ethology. *Z. Tierpsychol.* **20**, 410–433 (1963).
275. Krubitzer, L. In search of a unifying theory of complex brain evolution. *Ann. N. Y. Acad. Sci.* **1156**, 44–67 (2009).
276. Preuss, T. M. & Wise, S. P. Evolution of prefrontal cortex. *Neuropsychopharmacology* **47**, 3–19 (2022).
277. Smaers, J. B., Gómez-Robles, A., Parks, A. N. & Sherwood, C. C. Exceptional evolutionary expansion of prefrontal cortex in great apes and humans. *Curr. Biol.* **27**, 714–720 (2017).
278. Herculano-Houzel, S. *The Human Advantage: A New Understanding of How Our Brain Became Remarkable* (MIT Press, 2016).
279. Van Essen, D. C. et al. Cerebral cortical folding, parcellation, and connectivity in humans, nonhuman primates, and mice. *Proc. Natl Acad. Sci. USA* **116**, 26173–26180 (2019).
280. Herculano-Houzel, S. Numbers of neurons as biological correlates of cognitive capability. *Curr. Opin. Behav. Sci.* **16**, 1–7 (2017).
281. Charvet, C. J. & Finlay, B. L. Evo-devo and the primate isocortex: the central organizing role of intrinsic gradients of neurogenesis. *Brain Behav. Evol.* **84**, 81–92 (2014).
282. Clancy, B., Darlington, R. B. & Finlay, B. L. Translating developmental time across mammalian species. *NeuroScience* **105**, 7–17 (2001).
283. Takahashi, T., Nowakowski, R. S. & Caviness, V. S. Jr The mathematics of neocortical neuronogenesis. *Dev. Neurosci.* **19**, 17–22 (1997).
284. Zembrzycki, A. et al. Genetic mechanisms control the linear scaling between related cortical primary and higher order sensory areas. *eLife* **4**, e11416 (2015).
285. Cahalane, D. J., Charvet, C. J. & Finlay, B. L. Modeling local and cross-species neuron number variations in the cerebral cortex as arising from a common mechanism. *Proc. Natl Acad. Sci. USA* **111**, 17642–17647 (2014).
286. Nieder, A. Inside the corvid brain—probing the physiology of cognition in crows. *Curr. Opin. Behav. Sci.* **16**, 8–14 (2017).
287. Güntürkün, O., Ströckens, F., Scarf, D. & Colombo, M. Apes, feathered apes, and pigeons: differences and similarities. *Curr. Opin. Behav. Sci.* **16**, 35–40 (2017).
288. Güntürkün, O., von Eugen, K., Packheiser, J. & Pusch, R. Avian pallial circuits and cognition: a comparison to mammals. *Curr. Opin. Neurobiol.* **71**, 29–36 (2021).
289. Cowan, N. The magical mystery four: how is working memory capacity limited, and why? *Curr. Dir. Psychol. Sci.* **19**, 51–57 (2010).
290. Waltz, J. A., Lau, A., Grewal, S. K. & Holyoak, K. J. The role of working memory in analogical mapping. *Mem. Cogn.* **28**, 1205–1212 (2000).
291. Krems, J. A. & Wilkes, J. Why are conversations limited to about four people? A theoretical exploration of the conversation size constraint. *Evol. Hum. Behav.* **40**, 140–147 (2019).
292. Chater, N. & Oaksford, M. Ten years of the rational analysis of cognition. *Trends Cogn. Sci.* **3**, 57–65 (1999).
293. Geisler, W. S. Ideal observer analysis. *Vis. Neurosci.* **10**, 12 (2003).
294. Lieder, F. & Griffiths, T. L. Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behav. Brain Sci.* **43**, e1 (2020).
295. Griffiths, T. L., Lieder, F. & Goodman, N. D. Rational use of cognitive resources: levels of analysis between the computational and the algorithmic. *Top Cogn. Sci.* **7**, 217–229 (2015).
296. Sims, C. R. Rate–distortion theory and human perception. *Cognition* **152**, 181–198 (2016).
297. Bhui, R., Lai, L. & Gershman, S. J. Resource-rational decision making. *Curr. Opin. Behav. Sci.* **41**, 15–21 (2021).
298. Gershman, S. J. Origin of perseveration in the trade-off between reward and complexity. *Cognition* **204**, 104394 (2020).
299. Polanía, R., Woodford, M. & Ruff, C. C. Efficient coding of subjective value. *Nat. Neurosci.* **22**, 134–142 (2019).
300. Sims, C. R. Efficient coding explains the universal law of generalization in human perception. *Science* **360**, 652–656 (2018).
301. Hahn, M., Futrell, R., Levy, R. & Gibson, E. A resource-rational model of human processing of recursive linguistic structure. *Proc. Natl Acad. Sci. USA* **119**, e2122602119 (2022).
302. Cheyette, S. J., Wu, S. & Piantadosi, S. in *Proc. Ann. Meeting Cogn. Sci. Soc.* **43**, https://escholarship.org/uc/item/9hk7s32c (2021).
303. Yang, Y. & Piantadosi, S. T. One model for the learning of language. *Proc. Natl Acad. Sci. USA* **119**, e2021865119 (2022).
304. Hutter, M. *Universal Artificial Intelligence: Sequential Decisions Based on Algorithmic Probability* (Springer Science & Business Media, 2004).
305. Planton, S. et al. A theory of memory for binary sequences: evidence for a mental compression algorithm in humans. *PLoS Comput. Biol.* **17**, e1008598 (2021).
306. Grünwald, P. D. *The Minimum Description Length Principle* (MIT Press, 2007).
307. Solomonoff, R. J. A formal theory of inductive inference. Part I. *Inform. Control.* **7**, 1–22 (1964).
308. Miller, G. Human memory and the storage of information. *IRE Trans. Inform. Theor* **2**, 129–137 (1956).
309. Reber, A. S. Transfer of syntactic structure in synthetic languages. *J. Exp. Psychol.* **81**, 115 (1969).
310. Hick, W. E. On the rate of gain of information. *Q. J. Exp. Psychol.* **4**, 11–26 (1952).
311. Posner, M. I. & Rossman, E. Effect of size and location of informational transforms upon short-term retention. *J. Exp. Psychol.* **70**, 496 (1965).
312. Shannon, C. E. Prediction and entropy of printed English. *Bell Syst. Tech. J.* **30**, 50–64 (1951).
313. Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).

314. Macmillan, N. A. & Creelman, C. D. *Detection Theory: A User's Guide* (Psychology Press, 2004).
315. Chittka, L., Skorupski, P. & Raine, N. E. Speed–accuracy tradeoffs in animal decision making. *Trends Ecol. Evol.* **24**, 400–407 (2009).
316. Baddeley, A. The magical number seven: still magic after all these years? *Psychol. Rev.* **101**, 353–356 (1994).

Acknowledgements
The authors thank H. Clatterbuck for substantive comments and conceptual feedback on this manuscript, and C. Kaicher for help with scholarship, editing and formatting the article. This work was supported by the National Institute of Health (NIH) (grant number R01HD107840) and National Science Foundation (NSF) (grant numbers DRL2201843 and DRL2026416).

Author contributions
The authors contributed equally to all aspects of the article.

Competing interests
The authors declare no competing interests.

Additional information
Peer review information *Nature Reviews Psychology* thanks Stella Christie and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.