Cambridge Elements

Elements in the Philosophy of Mind
edited by
Keith Frankish
The University of Sheffield

INNATENESS IN MIND

Peter Carruthers
The University of Maryland

Imprint page

Innateness in Mind

Peter Carruthers
Department of Philosophy, University of Maryland
Author for correspondence: Peter Carruthers pcarruth@umd.edu

Abstract: This Element focuses on contemporary forms of nativism (belief in innateness), which mostly concern the existence of domain-specific learning mechanisms with innate structure and content. After sketching some innate capacities that are widely believed to be shared with other animals, the Element thereafter discusses a number of (alleged) distinctively-human ones. One concerns a faculty of language, another our capacity for representing the mental states of others (and derivatively, ourselves). It then turns to discuss some proposed innate adaptations that support culture. These include a number of learning biases, as well as affective learning mechanisms that enable swift acquisition of cultural values. The final two sections then discuss "tribal psychology." This may include an innate disposition to stereotype social groups as well as innate "tribal" motivations (both positive and negative). The over-arching thesis of the Element is that human nature might best be thought of as *culture-enabling* nature.

Keywords: affect, culture, mentalizing, language, stereotype

Contents

1 The Innateness Question

The debate that forms the background for this Element is almost as old as philosophy itself, often described as the question of "innate ideas" or "inborn knowledge." Plato (c.380 BCE) famously claimed that we have innate knowledge of the Forms (*Goodness*, *Beauty*, *The Perfect Square*, and so on), and Early Modern Rationalists such as Descartes (1637) and Leibniz (1704) argued that we have innate knowledge of the foundations of mathematics and geometry. In contrast, Empiricists like Locke (1690) and Hume (1739) claimed that all knowledge and all concepts must be derived from experience.

An initial challenge for early defenders of innate knowledge (hereafter "Nativists") was to explain where innate knowledge could come from. If some knowledge is "inborn" and unlearned then how did it get into the human mind in the first place? Plato's answer was the doctrine of recollection: our souls prior to birth had direct acquaintance with the Forms, and the challenge for philosophy is to recover (to "recollect") that knowledge. Descartes and Leibniz, on the other hand, believed that innate knowledge had been placed directly into the human soul by God, to be revealed by clear careful thinking. Empiricism, in contrast, was partly fueled by skepticism about these answers, influenced by the Empiricists' commitment to scientific naturalism (Carruthers 1992). Following the work of Darwin (1859), however, Nativists have acquired a much more plausible and scientifically coherent answer to the question of origins: the knowledge in question could be inherited from our forebears, a product of natural selection.[1]

It should be stressed at the outset that this Element will *not* be a survey and evaluation of recent debates between Empiricists and Nativists. In fact, very little will be said about the arguments that have been offered in support of Empiricism. These have been adequately responded to elsewhere. (See Pinker 2002; Marcus 2003; Laurence & Margolis 2024. Readers wanting to follow up on some of the varieties of contemporary Empiricism might consult Prinz 2012; Sterelny 2012; Heyes 2018; Buckner 2023.) Rather, our focus will be on reviewing and evaluating a range of different contemporary nativist positions with the aim of determining the strongest scientifically-informed case for Nativism. The goal of the Element is to lay out and discuss those aspects of the human mind that are most plausibly considered to be innate, for which the most convincing evidence has been provided thus far.

[1] Here and in what follows the term "knowledge" is used loosely (as it generally is in cognitive science) to refer to belief-like states of any sort—whether true or false, innate or learned, implicit or explicit—not only to justified or reliably acquired true beliefs.

Why does the question of innateness matter? Two main types of answer have been offered. One ties the innateness issue to the question of human perfectibility. Nativist answers have covered the full spectrum. At one extreme is Hobbes' (1651) claim that in the absence of the power of the state, human existence would be "nasty, brutish, and short," resulting from a "war of all against all." At the other is Rousseau's (1762) idealized "noble Savage." These provide very different conceptions of what an innate human nature might be like. Meanwhile, Pinker (2002) argues that the Empiricist picture of the human mind as a "blank slate" is perennially attractive because it is consistent with the Renaissance ideal of indefinite human progress and perfectibility. If humans are blank slates at birth, then it seems that they can be molded in whatever manner we choose, if only we can find the right kinds of cultural support and modes of upbringing.

It is important to note, however, that it is a fallacy to equate "innate" with "unalterable." While some innate properties may remain fixed across the life-span, many are modifiable. Some can be innately-structured initial learning systems that are thereafter precisified and/or elaborated through learning. Others may be just initial *defaults*—starting-state system-settings that can be modified by subsequent learning and experience. (Some initial settings of affective systems are like this. Bitter tastes are innately aversive, but one can be conditioned to enjoy them. Likewise for foods spiced with chili.) Alternatively, innate properties can be *biases*—fixed system-tendencies that can nevertheless be over-ridden with effort. (Perhaps the pre-Newtonian assumptions of our common-sense intuitive physics are of this sort; McCloskey 1983.) Note, too, that the extent to which an innate property may be changeable or over-rideable is very likely a matter of degree, and needs to be investigated on a case-by-case basis.

The second reason why debates between Empiricists and Nativists matter is simple: the question is intrinsically interesting. Anyone who is curious about human psychology will want to know what aspects of the human mind are *given* at the outset, and what results from learning and enculturation. The innateness question asks about the basic foundations of human psychology. Are those foundations general ones, in such a way that knowledge in different domains is acquired in essentially the same way (through association or Bayesian learning, perhaps)? Or are they domain-specific, with specialized learning systems embodying innately encoded knowledge operating in some domains but not others? This Element will touch on issues related to the first reason why the innateness-issue is interesting (specifically, the suggestion that our "tribal psychology" makes inter-group conflict well-nigh inevitable), but its main focus is to understand the foundations of our psychology.

It should be acknowledged, however, that not all innate knowledge need be given at the outset of

development; much may be acquired later in the absence of learning. Indeed, it is quite likely that some of our knowledge might be arrived at via biopsychological maturational processes (Elman et al. 1996; Oyama 2000; Tucker-Drob et al. 2013; von Stumm & Nancarrow 2024). But nothing further will be said about such possibilities here. We will mostly be considering evidence of early acquisition in infancy combined with arguments from likely adaptations. This is because it is, at present, extremely hard to disentangle knowledge that is acquired via maturation (especially experience-dependent maturation) from knowledge that is learned. (Think, here, of the effect of the lack of a father in the home on the age of a girl's first menarche; Webster et al. 2014, or the cognitive and motivational differences between males and females that emerge around puberty; Kretzer et al. 2024.) Put differently: our focus will be on knowledge that everyone agrees is learned somehow, and our question will then be *how* it is learned—via some sort of general-process learning or through innately structured domain-specific learning systems.

Before we embark on the task of evaluating the case for various forms of innate knowledge, however, more needs to be said about what innateness itself *is*. For some have argued that the innateness-concept is irredeemably confused, and should be retired from serious scientific discourse. Griffiths (2002) notes, in particular, that innateness-claims tend to cluster around three distinct sets of ideas. One is that innate properties are developmentally fixed, appearing during development in a wide range of environments. In a word, innate properties are "canalized" in the way they develop. The second cluster of ideas is that innate properties belong to a species' nature; they are universal to the species, and are found in all members of the species. Then the final set of ideas has to do with innate properties being evolutionary adaptations; they are the intended outcome of normal development; they are the properties that normal members of the species are *supposed* to have. Yet Griffiths provides many real examples where one or two of these properties are present in the absence of the others. They are not always instantiated together.

Griffiths (2002) argues, in fact, that the innateness concept is best understood as a manifestation of our folk-essentialism, which is itself pan-cultural and early-developing in children around the world (Gelman 2003). Essentialism is the view that members of any given living species share an internal physical *essence*. This essence reliably causes the manifest properties of individual members of the species (which are thus canalized); it delimits the boundaries of the species, such that all and only members of the species have that essence (it is universal, and the properties it causes are universal); and individuals that fail to manifest the essence-caused properties are *abnormal* members of the species—they don't have the properties they *should* have, that they are *supposed* to have by their nature.

Species-essentialism is widely agreed by biologists to be false. Members of a given species don't share

any single underlying property. Rather, their DNA is more-or-less closely similar, with significant individual variation, and it does its developmental work across variation in internal and external environments. Yet Griffiths (2002) argues that our implicit commitment to essentialism turns the concepts of innateness and human nature into natural attractors, pulling together the three different strands noted above in ways that are scientifically illegitimate and that lead to unsound inferences. As a result, he thinks the innateness-concept should be dropped altogether, and that we should just talk about canalization, universality, or adaptation as needed in any given context.

Notice, however, that the concept of innateness that Griffiths is discussing is intended to apply to biological properties generally, not just to properties of the mind. (This is made especially clear in a follow-up paper, Griffiths et al. 2009.) So it would be just as applicable to properties like having opposable thumbs as it is to properties like the capacity to learn language. And Griffiths is quite right that biologists have no real use for the innateness-concept. Rather, all biological features result from a complex interplay between genes, on the one hand, and intra-cellular and extra-cellular environments, on the other. Historically, however, innateness-debates were never about the properties of living beings generally. They were specifically about the properties of the human mind. If we look at what innateness-debates in philosophy and cognitive science have actually been *about*, and what has been at stake, we get a very different take on the concept of innateness that is at the heart of those debates.

Samuels (2002) argues that in the context of cognitive science, to say that some mental property is innate is to say that it is psychologically *primitive*. That is, it doesn't admit of any sort of psychological explanation (although it might have a neurological or biochemical one). Put differently, innateness is what marks the boundary of psychology—innate properties are those that enter into psychological explanation, but which don't themselves admit of any psychological explanation; they are what are psychologically *given*. In this sense, basic learning mechanisms are innate, of course, as Empiricists are happy to agree. For one cannot, from nothing, learn how to learn. The real debates are about the number and nature of the learning mechanisms that issue in mature human psychology. Are they few in number and general in scope, as Empiricists maintain? Or are there lots of them, specialized for learning in particular domains (such as language, mentalizing, navigating, and so on)?

There are obvious problems with primitivism, however, as Samuels (2002) himself recognizes. The main one is that it overgeneralizes, implying innateness where plainly none is present. For instance, localized brain damage can cause someone to believe that they are dead (Cotard's delusion) or to believe that their spouse has been replaced by an impostor (Capgras syndrome). Such beliefs are, obviously, not innate. But

they don't admit of a psychological explanation in terms of other mental states or processes. So in that sense they are primitive. Samuels' solution is to point out that such beliefs aren't *normal*. He thus claims that the correct account of innateness is that innate mental properties are those that emerge in ways that lack a psychological explanation (that is, without learning) *in the course of normal development*.

As Ritchie (2021) points out, however, it is by no means clear what counts as "normal" development. For example, depression among adolescents is quite common (normal?), but it generally lacks a psychological explanation (as opposed to one in terms of some sort of biochemical imbalance). Yet no one would say that adolescent depression is innate. Moreover, it is far from clear how one should police the distinction between psychological explanations and explanations of other kinds. To say that something is psychologically primitive (and hence innate, if appearing normally), for Samuels, is to say that it cannot be *psychologically* explained. But what qualifies as an instance of the latter? Is a computational explanation psychological, for example? And what about explanations that deploy neural-accumulator models, of the sort that are very widely used across cognitive science? Does the fact that they are modelled on the way neural activity builds up in local populations of neurons prevent them from qualifying as psychological, or not? It is unclear how one might give a principled answer.

In consequence, Ritchie (2021) defends the simple view: to say that a psychological property of any kind is innate is just to say that it is *not learned*. (See also Knobe & Samuels 2013, who show that both scientists and laypeople adopt this construal when considering general psychological properties.) For this is what debates between Empiricists and Nativists are actually about: what is or is not learned, and how extensive the learning mechanisms themselves are (which are themselves unlearned, of course). And as for what learning itself is, Ritchie suggests that we should be minimalists at the outset of inquiry: learning is any process, of any sort (associationist, Bayesian, inference to the best explanation, or whatever) that extracts information from input in accordance with some or other set of rules.

As for properties that result from brain damage or biochemical imbalances, Richie suggests that they are irrelevant in the contexts in which questions of innateness arise. No one thinks that Capgras delusion results from learning of any sort, whether domain-general or domain-specific; likewise for many forms of depression. So the innateness-concept doesn't need to be defined in such a way as to exclude cases like this. They just aren't relevant. Where questions of innateness *do* arise are in contexts where everyone thinks learning of some sort is in play, and the question then becomes: learning of *what* sort? Is it a matter of *general* learning? Or does the learning involve some domain-specific mechanism that implicitly or explicitly encodes unlearned information about the domain with which it deals (Laurence & Margolis

2024)? It is just such questions that will occupy us going forward. Accordingly, I, too, propose to use "innate" to mean just "unlearned."

Finally, a word about the methodology to be employed in later sections of this Element. The arguments we will be considering are all inferences to the best explanation from a combination of three sources of evidence. One is that the knowledge in question is just what one might predict would be an innate adaptation, given what we know about the selective pressures that likely operated on our ancestors. Another is that the knowledge in question is universal to humans, reliably emerging across a wide range of developmental and ecological conditions. (That is, the knowledge is apparently *canalized*.) And the third is that it emerges early in human infancy, despite little or no opportunities for learning. Note that these are all properties that Griffiths (2002) thinks are easily confused with innateness itself. In contrast, here they are to being used (especially in combination) as *evidence* of innateness.

2 Our Animal Inheritance

There are numerous unlearned systems and capacities of the human mind, many of which are shared with other primates, or more generally with other mammals (and some even with birds and insects). Indeed, we have known from the very earliest studies of brain anatomy that the human brain is by no means an equipotent association mechanism. On the contrary, there is an intricate system of partially-independent networks and regions that are designed to do specialized jobs, containing multiple types of neuron and involving a wide variety of neurotransmitter networks. There are innately-structured perceptual systems and homeostatic-monitoring systems. There are a variety of distinct or partially-distinct systems for storing previously-acquired information—episodic (or what-where-when) memory, semantic memory (including map-like representations of what is where in one's environment), prospective memory (for storing plans and intentions), sensorimotor memory (for storing skills), and so on. There are attentional systems that can target any sensory domain and orienting systems for attracting attention. And there are a variety of affective (emotional and motivational) networks that are likewise at least partially distinct from one another. Finally, there are executive networks that receive input from many others, take decisions, and direct the activity of attentional and motor networks. In short, human and other animal minds have an elaborate innate system-architecture (Carruthers 2006).

Contemporary Empiricists probably do not (and should not) deny any of this. But notice that, as a result, they are forced to concede the existence of a significant number of innate domain-specific learning mechanisms. The visual system enables us to learn the colors, textures, shapes, and positions of the objects around us. The auditory system enables us to learn of the nature and directions of local sound-

generating events. The tactile system enables us to learn about things that come into contact with our bodies. And so on. There are also systems for learning about ambient and bodily temperature, for learning of one's need for food or liquid, of one's need for sleep, and so on. These are all distinct from one another, are innate, and tacitly encode information about the domains with which they deal.

Traditionally, all learning in animals that takes place down-stream of the input systems mentioned above has been thought to result from gradual strengthening and weakening of associative connections.[2] But even quite simple creatures like ants and bees can learn in ways that cannot easily be explained in terms of mere association. Almost all living creatures (including humans) can navigate by dead reckoning, for example. A Saharan desert ant can leave its nest (a hole in the featureless floor of the desert landscape) to go foraging, taking a meandering route across the ground until it comes across an item of food. It then bites off a chunk, turns, and takes a direct path over approximately the correct distance towards its nest, at which point it stops and engages in a local search for the hole. Having deposited its haul, it can then turn around and head directly back to the food it had found, without having to repeat its previous search (Wehner & Srinivasan 1981). As we now know, the ant does this by keeping track of its heading each time it turns during the initial search (computed from the position of the sun in the sky), integrating this with a rough measure of distance traveled in each direction (computed from an estimate of the number of steps that it takes), to keep a running tally of its distance and direction from the nest. And then having returned to the nest, it can reverse the directional information to make its way back to the food again. This is a representational-computational learning system designed for operating in a particular domain of knowledge, not an associationist one (Gallistel 2000).

Moreover, there are well-established findings from the very heartland of theories of associative learning (namely behavioral conditioning) that are very hard to explain from an associationist perspective (Gallistel 2000). One is the time-scale invariance of conditioned learning. That is, it makes no difference to the average number of reinforcements needed for the acquisition of a behavior how extensive the interval is between the cue and the reward, provided that the inter-trial intervals are increased proportionately. This is quite hard to account for in terms of strengthening associations. For one would think that the increased time between the cue and the availability of reward would make it harder to

[2] For this reason, much of the discussion that follows in the remainder of this section contrasts Nativism with associationism. There are other options for Empiricists to adopt, of course, such as domain-general, unbiased, forms of Bayesian learning with no built-in priors. But it seems unlikely that such approaches would be any more successful in explaining the animal data we will be discussing than associationism is.

associate the two, and it is unclear why increasing the wait-time between trials should make any difference. From a computational perspective, however, the finding makes sense. For the predictive value of a cue remains the same as time increases provided that rewards are correspondingly sparsely distributed overall.

Relatedly, the number of rewards required for acquisition of a behavior remains the same no matter what reward-schedule the animal is on. For example, an animal on a 1-1 schedule (where there is a reward available every time the behavior is enacted following a cue) might take 40 rewarded trials to reach criterion, whereas an animal on a 1-10 schedule (where in nine out of ten cases no reward is given) might take 400. But it takes the same number of *rewarded* trials either way. (The equivalent phenomenon occurs in conditioned extinction.) This, too, is very puzzling from an associationist perspective. One would think that those nine-in-ten unrewarded trials would make it harder to establish an association between cue and reward; but it doesn't.

Moreover, we know that mice (as well as humans) can compute the approximate likelihood of a given outcome, adjusting their decision-making accordingly (Balci et al. 2009; Khefets & Gallistel 2012). For example, mice might be offered a choice among two locations where a nose-poke has a chance of securing a reward (where the value of the reward itself is fixed, whenever one is received). Not only do the chances of receiving a reward differ between the two locations, but at random intervals those chances themselves shift. So initially there might be a 20 percent chance of getting a reward for a nose-poke in location A and a 50 percent chance of reward from location B; but then after a random interval the chances shift (without being signaled in any way) so that there is now a 70 percent chance of reward from A and 40 percent from B; and so on. It turns out that mice (and humans in a similar sort of set-up) can track these changes about as swiftly and accurately as an ideal Bayesian reasoner could. Again, this is very hard to explain from an associationist perspective, and suggests the existence of an innate mechanism for computing and tracking likelihoods—although it could be explained in terms of a domain-specific Bayesian learning mechanism, of course, designed specifically for this purpose.

In fact, there are now known to be numerous domain-specific learning systems operative in non-human animals, as well as in human infants during the first year of life or earlier. Many of these are known collectively as systems of "core knowledge." The evidence concerning them is rigorously reviewed in some detail in Spelke (2022). Perhaps one of the best-known and most well-established is the system for computing and reasoning about the approximate number of items in a set, known as the "numerosity" of the set. These representations are subject to Weber's Law, enabling creatures to distinguish ratios but not

absolute differences. Thus a newborn infant might be able to distinguish a group of 5 things from a group of 10, and a group of 10 from 20, but not be able to distinguish between 15 and 20. Numerosity-discrimination starts out relatively crude (as here), but becomes increasingly precise with experience. So not only is this an innate learning system, but it is one that becomes better at its job with practice.

Both infants and animals can add and subtract approximate numbers, too. Thus an infant who watches 10 items disappear behind a screen to be joined by 10 more might be surprised when the screen is removed to reveal only 10 items in total.[3] Likewise if 20 items go behind the screen and then 10 of them come out, the infant might be surprised when the screen is removed to show that there are still 20 there. Moreover, members of many species of animal can divide approximate numbers as well as add and subtract them. In fact, there is an extensive literature on so-called "optimal foraging" which presupposes just this (Pyke et al. 1977; Stephens & Krebs 1986). In order to decide whether to stay in a given foraging site or move back to another that has recently been visited, an animal might need to calculate and compare the rate-of-return for each. This requires the animal to estimate the number of food-items retrieved and divide it by the amount of time spent foraging there. For to calculate a rate, one needs to divide the number of items by elapsed time (Gallistel 1990).

Another innate learning system well-studied in both animals and infants is a naïve physics system, charged with tracking and predicting the movements and interactions of physical objects. It is now thought that this is realized in the form of a physics emulator, similar to those used in many artificial game-engines, which is insensitive to the identity and exact shape of the objects that it tracks (Ullman et al. 2017). The latter fact explains a long-established but puzzling finding from the infancy literature, which is that an infant seeing a toy duck disappear behind a screen will show no surprise when a toy *truck* of approximately the same overall shape and size emerges from the other side after an appropriate interval (as if it had changed from the one to the other while travelling behind the screen). For the emulator is insensitive to object-identity, just tracking speeds, directions of motion, and the likely results of collisions. Only later do infants start to integrate the output of the emulator with their knowledge of object-identity.

[3] The measure used here, and in many of the experiments with infants discussed in this Element, is looking time. Infants will look longer at outcomes that conflict with their expectations than at outcomes that are consistent with their expectations. So in carefully controlled conditions one can use looking time to figure out what an infant had expected to happen. For an in-depth review of the violation-of-expectation looking-time paradigm, its pitfalls, strengths, and weaknesses, see Margoni et al. (2024).

Many creatures (again, including very young infants) distinguish sharply between the motions of physical objects (e.g. a ball rolling down a slope) and the goal-directed motions of agents. Infants use a number of cues to identify agency (either singly or together). One is self-initiated movement, such as a wooden block that starts moving and shifts directions of its own accord. Another is the presence of eyes or eye-like shapes. And yet another is apparent goal-directed movement, such as a square that moves around obstacles or jumps over a wall to end up next to a colored circle. Once a goal has been identified, infants will expect the agent to select the most efficient means to achieve the goal. Moreover, in more naturalistic settings, many animals are attuned to the distinctive properties of biological motion (Vallortigara et al. 2005; Bardi et al. 2011). So there appears to be an innate learning system that identifies agency and can predict the likely movements of an agent towards a goal.

There is also thought to be an innate system that can compute and store the geometry of a space. Indeed, both animals and pre-linguistic children mistakenly rely on this alone when re-orienting themselves, ignoring all other cues (Hermer & Spelke 1996). For example, children placed in a small rectangular room can be shown a toy being hidden in a corner adjacent to a colored or patterned wall (where the other three walls are white). The children are blindfolded and turned around a few times to disorient them. The blindfold is then removed, and they are invited to retrieve the toy. Children search equally often in the two geometrically-equivalent corners, ignoring the colored or patterned cue that would enable them (when combined with geometric information) to identify the correct one. Rats perform in very similar ways. This and a raft of other evidence suggests that there is a dedicated system for representing the geometry of an environment that plays an especially important role in some aspects of navigation.

Finally, in this catalog of some of the innate learning systems that we share with other animals, we should mention that there are a number of different systems for motivation and evaluative learning. These are not often discussed in the context of Empiricist-Nativist debates, but they should be.[4] There are multiple such systems, including for food, hydration, warmth, threat-avoidance (fear), aggression (anger), mating,

[4] One exception is the debate over so-called "basic emotions," where some have claimed that only the dimensions of valence (positive versus negative) and arousal (high versus low) are given at the outset of cultural learning (Russell 2003; Barrett et al. 2007). This "core affect" account has been successfully critiqued on multiple occasions (Izard 2007; Panksepp & Watt 2011; Scarantino & Griffiths 2011; LeDoux 2012). As a result, a basic-emotions framework will be assumed in what follows. Moreover, the core-affect account obviously fails to extend to basic homeostatic affective systems such as hunger and thirst—although admittedly, common-sense psychology doesn't treat these as emotions in the way that many affective scientists do (e.g. Rolls 1999).

nurturance of offspring (among mammals), and more. LeDoux (2012) calls them "survival circuits," though even a cursory look at the examples just given shows that they would be better described as *inclusive fitness* circuits. They are innate affective systems that have been selected by evolution and modified as appropriate across species because of the fitness-benefits that they confer.

Each of these affective systems responds to a different set of inputs (threats to one's physical integrity, in the case of fear; calorie depletion, in the case of hunger; noxious tastes, in the case of disgust; and so on), and each comes equipped with a distinctive set of behavioral dispositions as output (running or freezing according to the circumstances, in the case of fear; approaching and eating, in the case of hunger; avoiding and/or retching, in the case of disgust; and so on). Yet all also converge in producing as output some degree of valence (positive or negative) directed at the appraised object or situation. This provides a common currency for decision-making among options, as when bees adaptively trade off the strength of a sucrose solution against degrees of noxious heat (Gibbons et al. 2022). And valence in turn is best thought of as an analog-magnitude representation (that is: fine-grained and nonconceptual) of the contextually-modulated value (or disvalue) that the individual attaches to the thing or situation in question (Carruthers 2024).

Each of these systems comes equipped with some innate evaluative settings, referred to as "primary rewards and punishments" in much of the literature. Thus bitter tastes are innately repulsive to mammals, whereas sweet ones are innately attractive; pain sensations are innately unpleasant, whereas light stroking touch (at speeds of approximately three centimeters per second) is experienced innately as pleasant (Taneja et al. 2021); and so on. But each system is capable of learning new intrinsic values that are associated with, or predictive of, primary ones. (These are often referred to as "secondary" rewards and punishments.) Thus children can learn to like the bitter taste of broccoli by having it dipped in sugar or peanut butter (Yeomans et al. 2008). After a while they come to enjoy the taste of broccoli for its own sake, on its own. And rats will come to fear a light that is associated with electric shock, continuing to fear it thereafter even when it is no longer accompanied by pain.

Many of the value-learning mechanisms inherent in these innate affective systems are structured differently from one another. Thus learning to value novel tastes can require multiple exposures, whereas nausea-induced disgust results from one-off learning. The latter phenomenon will already be familiar to many readers. On waking up vomiting having eaten a chicken salad the night before, one is apt to find the very thought of eating chicken intrinsically repulsive (whereas salads themselves continue to be attractive). It seems that the disgust system is constructed in such a way that nausea initiates a search of

memory for the most likely culprit from among items recently eaten, immediately tagging the most likely candidate as disgusting. Interestingly, it seems that learning the positive value of certain foods can also operate retrospectively, albeit responding to cues that are deeply unconscious. Thus the reward-value of sugary and fatty foods is learned, in part, from signals received directly from the gut that respond to carbohydrates and fat respectively (Li et al. 2022). Since digestion takes time, this, too, would seem to require a retrospective search of memory for the most likely sources (such as a recently-eaten beef-burger as opposed to the accompanying broccoli).

Some of these systems have a complex inner structure, too, that may not be shared with others. Thus the fear-system in mammals, for example, comprises two partially-distinct circuits. One is wired directly into the motor system, and responds when a threat is imminent; the other projects forward to executive systems in prefrontal cortex and is operative when the threat is still distal (Mobbs 2020). Such a design makes good adaptive sense. If a predator is charging towards one, a response needs to be made immediately; there is no time to think or plan. In contrast, if one is being stalked by a predator that is still some distance away, then it may be more adaptive to plan one's best mode of escape (or alternatively, to initiate an aggressive threat-response).

We can conclude that Nativism is massively true of the human mind, at least when the human mind is considered as an instance of primate minds more generally. There are many unlearned capacities and systems in the primate mind, and many innate learning systems specialized for aspects of cognition and motivation. The remainder of this Element will consider whether there are any distinctively-human domain-specific learning systems as well (and if so, which ones they are), or whether the human mind should just be seen as a scaled-up version of the general primate mind.

3 A Language Faculty

Contemporary debates about human-specific innateness began in the 1950s, focusing especially on the question of an innately-structured faculty for the acquisition and use of natural language (Chomsky 1957). Chomsky was highly critical of the associationism assumed by then-dominant behaviorist theories in psychology. He argued that natural languages are governed by complex sets of hierarchical ("tree-structure") rules that permit some modes of word-combination while excluding others, and which distinguish possible from impossible interpretations of sentences. The view soon coalesced into the view that all natural languages share an innate universal grammar ("UG") that children bring to bear when learning their native language. Initially, the idea was that UG is richly structured from the outset, embodying a good deal of information about the universal features that all natural languages share

(Chomsky 1965). By the turn of the century, however, the idea had weakened to the possibility that UG might comprise only an innate expectation of hierarchical phrase-structure that enables phrasal components of sentences to be moved and re-combined (Hauser et al. 2002). But the nature and richness of UG continues to be a matter of active investigation.

Lidz & Gagliardi (2015) point out that there are two interlinked factors in phrase-structure grammars: the phrases, on the one hand, and the possibility of movement, on the other, enabling rearrangement of those phrases in a variety of different ways. Consistent with the suggestions of Hauser et al. (2002), there is evidence that infants come to the learning task already knowing this. For example, Takahashi & Lidz (2008) taught 18-month-old infants and adults a simple artificial language that they would have had no previous experience of. The training data contained some of the statistical signatures of phrases, but no examples of phrasal movement. Yet the infants were immediately able to distinguish sentences with moved phrasal constituents from those that involved moved non-constituents (as could the adults). This argument is not rock-solid of course, since it remains conceivable that the infants somehow generalized the idea that phrases are moveable from their nascent knowledge of their native language. But it remains to be shown how this could be done.

One of the traditional arguments for richer forms of UG is known as "the poverty of the stimulus." The claim is that children develop syntactic knowledge of their native language that was never present in the input, and that isn't generalizable from what can be learned directly from the input, either. Many such arguments have been offered over the years, although it is fair to say that none has convinced everyone. For example, Leddon & Lidz (2006) investigated young children's (and adults') interpretations of the following four sentences of English.

(1) Sally knew that Jane put a picture of herself on the wall.
(2) Sally knew that Jane was very proud of herself.
(3) Sally knew how proud of herself Jane was.
(4) Sally knew which picture of herself Jane put on the wall.

In all three of (1), (2), and (3) the reflexive pronoun "herself" can only be interpreted as referring to Jane. But despite its surface similarities to the others, in sentence (4) the pronoun's reference is ambiguous, and would need to be disambiguated from the context—it could either refer to Jane or to Sally. By the age of four years, children know this. But when Leddon & Lidz examined a corpus of child-directed speech containing over 10,000 phrases of the syntactically-relevant type (containing a wh-phrase, as [4] does), they found none at all containing a reflexive pronoun. The implication is that children must have come to the learning task with enough pre-existing grammatical knowledge to be able to predict the differing ways

in which sentence [4] can be interpreted.

It might be felt that these and other arguments for an innate universal grammar are moot, given the amazing successes recently achieved by large-language models (LLMs) and deep-learning AI systems, such as ChatGPT. These are multi-layered associative-learning networks that are not programmed with any knowledge of language in advance. Yet they can write convincing essays and stories, and can hold natural-seeming conversations in intelligible English (or whatever language they are initially trained on). They are an existence-proof that linguistic ability can be acquired through associative learning alone.

It certainly doesn't follow from this that *humans* learn languages associatively, however. For one thing, the amount of evidence available to human infants is many orders of magnitude smaller than the training-data used for large-language models. A human child before the age of two (when the so-called "language explosion" begins) will hear somewhere between 60,000 and 215,000 words per week (between 3 million and 11 million words over the course of a year), depending on their socio-economic status, or SES (Hart & Risley 1995). AI language systems, in contrast, are trained on millions of billions of examples, somewhat as if a child had been presented with multiple iterations of the entire contents of the internet.

Moreover, the numbers above vastly overestimate the extent of the learning opportunities that infants in the first two years of life can utilize when learning the meanings of words. For many of those words will be overheard, rather than addressed to the child, and we know that the latter is crucial for word-learning to occur and is often only effective when adult and child mutually attend to the same item (Baldwin 1991; Bloom 2000). For example, even when interacting with an adult who uses a novel noun or verb in the presence of the relevant object or activity, only a small proportion of those uses are instances of what Gleitman & Trueswell (2020) call "gems" as opposed to "junk." When adults and slightly older children view videos of such scenes (whether seen from a third-person perspective or from a camera mounted on the toddler's head), in which the sound has been removed but a beep occurs at the time the target word is uttered, in only a small proportion of cases are participants above-chance in their guess at the word's meaning. Those instances seem especially characterized by the exact timing of the adult-child social interaction. And the resulting learning is not gradual, as associative learning is, but (almost) one-off—the child makes a guess at the likely referent and compares that with the next "gem-like" occurrence of the same word, storing that meaning in their lexicon if the two match. It seems safe to conclude, then, that human children learn language in nothing like the way that an LLM does.

Even people who reject the existence of an innate universal grammar, however, need to acknowledge that

humans have a unique innate drive to pay attention to linguistic and other social cues, and to communicate with others. For example, although a chimpanzee can be trained to use a pictogram-language with a vocabulary of a few hundred words, and can combine those words into simple sentences governed by word-order (without hierarchical embedding), almost all of the chimp's uses of the language are imperative in nature, requesting things that it wants (Rumbaugh et al. 2003). Human children towards the end of their first year of life, in contrast, begin to draw their care-givers' attention to interesting phenomena by pointing, and many of their early uses of language are descriptive rather than imperative. Indeed, some have argued that humans are distinctive in having an innate motivation to "share intentionality" with others (exchanging thoughts and ideas; Tomasello 2010).

A particularly striking example of the human drive to communicate is provided by deaf children who grow up in hearing communities, isolated from other deaf people, and who thus have no exposure to any form of natural language. They nevertheless invent their own gestural systems of communication, called "home-sign" (Brentari & Goldin-Meadow 2017). These systems are not learned from the child's care-givers; on the contrary, many care-givers start to use crude versions of the gestures initially used by their child. Home-sign has some of the properties of a simple natural language, and can be used to comment on things that are displaced in time and/or space, to tell stories, to communicate hypotheticals, and even to comment meta-linguistically on the signs of the child and others. Indeed, recent evidence demonstrates that home-signers create communicative systems that possess hierarchical phrase-structure, too (Goldin-Meadow 2020). Thus, when deprived of any language to attend to and learn from, human children are driven to invent one for themselves.

Even more striking is what happens when home-signing deaf children are brought together and start to communicate with each other (Senghas et al. 2004). A natural experiment of this sort was studied in real-time by linguists in the 1980s, when the new government of Nicaragua founded the first school for deaf children and started bringing in such children from their isolated villages around the country, where each child had constructed their own version of home-sign. Within a year or two the first generation of children brought to the school had developed a common "pigeon" sign-system (a sort of shared home-sign, lacking many of the properties of grammatical natural language). But the next generation of students to arrive, exposed to the existing pigeon sign-system, spontaneously enriched it to include not only a more elaborate hierarchical phrase-structure, but also such grammatical features as verb agreement. It seems that the structures inherent in what is now known as Nicaraguan sign-language (a full natural language like any of the languages used by long-standing deaf communities around the world) came from the minds of the children themselves, in the absence of any model. Many think that this is perhaps the

strongest argument for some form of innate universal grammar.

We noted above that there is little doubt that humans (uniquely) are innately motivated to communicate with others, paying attention to, learning, and attempting to utilize the systems of symbols that others employ. There is also little doubt that humans are cognitively adapted for language, too, even if the language faculty turns out not to embody any kind of universal grammar. Some initial claims along these lines turned out to be false, however. For instance, humans are by no means unique in engaging in vocal learning. Many songbirds, too, differentially attend to and copy the songs of conspecifics. And likewise, many of the networks involved in the production and categorical perception of sound are shared across multiple species. Even the permanently lowered larynx of human adults, although unique among primates, is found in many other mammals, including lions and some species of deer (Fitch 2018). Moreover, careful measurement and modeling of the vocal tract of macaque monkeys shows that they are in principle capable of producing a range of vowel sounds—albeit not quite as many as humans (Fitch et al. 2016). So there appear to be no physiological impediments that prevent monkeys from speaking.

Indeed, the evolution of human language capacities might best be thought of as involving the recruitment and combination of resources for vocal learning, hierarchical-structure learning, and sociality that individually have deep evolutionary roots (Arnon et al. 2025). Nevertheless, there do appear to be a pair of human-specific brain networks involved in enabling human speech that are unique to humans (Fitch 2018). One is a set of direct monosynaptic links from motor cortex to laryngeal neurons in the brain-stem, enabling fast and direct control of laryngeal movements. The other is a vastly expanded bi-directional network linking regions of premotor cortex (specifically Broca's area, a ventrolateral region of the prefrontal cortex normally located in the left hemisphere) and regions surrounding auditory cortex (Wernicke's area, located on the upper surface of temporal cortex extending back to the lower regions of parietal cortex, also normally in the left hemisphere). The network in question is comprised of a very large long-distance bundle of white-matter neurons (which are myelinated for speed) called "the arcuate fasciculus."

Broca's area and Wernicke's area work together in reciprocal interaction to generate and comprehend meaningful speech.[5] Broca's area plans and initiates speech production, with the motor instructions

[5] Although I use the terms "Broca's area" and "Wernicke's area" for convenience and because of their familiarity, recent evidence suggests that the areas in question are only involved in the motor production and the perception of language, respectively, rather than the core linguistic processes involved in mapping meanings to forms (in

involved creating forward models of the sounds they would produce that are received in Wernicke's area prior the act of speaking itself. The latter area then interacts with widespread semantic knowledge in temporal cortex and elsewhere to check whether the resulting sounds would express the intended meaning. This is why brain-damaged people suffering from Wernicke's aphasia can generate smooth and grammatical speech, but speech that is often meaningless, of the "green ideas sleep furiously" variety. Speech comprehension, in turn, depends on signals sent forward through the network to Broca's area, which is thought to be critical for building the appropriate syntactic structures to help interpret the sounds in question (Hickok & Poeppel 2007).

One long-standing suggestion concerning the evolution of human capacities for syntax has focused on Broca's area in particular, which is uniquely expanded relative to the rest of the cortex in hominins, and especially in humans. The idea is that syntax may have evolved initially from the representations that underlie tool construction and complex kinds of action, issuing in forms of gestural communication in the first instance (Hauser et al. 2002; Tomasello 2010). For many skills, too, are hierarchically organized, constructed out of habit-like chunks that can be combined together in various sequences, just as linguistic phrases can be. (Thus one might chop and fry the onions before cracking and whisking the eggs when making an omelet, or vice versa.) And Broca's area seems to be involved not only in the syntactic aspects of speech production but also in hierarchical forms of action more generally.

Recent evidence suggests, however, that the regions of cortex that underlie syntax and those that are involved in action-planning more generally are adjacent but distinct (Gallardo et al. 2023). Specifically, the syntax region is just forward of the action-planning one. It is unique to humans, whereas the latter has a homolog in the brains of chimpanzees. It remains possible, nevertheless, that complex forms of action were a pre-adaptation for syntax. For evolution often operates by copying, duplicating, and repurposing genes and the structures they help build (Marcus 2003; Arnon et al. 2025). So it may be that the genes responsible for hierarchical action-planning in general were copied and used to build an adjacent cortical structure with similar computational properties but differing patterns of input and output, and possessing a distinct function—subserving language specifically.

We can conclude that there is an innate brain network that is uniquely developed in humans that underlies

---

production) and forms to meanings (in comprehension). They figure at the periphery of the core language network in the brain, in fact, which extends through much of the upper portion of temporal cortex as far as the temporal poles, as well as a number of regions above and below Broca's area in prefrontal cortex (Fedorenko et al. 2024).

human capacities for language (Fedorenko et al. 2024). It is also clear that humans are unique in being innately motivated to attend to and create communicative actions. Whether the resulting language faculty contains information about human languages beyond the general expectation that sentences are hierarchically structured out of moveable phrases is less certain. And indeed, recent modeling suggests that many of the universal or near-universal properties of the world's languages can be explained in terms of efficiency rather than an innate universal grammar, enabling people to learn and to communicate with language robustly and with minimal effort (Gibson et al. 2019). But at least an expectation of moveable phrases seems to be given at the outset.

4 A Mentalizing Faculty

Another cognitive ability that has generated significant debates between Nativists and Empiricists is the human capacity for attributing mental states to other people (and derivatively to ourselves).[6] It is often referred to as the "mindreading" or "theory of mind" faculty, although the preferred term in the field is now "mentalizing" (Quesque et al. 2024). It has as much right to be regarded as fundamental to human life as has the capacity for language. Indeed, comprehension of speech itself depends on it. Almost all speech interpretation requires pragmatic inferences to recover the intended meaning. This requires detecting the goals of the speaker, what the speaker already knows about the situation and about one's own existing knowledge, and so on. And mentalizing is constantly employed in daily life as one interacts cooperatively or competitively with others. Successful social living depends on it (as witnessed by the social difficulties that many autistic people experience—autism is well-known to be associated with weaker mentalizing abilities).

Debates about mentalizing began with a question about the "theory of mind" abilities of chimpanzees (Premack & Woodruff 1978). But they soon transitioned thereafter into the field of developmental psychology (Wimmer & Perner 1983). There then followed twenty years of intensive experimental investigation. Most researchers in the field embraced an Empiricist position. The view was that mentalizing ability is acquired gradually over the first four or more years of life using domain-general learning abilities that are akin to scientific forms of hypothesis-formation and inference-to-the-best-explanation, albeit partially scaffolded by the mentalizing speech of older children and adults (Wellman

[6] Although introspection-based forms of simulation-theory received some attention in the early days of these debates (Goldman 1989, 2006), they no longer seem to be regarded as serious contenders in the field; nor is the idea that capacities for self-knowledge and other-knowledge might be independent of one another, proposed by Nichols & Stich (2003).

1990; Gopnik & Meltzoff 1997). But a minority view was that mentalizing is subserved by an innate module, with defenders drawing especially on evidence from autism-spectrum disorder, which is a highly-heritable syndrome involving significant mentalizing difficulties (Baron-Cohen et al. 1985; Leslie & Thaiss 1992; Baron-Cohen 1995).

Up until the year 2005 everyone in the field agreed that mentalizing ability develops gradually (whether supported by an innate module that matures slowly or using forms of science-like theory-construction), and it was agreed that competence emerges in stages, with simple forms of goal-directed action being understood first, followed by understanding of the significance of perceptual access, that people have differing desires, then grasp of the difference between knowledge and ignorance, and then finally (around age four) an understanding that people can act on false beliefs and have subjective mental states capable of misrepresenting the world. Failure to represent and reason from the false beliefs of others prior to the age of four seemed an especially robust result, supported by dozens and dozens of studies, as well as being replicated cross-culturally (Wellman et al. 2001). But all of the tests that had been developed up to that time involved children's elicited verbal or other communicative answers to questions (such as pointing to where an agent with a false belief would go to retrieve their goal-object).

Since 2005 the field has been completely transformed, however. For that was the year in which the first expectancy-violation looking-time study was published (Onishi & Baillargeon 2005), suggesting that 15-month-old toddlers could track an agent's false belief and form appropriate behavioral expectations accordingly.[7] Since then there have been dozens of non-verbal studies conducted with infants and toddlers between the ages of six months and three years in which spontaneous responses of one sort or another have seemed to show competence with false-belief reasoning. These experiments have been conducted across a number of different labs and have used a variety of output-measures, including not just expectancy-violation looking-time, but also anticipatory looking, spontaneous helping behavior, interpretation of another's communicative gestures, and motor-related brain activity, among others (Baillargeon et al. 2016; Scott & Baillargeon 2017.)

There have been some failures of replication of some of the early studies, however, which have provided ammunition for those in the Empiricist camp (for examples: Dörrenberg et al., 2018; Kammermeir &

[7] I do not mean to imply that the specific results found using Onishi & Baillargeon's paradigm are robust and replicable. Indeed, I am myself confident that they are not, since the working-memory demands of the experiment seem to me to be too great. Nevertheless, they served to kick off an extremely fruitful body of infancy research.

Paulus, 2018). But Baillargeon et al. (2018) point out the methodological weaknesses of many of these attempted replications, while also acknowledging that some methods (specifically anticipatory looking) might not be reliable. (We will return to discuss the anticipatory-looking paradigm shortly.) But there have also been a number of successful replications, as well as many new studies supporting early false-belief understanding, employing even-more stringent methods and statistical criteria (Király et al. 2018; 2023; Buttelmann & Kovács 2019; Forgács et al. 2019, 2020; Burnside et al. 2020; Kovács et al. 2021; Kampis & Kovács 2022; Schulze & Buttelmann 2022; Woo & Spelke 2022, 2023).

As a result, many in the field agree that *something* real and important about human mentalizing abilities has been discovered post-2005. But they divide into two main camps. One maintains that there are *two* systems for mentalizing. One of these is innate, present in early infancy, and can track some of the mental states of others but without representing them *as such* (Apperly & Butterfill 2009; Butterfill & Apperly 2013). Its outputs have only limited availability to executive systems for reasoning, decision-making, and reporting, which is why young children fail at verbal tasks. This system is held to continue to exist into adulthood largely unchanged, operating alongside a second system. The latter, in contrast, is said to be built slowly through general learning mechanisms, is flexible in its operations, and can guide executive decision-making and speech-production. The contrasting view is that there is just a single mentalizing system with significant innate structure, which is elaborated and becomes more sophisticated through learning and development (Carruthers 2015; Scott & Baillargeon 2017).

Evidence for the existence of two systems is (allegedly) provided by cases in which both young children and adults fail to attribute states with the properties said to be proprietary to the later-developing System 2 (specifically an understanding of the "aspectuality" and subjectivity of perception and belief), while in otherwise-similar circumstances they show sensitivity to states distinctive of System 1. For example, when people are required to respond whether or not they can see two dots on the walls in an image, they are slower and make more errors when there is an avatar in the image who can see just one dot (Samson et al. 2010). It seems that people automatically compute, and are biased by, what is in the line of sight of another agent. (This is said to be the work of System 1.) In contrast, when people are required to indicate whether a numeral they see in an image (a "6" or a "9," say) is odd or even, they *fail* to be biased when the avatar sees the same numeral from the opposite perspective (in which case they would see the "6" as a "9"). They thus fail to automatically take into account the other's differing subjective perspective, which is said to require System 2 (Surtees et al. 2012).

Unfortunately, this and many similar experiments confound mental-state understanding with working-

memory demands. In order to judge that the avatar in such circumstances sees a "6" as a "9" one has to mentally rotate one's own image of the numeral through 180 degrees and identify the result. This is not something that can normally happen automatically, whereas one can take in at a glance whether someone can see one dot or two. (Nevertheless, there is evidence that people will spontaneously represent what others are seeing something *as* when engaged with them in a joint task for a significant amount of time; Elekes et al. 2016, 2017.) No doubt there is a real distinction between mentalizing tasks that are cognitively effortful and those that are effortless. But this does nothing to support the existence of two distinct systems with distinct conceptual resources (Carruthers 2015, 2017). Moreover, many of the other experiments that have been thought to provide evidence of two systems have used anticipatory looking as the output measure (e.g. Low & Watts 2013). But this is the measure that has generated the most replication failures, and is now agreed to be unreliable (Baillargeon et al. 2018; Kampis et al. 2021), for reasons we will discuss shortly.

While some of these debates are ongoing, my own view is that the overall evidence supports a one-system view. There is an initial innate learning system that matures (without learning) during the first few months of infancy. It contains a number of conceptual primitives, including perhaps SEE, INTEND, THINK, and TELL. And it contains a number of simple attribution-rules, such as, SEES X → THINKS X, as well as an action-emulator that can compute an efficient means of achieving a goal once the latter has been attributed to an agent (Jara-Ettinger et al. 2016). The conceptual primitives will be enriched through subsequent learning, however (both by adding new concepts like CURIOUS and by acquiring new information about the states in question and their interactions), and new attribution rules and inference rules will be learned.

At early stages in the development of the system, representing false beliefs can be left implicit in the updating rules for belief. So if an agent has seen an object O placed in location A, and the object is moved to B in their absence, the original attribution THINKS O IS IN A is left unchanged; whereas if the move occurs in the agent's line of sight, the attribution is updated to THINKS O IS IN B. Only later in development is an explicit concept of falsity acquired. (This can be by definition: THINKS FALSELY THAT X = NOT-X & THINKS THAT X.) So there can be a single representation THINKS (or BELIEVES) that remains constant throughout development, whereas the child's understanding of what beliefs are, how they can be formed, and how they interact with other states can be greatly enriched over time.

If there is a single system capable of representing false beliefs from infancy, however, then why is it that children systematically give mistaken reality-based answers in verbal false-belief tasks prior to the age of

about four? A number of explanatory proposals have been offered. (These are mutually consistent with each other. All three might contribute.) One has to do with the executive demands of resisting pre-potent reality-based responses to questions, since executive-function abilities are relatively slow to mature (Scott & Baillargeon 2017). Another is that because successful communication itself depends on mentalizing-based pragmatic abilities, explicit mentalizing tasks might overwhelm the resources of the early-developing mentalizing system, since it then has to tackle a number of distinct mentalizing tasks at once (Carruthers 2013). And a third suggestion in the same vein is more specifically that standard communicative tasks are pragmatically misleading for young children (Westra 2017; Westra & Carruthers 2017). For example, the term "think" is most commonly used in ordinary speech as a mere qualifier of world-directed assertion. ("I think he's in the kitchen" just means, "He's in the kitchen, probably"—one isn't talking about one's own mental states.) So when the child is asked, "Where does Sally think her ball is?" this might be heard as asking, "Where is Sally's ball, probably?" All three of these explanations have significant direct and/or indirect support.

If we grant that humans possess a mentalizing faculty with initial innate structure that is then elaborated and extended through domain-specific learning, as I have suggested here, then the question arises whether this faculty is uniquely human. Field reports of deceptive behavior among other great apes (that is, behavior intended to induce a false belief in another) initially suggested a negative answer (Byrne & Whiten 1988). But then a series of carefully controlled experimental studies with chimpanzees suggested that they could not even represent and reason about the perceptual access of others, let alone represent false beliefs (Povinelli 2000). A break-through was achieved, however, when experimenters began to use competitive situations for their tests (Hare et al. 2000). For example, subordinate chimpanzees will not attempt to retrieve some food when a dominant can also see the food's location, but will do so when the food is obscured from the dominant by a barrier.

Until recently, the consensus in the field was that chimpanzees (and other great apes, as well as at least some other primates) can represent and reason about the perceptual access of a competitor agent (both visual and auditory), can represent and reason about knowledge versus ignorance, and can distinguish between intentional and accidental actions and their consequences; but that they cannot represent beliefs, false beliefs, or the aspectuality of perception (Call & Tomasello 2008). The working-memory demands of many of the tasks in question were quite considerable, however (especially the tests of false-belief understanding). Only recently have experimenters begun to use eye-tracking technology to test whether non-human primates can represent false beliefs, especially using anticipatory looking as an outcome measure, with positive results for great apes (Krupenye et al. 2016; Kano et al. 2019), and even for

macaque monkeys (Hayashi et al. 2020).

How confident can we be in these most recent results, however, given that anticipatory-looking methods have proven so unreliable and hard to replicate as measures of human mentalizing? Answering that question requires us to delve into the reasons *why* anticipatory-looking results have often failed to replicate. One important data-point is the high exclusion rates among both children and adults in these experiments, specifically participants who fail to look in anticipation of an expected outcome during the familiarization (true-belief) trials, and so who are not included in the test trials (Schuwerk et al. 2018). A natural interpretation of this finding is that many participants are either too confident of the outcome-event to bother looking, or are not interested enough in the outcome to re-direct their attention. There is an obvious contrast here with violation-of-expectancy paradigms, given that a surprising event, once it has been observed, always provides an opportunity for learning, and will thereby be of relevance. One might predict, then, that by increasing the importance or relevance of the outcomes in question, one could increase the inclusion-rates for anticipatory-looking paradigms significantly. And indeed, the ManyBabies consortium found just this, by using competitive rather than the usual neutral stimuli, with a bear chasing a mouse to one of two possible destinations (Schuwerk et al. 2022). Despite this, the ManyBabies consortium has reportedly still failed to replicate a key expectancy-looking study of infant understanding of false belief—perhaps (I suggest) because the stimuli were not appraised as relevant *enough*.

In contrast, all of the anticipatory-looking experiments conducted with other primates have used highly engaging competitive stimuli. One of the experiments reported in Krupenye et al. (2016), for example, involved a human agent trying to attack another agent (a human dressed in a gorilla suit) with a stick. The gorilla could hide in one of two locations, about which the human agent could have a true or a false belief. When the human reappeared from behind the door and paused in a central position with his stick raised, most of the apes looked in anticipation towards the location where the human falsely believed the gorilla to be. There is likely a general moral here for tests of mentalizing ability among human infants and toddlers, too. The stimuli used should be made socially meaningful, employing not just competitive situations but also (for humans) ones that might show something about the pro-social or anti-social intentions of the protagonists (Woo et al. 2023).

It is possible, then, that the innate human mentalizing faculty is not uniquely human but shared (at least in part) with other primates. The main difference might be motivational rather than cognitive in nature. As noted earlier, among other primates mentalizing abilities seem only to reveal themselves in competitive situations. Humans, however, are unique in the extent to which they cooperate with and offer social

support to non-relatives. (This point will loom large in subsequent sections.) And there is no doubt that human infants are deeply interested in social stimuli of all sorts (faces, speech, emotions) from an early age. So it may be that, given the same initial starting-state mentalizing faculty, infants' abilities rapidly outstrip those of other primates, to the point that mentalizing becomes almost ubiquitous in human life whenever humans are to be found in one another's company.

It also remains possible, however, that the initial state of the human mentalizing faculty has been enriched relative to our primate ancestors. (It also remains possible that the few studies so far showing positive results with other primates will fail to replicate, of course.) It seems especially likely, in particular, that a conceptual primitive like TELL, combined with some initial attribution rules, might have been added in the course of human evolution. For we noted in Section 3 that humans are unique among primates in freely communicating information to others, and that the drive to communicate is so strong that isolated deaf children set about inventing their own home-sign systems for the purpose.

Moreover, there is evidence that infants as young as three or four months follow someone's eye-gaze more quickly and reliably when it is preceded by speech as opposed to non-speech sounds, and that speech (again, as opposed to non-speech sounds) facilitates their acquisition of novel categories (Ferry et al. 2010; Marno et al. 2015). There is also evidence that infants as young as six months understand the communicative function of speech. For example, suppose an agent has displayed a preference for one of two objects, but can now no longer reach them. Infants look longer (seemingly in surprise) when an observer hands over the non-preferred object after the agent has looked at her and uttered something that sounds like a word ("koba"), in contrast with the case where the agent merely coughs (Vouloumanos et al. 2014; Yamashiro & Vouloumanos 2018). This suggests that infants in the former circumstances understood that the agent had communicated to the observer what she wanted. Indeed, even newborns are sensitive to turn-taking speech, in contrast with a single speaker repeating the same sounds or two speakers echoing one another with the same sounds (Forgács et al. 2022). Taken all together, the evidence suggests that TELL is likely be a uniquely-human conceptual primitive.

Before concluding this section we should consider one other feature that has been claimed as an innate component of the mentalizing faculty. This is a commitment to an ontological separation between mind and body (Bloom 2004). There is no doubt that such beliefs are a human universal.[8] All hunter-gatherer

[8] As with other human universals (Brown 1991), this doesn't mean that every individual human has believed in the separation of mind and body and the possibility of life after death. Rather, these beliefs are universal in the same

groups are characterized by *animism*, or the belief that the world is imbued with spirit-like forces (Peoples at al. 2016). And until very recently, with the advent of modern science, all humans in all cultures have believed that the minds and bodies are distinct existences (Boyer 2001; Cohen et al. 2011; Roazzi et al. 2013). They have thought that mental properties can be tokened independently of physical ones, and have believed that the self is a distinct thing, independent of the body. Even among people who explicitly reject them, dualist beliefs are apt to emerge indirectly in their intuitive judgments (Chudek et al. 2018), or to manifest themselves when people are placed under cognitive load (Forstmann & Burgmer 2015). As a result, belief in some sort of afterlife has been (and continues to be) common, and is found in around eighty percent of hunter-gatherer societies (Peoples et al. 2016). The afterlife can be merely spiritual in form, or involve resurrection of one's original body following a period of attenuated mental existence, or reincarnation into a distinct body.

My own view is that dualism is not directly innate, but is rather a very natural by-product of a clash or disconnect between our innate physics system and the innate mentalizing system (Carruthers 2020). Everyone knows that mentality is at least *causally* tied to the physical, of course. We know that we see things by opening our eyes, that hearing depends on the impact of sound on our ears, and that felt touch requires physical contact with our bodies. Moreover, we know that we can effect change in the physical world by deciding to move our limbs. But interactions among our mental states themselves are *sui generis*, conforming to none of our familiar models of physical causality. Paradigm cases of everyday physical causation are mechanical, involving pushing, pulling, and motion in space. But nothing remotely like that holds for mental-to-mental causal relations. When a perceptual state sparks a memory, or a thought makes one sad, these causal relations strike us as *immediate*, and certainly not as involving anything like mechanical contact.

Even more importantly, our mentalizing system seems not to require that mental states should occupy

---

sense that mentalizing itself and sensitivity to rhythm are universal: they are possessed by almost all individuals across all or almost all cultures. Note, too, that the separation in question is ontological, not functional. Barlev & Shtulman (2021) go wrong on just this point when trying to develop an argument against any sort of inherent dualism. Of course everyone (even a little infant) knows that mental and physical states interact with one another causally and in rich ways. The real point is that intuitive psychology doesn't conceptualize causal relations among mental states themselves in physical or mechanical terms, and isn't committed to physical locations for those states, as we will see. Moreover, since all the modes of interaction with the physical world recognized by intuitive psychology involve the body, it is hardly surprising that many beliefs about the spirit-world involve agents with body-like properties, such as capacities to see and to hear.

positions in space, which is the defining feature of physical objects and physical processes. Indeed, it comes close to entailing that mental states *don't* occupy spatial positions (McGinn 1991). For example, statements such as, "My thought about my mother is two inches behind my right eye" seem hardly even to make sense. The physical subject or bearer of mental states is believed to be spatially located, of course. My thought about my mother is located where *I* am located. And most people in scientific cultures know enough about the brain to realize that it has a special role to play. But even this is apt to be expressed in causal rather than constitutive terms. ("Something happening in my brain caused me to think it" rather than, "An activity of a specific physical network in my brain *is* me thinking it.")

It seems, then, that there is a clash between our core knowledge of physics and our core knowledge of psychology, which gets set up in terms of a contrast between a world of space-occupying objects, events, and causes, on the one hand, and a set of apparently *non*-spatial mental states and processes, on the other. This makes it entirely natural to think that there is a deep ontological separation between them, and will support afterlife-beliefs in cultures that articulate them. Such beliefs might be expected to exert a deep attractor effect on cultural evolution, being sustained and transmitted both because of their apparent *naturalness* given the underlying core-knowledge clash, and because of their role in terror-management, enabling people to become more reconciled to their own eventual deaths (Atran 2002).

5 A Cultural Creature

There is a long tradition in philosophy of characterizing what is distinctive about human nature in terms of reason and rationality, dating back to at least Aristotle (c.350 BCE). "Man," it is said, "is the rational animal." There is obviously an element of truth in this. Humans are unique in the extent to which they make long-term plans and can (and sometimes do) consider a wide range of alternatives to any given choice that they make (Sripada 2016). But the claim is also deeply misleading. For what enabled *Homo Sapiens*—uniquely among primates—to flourish and succeed in such a wide range of differing ecologies around the world (savannas, rain forests, temperate forests, deserts, swamps, mountains, and even the arctic) is not so much how smart we are, but rather the fact that we are adapted for cultural learning and cultural living (Henrich 2016; Boyd 2018). Moreover, much of what people count as rationality in the large-scale civilizations (such as Aristotle's Ancient Greece) that grew up following the introduction of farming around 10,000 years ago is itself culturally constructed and culturally acquired through learning (Henrich & Muthukrishna 2024). Indeed, farming, too, appears not to have been a matter of relatively fast *invention*, but rather a very gradual process of cultural evolution that happened in fits and starts in different locations (initially in the middle East) over the course of millennia, co-evolving with changes in the plants and animals that were gradually domesticated (Allaby et al. 2017; Fuller & Lucas 2025;

Spengler 2025). It is better to say that human nature is culture-enabling nature. The remainder of this section will elaborate on this point and the evidence that supports it, before later sections consider its ramifications in three specific domains.

It needs to be stressed at the outset, however, that culture in the broadest sense—that is, social learning of novel behavior—is by no means unique to humans. On the contrary, it has been observed across multiple species and taxa, from bees, to birds, to orcas, elephants, many primate species, and especially chimpanzees (Whitehead et al. 2019; Whiten 2021). Many species show some of the same social-learning biases that are found among humans, too, such as differentially copying the behavior of the majority, or copying the behavior that has the greatest perceived success or payoff. So for sure there is evolutionary continuity here. But all human cultures are many orders of magnitude richer than those found in other animals. While chimpanzees, for example, have been found to have a few dozen socially-transmitted behaviors, these are mostly fairly simple in nature, ranging from methods of grooming, to use of leaves as sponges, to nut-cracking with a stone and anvil, to termite fishing with sequential use of a stout puncturing stick and a termite-extracting frilly wand. So rather than say that human nature is distinctively culture-enabling, it might be better to say that it is *maximally* culture-enabling, or *fundamentally* culture-enabling, or something of the sort.

A combination of anthropological, archaeological, and formal-modelling evidence shows that cultural accumulation and improvement were originally slow processes (albeit fast when compared with the speed of evolution by natural selection), and were heavily dependent on happenstance (as is natural selection itself, of course). It also required critical levels of population density, needing a sufficient number of people who possessed the technical skills embodied in the culture to serve as targets for learning, compensating over generations of learners for copying errors made by the people who learned from them (Kline & Boyd 2010). One dramatic example (among many) is provided by what happened on the island of Tasmania.

Humans probably first reached Tasmania by walking across a land-bridge from mainland Australia more than 30,000 years ago. But then about 10,000 years ago the island was cut off by rising sea levels, and the inhabitants lost all contact with other indigenous groups. A combination of direct archaeological evidence and inference from what is known about their ancestors on the mainland demonstrates that at the time of their isolation the people of Tasmania had an elaborate and sophisticated material technology. This included a wide range of bone tools, cold-weather clothing, barbed spears, fish traps, nets, spear-throwers, and boomerangs. But by the time the first European explorers arrived thousands of years later, all this had

been lost. To hunt and fight, Tasmanian men were found to use only one-piece spears, rocks, and clubs (Jones 1995). (Nevertheless, many other low-skill cultural practices and beliefs, including kinship systems, religious beliefs, dietary traditions, and so on remained richly elaborated.) How could this happen? For these later Tasmanian people were presumably no less intelligent than their forebears or their distant cousins in Australia.

Drawing on the modeling framework developed by Boyd & Richerson (1985), Henrich (2004) shows how this could (and probably did) happen. The model assumes that people tend to learn from, and acquire skills from, the most prestigious individuals in the group. It also assumes that the copying process is imperfect. Sometimes, through accidental discovery or individual experimentation, the products made by a learner are an improvement on those modeled by the teacher. But more often what gets produced by the learner is somewhat worse. Reasonable, empirically-informed, assumptions about the error-rates involved can be used to show how cultures can gradually improve their technological tool-set, thereby becoming more adapted to the environments they find themselves in, and explaining how humans can be successful in such a wide range of ecologies around the globe. But this process is critically dependent on population size and the extent of social interaction and inter-dependence within the culture. In sparse populations that interact less commonly, the result can be gradual *loss* of technological knowledge and skills, which is probably what happened in Tasmania and other places (such as North-West Greenland) where similar phenomena have occurred.

A number of factors suggest that humans are adapted for social learning, in fact, just as this account suggests. Not only are humans biased towards copying the majority and copying successful behaviors (as are many other animals), but they are also biased towards imitating those who are socially prestigious. This latter motivation appears to be innate rather than strategic, since it operates indiscriminately (Henrich & Gil-White 2001). Thus people are disposed to wear what famous athletes they admire wear or recommend, they adopt the mannerisms of venerated teachers, and so on. This makes sense from an evolutionary perspective, because so much of human technology and of human cultural practices are opaque to their users. When someone in the community is a prestigious hunter, for example, it can be hard to see what it is that makes him especially successful. Is it the way he prepares before the hunt? The particular way in which he strings his bow? The way he carries himself when stalking prey? What he wears when hunting? Or what? And the hunter himself may be unaware of some of the sources of his success, too.

Humans seem also to be innately equipped for what is called “natural pedagogy” (Csibra & Gergely

2011). There are signals of communicative intent that are used across cultures and that are responded to even by very young infants, facilitating learning (Senju & Csibra 2008; Hewlett & Roulette 2016; Boyette & Hewlett 2018). These include intentional eye-contact and use of the infant's name to attract attention. And young children across very different cultures who receive such signals prior to a skill being modelled for them (such as how to open a box) will copy all components of the adult's actions, even those that are causally irrelevant (Nielsen & Tomaselli 2010). Chimpanzees and children who do not receive ostensive signals, in contrast, just copy the components that matter. This makes good sense, of course, given that the functionality of much of what children have to learn about a culture is opaque, even to experts (Derex et al. 2019; Harris et al. 2021). Adults also provide opportunities for observational learning and tend to modify their skilled actions to increase an observer's learning of the component processes—such as by slowing down or exaggerating the movements involved (Lew-Levy et al. 2017).

Moreover, humans are the only creatures known to systematically practice the skills that constitute and support their society's culture and cultural products, taking skill-acquisition as an explicit goal. Many animals are capable of fine-grained forms of motor control, of course. (Think, here, of a squirrel leaping from branch to branch high in the canopy, or a bird that lands smoothly on its perch despite being buffeted by the wind.) And many of these behaviors improve with practice during the animal's initial development. Animals can also be trained to *acquire* specific abilities, of course. (Think of a seal trained to balance a ball on its nose or a dog trained to walk on its hind legs.) But none is capable of the range of skills that humans can acquire and flexibly deploy, enabled by uniquely extensive projections from human motor cortex to sub-cortical motor neurons in the medulla and spinal cord (Striedter 2005). Nor do any animals take skill-acquisition itself as an intrinsic goal, repeatedly practicing and making adjustments by observing their own performance.

While many animals are biased towards copying the behavior of the majority in specific domains (such as mate-choice or foraging behavior), humans are apt to be indiscriminate in copying the practices of the majority—a so-called "conformity bias." This, too, makes good evolutionary sense. For human cultures are, in general, well-adapted to their environments. So what the majority of people do is likely to work. Moreover, cultural members are apt to look askance at those who fail to conform, especially in small-scale hunter-gatherer communities, which tend to be highly uniform in everything from modes of dress, to the procedures used in food preparation, to ways of handling and using technology. Indeed, people in general seem to assume by default that what is commonly done is what *ought* to be done, following a "common-is-moral" learning heuristic (Lindström et al. 2018).

As we briefly noted earlier, too, many of the reasoning skills that are so prized and emphasized by philosophers are themselves cultural products acquired through cultural imitation and (now) explicit forms of intellectual training (Henrich & Muthukrishna 2024). While some rational capacities appear to be innate, and are present in pre-linguistic infants (Cesana-Arlotti et al. 2018, 2020), at least some of these are shared with other animals, and so are not distinctively human (Ferrigno et al. 2021). But many others—especially those involved in so-called "System 2" reasoning—are culturally acquired (Kahneman 2011; Henrich et al. 2023). And even if Mercier & Sperber (2017) are right in claiming that humans possess an innately channeled "adaptation for argumentation," this is best seen as an adaptation for cultural living, rather than as having evolved to improve individual cognition or individual decision-making.

Behavioral and cognitive skills are by no means the full extent of what humans need to learn, of course. On the contrary, it is just as important for individual fitness for people to learn and internalize the norms and values of the surrounding culture. (This point will loom especially large in Section 6.) All human societies are imbued throughout with *norms*—that is, things that one must or must not do. And in the kinds of small-scale hunter-gatherer communities in which humans lived from at least 200,000 years ago until the emergence of agriculture a mere 10,000 years ago, there would have been little or no internal variation in the norms that govern each group. While there would have been separate norms governing the behavior of men and women, those norms would have been known and endorsed by everyone in the group, even when they only apply to others. And people would be fully prepared to enforce them, through gossip, loss of reputation, and withdrawal of cooperation in the first instance, and ultimately through violence or exile from the community. A crucial adaptive problem for an individual growing up in such a society, then, is not only to learn what the prevailing norms *are*, but to *internalize* them—coming to value compliance with them (both for oneself and others) for its own sake. For merely strategic compliance and enforcement is likely to be recognized as inauthentic, as well as leading to a greater number of failures to comply when people are faced with conflicting motivations.

The upshot is what some have called an innate "norm psychology" (Sripada & Stich 2006; Chudek & Henrich 2011; House et al. 2020). This is a faculty for identifying and learning the norms of the group, for storing them and then accessing them in appropriate circumstances, and for creating intrinsic motivation to comply with them and to punish those who fail to comply with them. Norm psychology likely co-evolved with the emergence of widespread cooperation in human societies, much of which takes place with unrelated individuals. All extant human societies depend on such cooperation, and probably always have. This is made possible, in part, by human norm-psychology and the behavior that it supports,

ensuring that cheaters and free-riders are identified and punished, and providing individuals with the intrinsic motivation to inflict such punishment, as well as to act in accordance with the group's norms in the face of temptations to avoid doing so.

Once norm psychology became established, involving both informal punishment and meta-punishment (that is, punishment of those who fail to punish breaches of societal norms), then social norms could proliferate and be sustained every which-way (Boyd & Richerson 1992). The result is what we now see across cultures. While there are some commonalities, of course (norms against murder, theft, and so on), otherwise what we see is immense variation in required modes dress, food-preparation, religious belief, ritual, and much, much, more. Even objectively harmful norms like child-sacrifice and female genital cutting can be sustained—although practices that make a culture more successful overall can gradually proliferate through processes of cultural competition and cultural evolution (Henrich 2016).

If humans are fundamentally cultural creatures, as I am suggesting, then the innate capacities for language and mentalizing discussed in Sections 3 and 4 may best be viewed as belonging within the same framework. For language is, of course, one of the primary means through which information is transmitted within a culture. Much of what humans learn about the world and the people around them, as well as about the culture they inhabit, is acquired through language. And language is vitally involved in all but the simplest forms of group cooperation, coordination, and planning. Likewise, mentalizing is vital, not only for the pragmatics of linguistic communication itself, but also for tracking the intentions of others and evaluating their actions and status in the community.

The claim that human nature is fundamentally culture-enabling nature should not be taken to exclude contributions from distinctively-human general intelligence, of course. And there is no doubt that this, too, has been selected for (Cantlon & Piantadosi 2024). Individual humans engage in many forms of complex problem-solving. Examples include mediating social disputes, planning trap placement, and reading and interpreting animal spore when hunting. Indeed, the reasoning processes displayed by skilled trackers involve a form of inference-to-the-best-explanation, and have much in common with core kinds of scientific reasoning (Liebenberg 2013)—except that the latter are directed at generalities, whereas the former are mostly aimed at particulars, such as whether the animal that made a specific set of tracks is injured or not.

If human nature has been heavily adapted for culture and cultural learning, then we might expect to see a number of further evolved adaptations for cultural living in addition to language, mentalizing capacities,

and the learning biases discussed in this section. The remaining three sections of this Element explore the evidence for innate systems in humans for social acquisition of novel values, for social-group cognition, and for social-group-directed motivations.

6 Evaluative Learning

This section will suggest that humans possess a domain-general adaptation for evaluative learning, which operates on top of the domain-specific systems discussed in Section 2. (For more detailed elaboration and support, see Carruthers 2025.) It facilitates acquiring the evaluative components of the local culture (what foods to like, what forms of dress and body-decoration to admire, what kinds of dance and music to enjoy, as well as what kinds of normative behavior one should feel intrinsically motivated to comply with and enforce). As we noted in Section 5, small-scale and hunter-gatherer societies tend to be highly uniform internally in the things and modes of behavior that they value and disvalue. Some of these values represent the collective wisdom of the society in question, culturally adapted to a given ecology (what things are good to eat, what things are dangerous, and so on). But many will have grown up over time in ways similar to the growth of material technology and will either have become norms in their own right, or will at least be expected and assumed by the community. This would have created intense selection pressure for swift across-the-board evaluative learning. The resulting domain-general mechanism, I suggest, is a greatly-strengthened influence of belief and expectation on affective evaluation.

The existence of placebo effects on various forms of illness has been known about for centuries. (Thomas Jefferson's doctor is reported to have remarked that he had prescribed more sugar-pills than real medicines over the course of his career.) And such effects can be remarkably powerful—by some estimates accounting for around 50 percent of the benefit provided by many established medications and procedures. Placebos are now known to have their strongest effects on affective forms of illness, including chronic pain, depression, and anxiety disorders (Ashar et al. 2017; Petrie & Rief 2019). Indeed, it appears that their benefits for other kinds of illness come via the positive impacts they have on pain, stress, and anxiety, thus reducing inflammation and speeding healing (Liu et al. 2017). In fact, meta-analyses of existing studies suggest that as much as 80 percent of the benefit provided by either drug-treatments or psychotherapy for affective illnesses is accounted for by a combination of placebo effect and spontaneous recovery (Cuijpers et al. 2012; Khan et al. 2012).

Placebos work by creating an *expectation* (or partial expectation) of improvement, especially when accompanied by the rituals associated with medical healing. (Since clinical trials with a placebo-control require participants to be informed that there is only a 50 percent chance that they will be assigned to the

test-medicine condition, participants should not believe outright that they are taking a real medicine.) But those expectations have direct and powerful effects, not just on people's reports of their affective state, but also on the neural networks that underlie reward and punishment (Ashar et al. 2017). So placebos provide an instance of powerful top-down effects of belief (or partial belief) on affective experience. (As do nocebos, where expectations of a bad outcome induce one to feel worse.) Indeed, Ashar et al. (2022) found that encouraging patients with chronic back pain of no known biological origin to reconceptualize their pain as a non-dangerous brain-produced false-alarm—rather than as resulting from peripheral tissue damage—produced very large reductions in the amount of pain felt (compared to both open-label placebo treatments and treatment-as-usual). And these effects were still large (albeit somewhat reduced) at the time of a one-year follow-up.

Placebo and nocebo effects on pain, depression, and anxiety are now known to be just two of the ways in which beliefs and expectations can impact affective experience. Thus expecting a neutral odor (or even clean air) to smell like body odor makes it seem unpleasant (de Araujo et al. 2005), expecting something to taste good makes it taste better (Grabenhorst et al. 2008), expecting a touch to feel pleasant makes it more so (McCabe et al. 2008), and expecting a wine to taste better (as indicated by its price) makes its taste more enjoyable (Plassmann et al. 2008; Schmidt et al. 2017). Moreover, evaluative statements about a novel group of people (e.g. "Squarefaces are good, Thinfaces are bad") are sufficient to induce new evaluative attitudes in children (as measured by the implicit attitudes test), whereas associative pairing with positively or negatively valenced items are not (Charlesworth et al. 2020). It appears that the mechanisms via which expectations impact evaluative experience are domain-general, operating across all forms of affective experience.

There is one other massively-replicated finding that is best explained as a placebo-like effect of expectation on value. This is the choice effect. People who are forced to choose between two equally-rated options thereafter shift their evaluations of those options accordingly—liking the chosen item more strongly and/or liking the rejected item less (Brehm 1956; Enisman et al. 2021). And just as with the expectation-based effects reviewed above, these effects aren't merely behavioral in nature, but also involve changes in underlying reward networks (Sharot et al. 2009). A range of explanations have been offered over the years for the choice effect, including post-choice motivated reasoning, biased reflection on the properties of the two options, or as resulting from an attempt to protect one's sense of self-integrity or self-esteem. But these explanations appear to be ruled out by a number of recent findings.

The choice effect continues to operate in people with severe amnesia (who won't remember their earlier

choices for more than a minute or two) and in people who are placed under cognitive loads that prevent reflection (Lieberman et al. 2001). The effect can also be found in preschool children, who are unlikely to engage in post-choice reasoning or reflection (Egan et al. 2007). Moreover, the effect is just as powerful when people are tricked into *believing* that they have chosen one thing over another (without really having done so) or when they choose blindly, in ignorance of the identity of the two options (Egan et al. 2010; Sharot et al. 2010). Even more striking, the effect of believing one has made a choice on subsequent evaluation is found among human infants in their second year of life (Silver et al. 2020), and the effect of blind choosing has been established early in the third year, and turns out to be unrelated to capacities for self-identification (Wiesmann et al. 2022).

Given these findings, I suggest that the best (albeit currently untested) explanation of the choice effect is as follows. One's mentalizing system takes as input one's own observed behavior (as it always does; Carruthers 2011), and deploys the principle, *if S chooses A over B then S thinks A is better than B*, or perhaps, *if S chooses A over B then S prefers A to B*. The latter is known in economics as the principle of "revealed preference," and plausibly plays an important role in observational value-learning—observing another agent choose one thing over another, one infers that the chosen option is likely to be the better or more desirable of the two. So when an adult, child, or infant chooses one thing over another they form the belief that the former is better than the latter from tacit interpretation of their own behavior, and that belief then exerts a top-down influence on the agent's evaluation of the two options. We know that infants can engage in this sort of simple mentalizing, and we know that the mentalizing system continues to operate in the first person, so this explanation makes good theoretical sense.

It should be stressed that these top-down effects of expectation on value are not just ephemeral effects on current affective experience. On the contrary, they are apt to give rise to long-term evaluative change. Thus the effects of choice on value are still discernable three years later (Sharot et al. 2012), and one's stored affective evaluation of a (fictional) group of people induced by reading a single short narrative can survive many rounds of counter-conditioning (Gregg et al. 2006). Indeed, we have known for many years from the animal conditioning literature that novel values, once acquired, tend to be permanent or semi-permanent (Quirk & Mueller 2008). For the phenomenon known as the "extinction" of a secondary (learned) reward-value is best understood as a form of contextual learning rather than as a loss of that value. For the original valuation can return immediately once the secondary reward-stimulus is paired with a primary reward once again. It doesn't need to be re-learned. During extinction the animal has learned that the stimulus object is not valuable *in the current context* or *at the current time*, rather than ceasing to value it altogether.

There is no evidence of top-down induced (as opposed to conditioned) placebo and nocebo effects in animals, although the issue has not been much investigated (Stewart-Williams & Podd 2004; Meeuwis et al. 2020). So top-down influences of expectation on value may well be uniquely human. And for sure they have become massively expanded in frequency and power among humans. This gives rise to something of a puzzle. For if, as I have argued elsewhere (Carruthers 2024), one of the primary outputs of affective valuation (namely positive or negative valence) is a *representation* of the current, contextually-modulated, value of the thing or event being evaluated, then one would think that there would have been intense selection pressure to secure *veridical* representations of value. While there are top-down effects of belief and expectation on visual representations, for example (Ogilvie & Carruthers 2016), these tend to be tiny in comparison; and as a result, vision tends to be highly reliable.

A number of points can be made in resolution of this puzzle. One is that the true value of a thing or a behavior is often opaque, especially in the short term or in one's own initial experience. Many foods that are actually highly nutritious aren't immediately attractive when first tasted, for example. And the benefits of many behaviors may only become apparent in the long term, if they are noticeable at all. So it makes sense that human value-learning mechanisms should give considerable weight to the culturally-evolved and established values of members of the surrounding community. So when one sees others choosing something, one should believe that it is likely to be valuable, and one should come to value it accordingly. Likewise when community members tell one that something is good, or that an action is good or bad to do. The resulting shifts in value provide one with a head-start on acquiring the values of the community, which can then be reinforced and strengthened over time through normal reward-based learning (especially through the approval or disapproval of others)—likely involving an iterative feedback loop that further strengthens one's beliefs about value.

Moreover, given widespread expectations of behavioral and evaluative conformity in the small-scale and hunter-gatherer communities in which humans evolved, it would have been adaptive to swiftly acquire and internalize the values of one's group. (Even infants in the first year of life expect members of social groups to behave alike and make similar choices; Powell & Spelke 2013. And 14-month-old infants expect agents who belong to the same group to share the same food preferences; Liberman et al. 2016.) Failure to do so would likely have resulted in ridicule or ostracization, and if one attempted to merely mimic the behaviors and choices of others without sharing their values, this would likely have been recognized as inauthentic, as well as requiring continual vigilance to insure compliance.

These points can be further strengthened when one recalls the central place of norms in all human cultures. Human societies are richly imbued with norms that govern people's behavior. There are things that one must do, on penalty of social punishments of various sorts (primarily loss of reputation), and there are things that one must *not* do, or else face ridicule, ostracization, and loss of potential cooperative partners (Boehm, 2001; Wiessner, 2005). As we noted in Section 5, the evidence suggests that these powerful selection pressures led to the evolution of so-called "norm psychology." This enables one to swiftly identify the norms of the surrounding society and to internalize the values that insure norm-compliant behavior (or rather, that make such behavior more likely in the face of competing motivations). Since one can operationalize things one *must* do as things it would be bad *not* to do, and things one must *not* do as things it would be bad *to do*, learning a new norm initially involves forming new beliefs about value. As a result, one's subsequent internalization—or affective valuing—of the norm can draw on the very same top-down mechanisms that underly placebo and nocebo effects generally.

I have suggested that humans have an adaptive innate system that uses beliefs and expectations about value to shift one's affective experiences and stored affective valuations accordingly. This system appears to have emerged at some point during the course of hominin evolution. The mechanism in question is a domain-general one, however—at least in comparison to the set of specialized emotional and homeostatic affective systems noted in Section 2—since it operates across all forms of affective valuation. So it is perhaps unclear whether or not it qualifies as Nativist or Empiricist, when these are understood in the terms discussed in Section 1. For although it is a learning system that is specific to the evaluative domain, that domain itself is extremely broad. Actual Empiricists don't mention it or claim it as their own, however. And so far as I am aware it has not been previously discussed in the context of debates between Nativists and Empiricists. Perhaps the question of classification doesn't matter very much. The important point to note is just that we have identified a key component of our innately-channeled human nature.

What *would* be worrisome, from a Nativist perspective, is if it could be shown that the top-down effects of expectation on value are a by-product of our evolved general intelligence of the sort that all Empiricists are already committed to, or are perhaps a by-product of the overall expansion of the human neocortex, and especially the greatly-expanded prefrontal cortex that houses our executive-function abilities. If either of these things were true, however, then one might expect that individual differences in intelligence (IQ) or executive control abilities would correlate with the strength of placebo effects. I am aware of no evidence that this is so. Indeed, although the literature is somewhat tangled, it seems that the strength of people's placebo responding either correlates or anti-correlates with a range of personality variables, including optimism, need for cognition (thoughtfulness), pleasure seeking, and (negatively)

somatosensory bodily awareness (Geers et al. 2007; Plassmann & Weber 2015; Kern et al. 2020). So there is no reason to think that the top-down effects of expectation on value are a by-product of things that Empiricists already accept.

7 Tribal Thinking

Sections 5 and 6 have developed the case for maintaining that humans are cultural creatures, adapted for cultural learning and cultural living. But humans are also *tribal* animals. For the vast majority of our evolutionary history we lived in ethnic groups within which humans cooperated and sought mates, marked especially by language or dialect. And a combination of archaeological and anthropological evidence enables us to approximate what tribal life was like, at least for the last 70,000 years, and probably for a great deal longer. (Humans first evolved as a separate lineage in Africa sometime between 200,000 and 300,000 years ago.) Moreover, recent evidence suggests that most of the genetic changes in human populations that have happened in the last 10,000 years are related to adaptations to the dietary shifts that followed the emergence of farming, and to changes in human immune systems driven by population expansion and the existence of settled communities and cities (Mathieson et al. 2015; Kerner et al. 2023). Human adaptations of the sort discussed in this Element must therefore have evolved prior to that, when all humans lived as hunter-gatherers.

The most direct evidence comes from Australia, where the indigenous inhabitants would have had no previous contact with sedentary farming cultures—direct or indirect—prior to the arrival of European anthropologists. There were about 600 different tribes living in Australia when Europeans first arrived, speaking at least 250 different languages as well as many further dialects (Flood 2019). But much can also be learned from anthropological reports from other isolated, or relatively isolated, populations combined with archaeological evidence from earlier eras (Marlowe 2005). Tribes in general were marked by language or dialect, modes of dress and body-decoration, and were territorial in nature.

In Australia, as elsewhere, hunter-gatherer tribes had three levels of organization. There was the tribe itself, comprised, on average, of around 1,000 adults. They may or may not have gathered together as a whole periodically (e.g. for ceremonial purposes or for collective activities such as the construction of a weir or a communal hunt; Boyd & Richerson 2022). Then there were traveling bands of around 30 adults, who camped together, shared food together, and shared child-care duties. The composition of these bands changed fairly frequently, as people moved to visit with relatives, or as a result of personal animosities within the group. (Evidence suggests that hunter-gatherers might have interacted with over 300 different same-sex same-tribe adults over the course of a lifetime; Hill et al. 2014.) And finally, there were smaller

groups of around five adults who foraged together during the day, bringing most of what they acquired back to a central camp to be cooked and consumed. Children who were too heavy to carry but not mobile enough to participate in foraging trips remained in the camp, overseen and protected by some of the adults.

Importantly for our purposes, attitudes of tribal members towards neighboring tribes were at best wary and at worst outright hostile. Archeological evidence of transport over long distances of materials such as flint, shells, or mammoth tusks—sometimes very long distances, probably exceeding the home-range of any given tribe—suggests the existence of inter-tribal trading networks (Khatsenovich et al. 2020; Golovanova et al. 2021; Boyd & Richerson 2022). Whether or not trading was conducted by groups that specialized in the practice (somewhat like the Roma—"gypsies"—in Medieval Europe) or happened near the borders of adjoining tribal home-ranges, this suggests that interactions with out-of-tribe individuals could be tolerated in the right circumstances.

Despite occasional contacts, inter-tribe warfare was common, often comprising sneak raids (generally at night) from which the attackers themselves rarely received injuries, but sometimes involving larger-scale pitched battles in which many warriors participated (Allen & Jones 2014; Boyd & Richerson 2022). Sometimes the killing was indiscriminate, sometimes focused mostly on males, and women were sometimes kidnapped to serve as wives for some of the attackers. Rates of violent death were high in most ancestral hunter-gather communities, occasionally resulting from intra-tribal feuds and jealousies, but mostly from inter-tribal attacks, either of isolated individuals, or through sneak raiding. Rates of violent death were around 160 per 100,000 individuals each year on average (or 0.16 percent of the population), *much* higher than the homicide rate in the United States, which is around 6 per 100,000, or 0.006 percent (Gat 2015). Indeed, it may have been these high violent-death rates that prevented local human populations from expanding very much prior to the emergence of agriculture some 10,000 years ago. The traditional conception of hunger-gatherers as "noble [largely peaceful] savages" is a myth (Pinker 2011).

Given that this is the context in which humans evolved, we might predict that our minds would be pre-prepared to learn, not just about our own culture (our own tribal group), but about the properties of surrounding tribes as well. One should rapidly learn to identify someone as a member of another tribe, as well as to which tribe they belong. And one should be ready to draw inferences about out-group individuals that are specific to the tribe in question—especially whether they constitute an immediate threat, under what conditions they might become a threat, the weapons and materials they have available

to them, and more. As we will see, these predictions seem to be borne out by a range of findings from modern-day populations.

We have already noted that human infants seem pre-prepared to think in terms of social groups, one cue to which is language (even at ages when they know very little about their own native language). For example, 9-month-old infants expect speakers of the same language to affiliate, and are surprised when speakers of different languages do (Liberman et al. 2017); and at 14 months they identify in-group versus out-group individuals by the languages they speak (Buttelmann et al. 2013). Moreover, 8-month-old infants identify animated shapes (which infants are well-known to see as agentive) as belonging to the same group on the basis of previous joint engagement in a synchronous activity, whether or not they otherwise have a similar appearance (Powell & Spelke 2013). And other labs, too, have found that infants sort agents into groups on the basis of a history of previous joint activity, cooperation, or affiliative behavior—at 9 months (Pun et al. 2021), at 14 months (Liberman et al. 2016), or at 16 months (Rhodes et al. 2015). Furthermore, by the age of 12 months, infants will sort people into groups on the basis of the distinctive and unusual clothing they wear (Ting et al. 2019), provided the clothing doesn't serve an obvious instrumental purpose (Bian & Baillargeon 2022).

Since infants at these early ages are unlikely to have had any experience of social groups as such, it is reasonable to conclude that there exists an innate domain-specific learning mechanism designed to identify social groups (originally tribes), built around conceptual primitives like LANGUAGE, JOINT-ACTION, AFFILIATION, and APPEARANCE / CLOTHING. The same mechanism seems also to encode some initial expectations concerning the social significance of group membership. We have seen that infants expect members of the same group to share food preferences (Liberman et al. 2019) and to share the same behaviors (Powell & Spelke 2013). They also expect that group members will help one another, but not members of other groups (Jin & Baillargeon 2017). And they expect that an agent belonging to one group whose members have previously fought with a member of another (wrestling over opening versus closing a box) will behave similarly, also fighting rather than cooperating with a member of the out-group (Rhodes et al. 2015). They thus seem to expect that conflict is a property of groups, and not just of individuals.

Even more striking, 12-month-old infants expect a member of one group to withhold support from (to "punish") an out-group member who had previously harmed another member of the agent's own group; whereas they expect an agent to help another member of their group who had previously harmed an out-group member (Ting et al. 2019). And at roughly the same age, infants expect an agent to help an in-

group member who is in conflict with a member of another group (bumping up against one another trying to cross a bridge in opposite directions), and are surprised when the agent helps the out-group member instead (Pun et al. 2021). It seems likely that infants' expectations of in-group loyalty in situations of inter-group conflict are innate. For at this early age they will have had little or no opportunities to learn about such properties of groups (or even that there are such things as social groups, come to that).

As one might expect, then, given our extensive evolutionary history of tribal living and inter-tribe hostility, humans appear to have an innately-channeled system for identifying and learning about other social groups. In the modern world—at least among developed nations—there are no such things as tribes, of course. But it seems that the social-group system continues to be cued into activity by indications of various kinds of group interaction and group alliances, and is apt to have a significant influence on people's behavior in contemporary societies. For there are, of course, still ethnic groups marked by language (e.g. Arabic) or dialect (e.g. African-American), as well as groups marked by cultural belief (e.g. religion), or by alliances of various kinds (political, sporting, and so on). In the modern world, many of these groupings—often described using the language of "identity"—cross-cut one another, so people will generally belong to many different groups (and will thus have "intersectional identities"), shifting among them as the occasion requires (church, political rally, football game, and so on).

The upshot of the tribal-group system operating in the modern world is an extensive set of group stereotypes. ("Asians are smart," "Arabs are terrorists," "Black men are violent," "White people are racists," and so on.)[9] These are known by almost everyone in the population, including members of the stereotyped groups themselves. (Hence the phenomenon of stereotype *threat*, where people seem to underperform in some task through fear of conforming to their group's stereotype; Spencer et al. 2016.) But this doesn't mean that those stereotypes are *endorsed* by everyone, of course. On the contrary, many are explicitly rejected. Nevertheless, they continue to exert an influence on people's behavior, issuing in tacit inductive inferences about members of various groups. As is now widely known, these implicit stereotypes can cause a great deal of damage in modern societies, influencing such factors as the rates and severity of sentencing for crimes, the implementation of stop-and-search policies, the outcomes of hiring

[9] There are also gender-based stereotypes, of course. ("Women are kindly," "Men are leaders," and so on.) It is unclear whether or not these result from the operations of the same social-group learning mechanism. For members of one's opposite sex are not at all similar to an out-group tribe (despite cultural tropes such as "Women are from Venus, men are from Mars"). Nevertheless, all small-scale societies and hunter-gatherer tribes have always been heavily gendered in their internal organization, and encoding of gender is an automatic and seemingly-mandatory component of person-perception (Campanella et al. 2001; Weisman et al. 2015; Martin et al. 2024).

and promotion decisions, and much more (Amodio & Cikara 2021).

Notably, group stereotypes generally take the form of what linguists call *generics*. A generic statement is comprised of a noun or noun-phrase—unqualified by quantifiers like "some," "many," "most," or "all"—combined with an adjective or adjectival phrase. There is a significant convergence here with people's naïve biology.[10] Beliefs about biological kinds, too, are generally expressed in the form of generics. ("Birds fly" "Mosquitoes bite" "Deer-ticks carry Lime disease" "Tigers are striped" and so on.) These are accepted and endorsed even though they aren't true of all instances of the kind. Indeed, some aren't even true of a majority (only female mosquitoes bite), or are true of only a tiny minority (as is the case with deer-ticks and Lyme disease). It seems that generics are the default mode of storing information about both biological and social kinds in semantic memory. In fact, even when information about a novel kind is presented to people in qualified terms ("Some Zorks live in trees") it is apt to be recalled as a generic (Leslie & Gelman 2012; Sutherland et al. 2015; Gelman et al. 2016).

Generic beliefs in both the social and biological domains are learned early, with learning being facilitated by hearing adults make generic statements about the kind (Gelman & Roberts 2017; Rhodes & Baron 2019). Learners are thereafter apt to assume that kind-membership is fixed and immutable, and that members of the kind in question will share many other unknown properties. This is true even when the generic statements in question are intended to be counter-stereotypical, as in "Girls are great at math" (Rhodes et al. 2025). Moreover, both children and adults are especially quick to acquire negative stereotypes about both social and biological kinds (such as "Arabs are terrorists" post 9/11; "Sharks attack bathers," "Deer ticks carry Lime disease," and so on). This makes good sense from an evolutionary perspective. Children should rapidly learn that snakes are poisonous, for example, even if many local species are not, and in advance of learning to recognize the specific kinds that are (Barrett & Broesch

[10] Although there isn't space to pursue this here, many have argued that humans have an innately-structured domain-specific learning mechanism for acquiring information about the natural world, especially plants and animals (Atran 1998; Gelman 2003; Medin & Atran 2004). People everywhere sort the living world into hierarchical tree-structures roughly corresponding to kingdom (animal versus plant), life-form (e.g. mammal), generic (e.g. cat), species (e.g. tiger), and sub-species (e.g. Sumatran tiger); as do young children. And people everywhere use these structures to undergird their expectations about novel instances. For instance, properties of organisms at a higher level in the tree are automatically transferred downwards (if mammals have livers then so do tigers), but people don't expect the reverse. Moreover, people everywhere are tacit essentialists about living organisms. And for sure (as one might expect) hunter-gatherer communities have extensive knowledge of the plants and animals living around them (Liebenberg 2013).

2012). They should also quickly learn that members of tribe X are murderous and to be avoided at all costs (even though females and children belonging to tribe X are unlikely to be so).

People's generic beliefs about biological kinds have a tendency to "bleed into" their beliefs about social kinds, influenced especially by the use of generic language when referring to the latter (Gelman & Roberts 2017). As noted in Section 1, people are naïve *essentialists* about biological kinds: they think that each kind has a hidden essence or inner nature, which causes the surface properties and species-specific behavior of the kind. (Post-Darwinian biology has shown this to be false, of course.) So ducks, for example, have an inner nature that causes their coloring, their body shape, the fact that they quack, lay eggs, and much more (Gelman 2003). Indeed, even 8-month-old infants expect animals to have insides rather than being hollow or rattling when shaken (suggesting that they are mostly hollow), whereas they have no such expectations about artefacts (Setoh et al. 2013). The use of generic language reliably leads young children to essentialize social kinds, as well, even when the generic statement in question suggests a non-biological cause, as in "Zarpies have striped hair because they are taught to paint stripes into their hair" (Benitez et al. 2022). Adults, on the other hand, are more discerning, and are only likely to draw essentialist conclusions when the generics they hear are suggestive of a biological cause, such as "Zarpies have beards" in contrast with "Zarpies are underpaid for their work" (Noyes & Keil 2019).[11]

The human tendency to stereotype and essentialize social groups would almost certainly have been adaptive throughout most of our history (Gil-White 2001). It would have been vitally important to draw a fundamental distinction between own-tribe and other-tribe (in-group versus out-group), as well as to swiftly identify the tribes to which people belong, and to know which out-groups are direct competitors or represent an existential threat. Moreover, stereotype-inferences would also have been much more reliable among hunter-gatherer tribes than they are today. For as we noted in Section 5, traditional societies are generally highly uniform internally. Members are apt to share the same religious beliefs, the same norms, the same tastes in music and dance, the same technology, the same modes of dress and body decoration, the same methods of food preparation, and more. So a fact learned about one member of another tribe is highly likely to extrapolate to the others (with due allowance made for the gendered structure of all small-

[11] Just as generic language is reflective of essentialist beliefs when used in the biological domain and encourages them when used in connection with social groups, so the language of "identity" that dominates discourse about groups on the political left is strongly suggestive of just the kind of group essentialism that liberals are otherwise explicitly critical of, and may well serve to further reinforce it. It seems that essentialist thinking continues to operate implicitly even among those who explicitly reject it.

scale societies, of course).

8 Tribal Feeling

Section 7 concluded that humans would seem to possess an innate domain-specific learning and reasoning system that is designed for tribal thinking. The present section will argue that there is a parallel system in the domain of motivation. This one splits into two very different components, however. One is comprised of innate pro-social motivations directed towards one's in-group (members of one's own tribe); the other is a preparedness to denigrate, harm, and ultimately to destroy out-group members (people who belong to other tribes). These can be thought of as the "good angels" and "bad angels" of human nature. We will begin with the good.

Humans are unique in the animal kingdom in the extent to which they cooperate with non-relatives. Although some other species cooperate, it is almost always with close relatives (as is the case with eusocial insects like termites and honeybees), or is restricted to particular domains, like the border-patrols of male chimpanzees (although many of these males will be related, too, since it is the females who leave the troop when they reach sexual maturity). As we noted in Section 7, there is frequent movement in and out of hunter-gatherer travelling groups, so people will often find themselves cooperating with tribal members who are strangers. And such groups are mostly composed of people who are not direct blood-kin—in fact only around one quarter of them are (Hill et al. 2011). Moreover, uniquely among primates, brothers and sisters will often live together in the same local group.[12]

Humans are also unique in the animal kingdom in their life-history profile, with an extended juvenile period during which children and adolescents need to be supported by the adults in the group (Kaplan et al. 2000; Marlowe 2005; Hill & Hurtado 2009). Hunter-gatherer males don't begin to make a net contribution to the calorific needs of the group until their late teens, and male productivity doesn't peak until the 30s or 40s (testifying to the extended learning process needed to become a successful hunter). And in many tribes, female foragers never make a net positive contribution until they pass reproductive age, because of the calorific demands of pregnancy and then breast-feeding their infants up to and beyond the age of three. Moreover, humans are uniquely long-lived among primates, with a mean adult life-span

[12] Interestingly, Dyble et al. (2015) provide a mathematical model that approximates the composition of many hunter-gatherer co-resident groups quite closely. The model assumes that individuals prefer to co-reside with relatives, but that men and women have equal influence over the outcomes, consistent with the egalitarian structure of all hunter-gatherer societies.

about 2.5 times longer than chimpanzees, and with around 30 percent of the population surviving beyond the age of 60, with both men and women making a net positive contribution over much of their final twenty or thirty years beyond the age of 40.

Humans also routinely share food with one another. This, too, is quite unusual among primates. Chimpanzee mothers will not even attempt to provision their newly-weaned offspring, although chimpanzees do sometimes get food from others through tolerated scrounging. (It can be easier to allow another individual to take some of one's food than to put up with constant harassment and begging.) Not only is food (especially meat) shared equitably throughout the hunter-gatherer band to the benefit of all, but infants and toddlers are cared for by networks of individuals within the group (in addition to the mother), both related and unrelated—a practice known as "alloparenting" (Hrdy 2009; Rosenberg 2021). Nothing like this is found among other apes, and alloparenting of any kind is quite rare among primates generally. Exceptions include marmosets and tamarins, but these animals live in small family groups composed mostly of related individuals.

Burkart et al. (2009) argue that one can explain (or partially explain) how humans made the transition to becoming the cultural creatures that they are through a combination of the cognitive abilities characteristic of great apes (including, especially, nascent forms of mentalizing) with the motivational profile distinctive of cooperative breeders like marmosets and tamarins. These primates routinely engage in wide range of cooperative behaviors with both related and unrelated individuals in the group, including infant carrying, shared vigilance and defense, and spontaneous (unsolicited) provisioning with high-value food items. Combining these pro-social, altruistic, motivations with capacities for mentalizing may have facilitated intentional provisioning with information as well as with food, thus creating the conditions for culture and cultural learning to evolve and begin to accumulate.

Whatever might be true of the evolutionary history, there is little doubt that humans today have innate pro-social dispositions. Note, however, that although the experimental work done with young children to be reviewed below might suggest that human children are indiscriminately pro-social, in fact almost all of that work has been conducted in industrialized societies in which group-membership is only weakly marked. Moreover, although group membership is not made explicit in such experiments, children are likely to assume from the way their caregiver gets a friendly greeting from the experimenters that the latter, too, belong to the child's in-group. (Indeed, we know that infants are apt to make exactly this inference; Thomas et al. 2022.) So the findings in question only provide evidence of innate in-group pro-sociality.

Pro-social motivation cannot be measured directly in infants, who have little control of their own movements; but there is at least evidence that infants in the first six months of life prefer pro-social agents to anti-social agents.[13] They prefer an agent who has helped another over an agent who has hindered another (Hamlin & Wynn 2011), and they prefer an agent who interferes to prevent one agent from harming another to one who does not interfere (Kanakogi et al. 2017). And by the age of ten months, these pro-social evaluations are integrated with the output of the infants' mentalizing systems, too, leading them to prefer agents who have positive intentions over those who accidentally achieve positive outcomes (Hamlin 2013; Woo et al. 2017; Woo & Spelke 2023). Moreover, by the age of about 12 months (when infants become capable of pointing) they helpfully point out information to others, communicating facts to ignorant (but not to knowledgeable) adults engaged in a search task (Liszkowski et al. 2008).

As soon as young children become toddlers and are capable of independent motion, they display spontaneous and unsolicited altruistic-helping behavior, such as assisting an adult whose hands are occupied to open a cupboard door, or picking up a dropped object for an adult who can't reach it (Warneken & Tomasello 2009). Moreover, such behavior is intrinsically motivated, driven by empathy, rather than done in expectation of any reward. For helping-behavior is actually undermined by provision of rewards, rather than strengthened (Warneken & Tomasello 2008), and toddlers are just as pleased when someone else does the helping as when they do it themselves (Hepach et al. 2012). It is only at later ages (from around three), when children become capable of planning ahead, that they become influenced by reputational concerns in their helping behavior (Grueneisen & Warneken 2022); and House et al. (2013) show that this finding holds good across a range of different cultures.

Further support for the suggestion that humans are intrinsically motivated to help group members who need it comes from the finding that human adults display more spontaneous pro-social behavior under speeded conditions, when they don't have time to reflect (Rand 2016; Yamagishi et al. 2017); and also by the finding that adults exhibit more pro-social behavior when they are put under cognitive load, when reflection on alternative options is more difficult (Cornelissen et a. 2011; Schulz et al. 2014). And although these findings relate to people's behavior in anonymous economic games, where direct cues of

[13] The seminal experiment by Hamlin et al. (2007) has recently failed to replicate in a large multi-lab replication attempt (Lucca et al. 2025). But since that initial study there have been multiple findings of infants' evaluations of pro-social agents in many other paradigms across multiple different labs. See Woo et al. (2025) for discussion.

in-group status are not available, they are at least engaging in an activity with others, which they might well treat as a cue to in-group membership.

Indeed, one of the most remarkable findings in psychology is how easily pro-social behavior and positive evaluation of strangers can be created by the most minimal of cues of in-group status. This the minimal group effect (Dunham 2018). Originally discovered in what had been intended simply as a control-condition in another experiment (Tajfel 1970), the effect has been extensively investigated over subsequent decades. The finding is that people who are randomly assigned to a small group to engage in some task together immediately form positive expectations and attitudes towards their fellow group members. Membership of these groups can be left anonymous, or they can be marked by some arbitrary property (like wearing green T-shirts). Among other effects, people preferentially allocate resources to members of their own group, as opposed to others; and they have more positive expectations regarding the character and behavior of their in-group members.

The effect has been demonstrated in five-year-old children and younger, with children immediately showing both explicit and implicit preferences for their in-group, better expectations of the behavior of in-group members, and even biased encoding of positive versus negative information about in-group versus out-group (Dunham et al. 2011). The effect-sizes are moderate-to-large, and are at least half as strong as gender biases (which are quite powerful at this age), and equally as strong as racial biases (Yang et al. 2022). Indeed, even one-year-old infants will form a preference for individuals who like the same foods as them, or who like the same color mittens as them, which can likewise be interpreted as cues of in-group membership (Mahajan & Wynn 2012).

Dunham (2018), in his review of the literature, details a wide range of effects that follow from minimal-group membership (44 in all). These include both implicit and explicit in-group preferences and liking, more positive expectations of in-group members, and more positive traits attributed to in-group members; greater empathy for pain and more overall empathy for in-group members; more favoritism in costly giving, more trust, and greater willingness to overlook in-group member transgressions; greater willingness to accept testimony from in-group members, better face memory, better recognition of emotional facial expressions, and more. In addition, Hackel et al. (2017) show that not only do people provide more resources to anonymous members of a perceived in-group, but the extent to which they do so correlates with degrees of activity in the ventral striatum (the brain's main reward center) when they learn that an in-group member has received something good (from whatever source). It seems that people find it intrinsically rewarding when in-group members do well.

These findings make good sense in light of our history of tribal living. For membership of a cooperative group of any sort would have been a powerful cue of shared tribal membership. Hunter-gatherers would often have found themselves cooperating with strangers from the same tribe, when new members join their local group, when they themselves join a new group, or when the tribe as a whole gathers for a cooperative activity of some sort. But rarely, if ever, would they have engaged in a joint activity with members of another tribe. Since people's inclusive fitness would have been strongly impacted by how successfully they integrate with fellow tribal members, supporting and being supported by them in turn, one would expect intense selection for a default positive evaluation of one's in-group. And that is exactly what we find in contemporary populations.

There are two possible (mutually consistent) accounts of the mechanisms underlying the minimal-group effect. One is that it is yet another instance of the kinds of top-down influence of belief on affective value of the sort discussed in Section 6. It might be that recognizing someone as an in-group member causes one (innately) to believe that they are good, which in turn creates a positive valuation of them in a top-down manner. Or it could be that an appraisal of them as an in-group member directly causes (innately) positive affective valuation of them, which in turn causes an expectation that they are good. (Or both.) I am unaware of any evidence that directly adjudicates between these two causal pathways. But we do at least know (as we saw in Section 7) that human infants in their first year of life have expectations of in-group loyalty and support, suggesting that the top-down route is a possibility. In any case, on either account it seems that humans have innate pro-social dispositions caused by recognition of in-group membership.

It appears that the effects of minimal-group status in children result from immediate positive evaluations of their in-group, rather than negative evaluations of the out-group, at least initially (Buttelmann & Böhm 2014). And in adults, too, the evidence suggests that the default effect of in-group membership is positive evaluation of members of one's own group, while one's attitude to the out-group is one of disregard or indifference, rather than active dislike (Brewer 1999; Bailliet et al. 2014). The same is true of the implicit evaluative attitudes towards racial and ethnic groups that have attracted so much attention in recent decades (probably because their discovery was surprising, and also because of their insidious societal effects, since people are generally unaware of their existence). For it turns out that these positive in-group attitudes are only manifested in socially dominant groups (Whites in the US and South Africa; high-caste Hindus in India, and so on; Dunham et al. 2008). They result from combining a positive evaluation of one's own group with a prestige bias. Among Black people in the US and South Africa and in lower-caste

Hindus in India these two effects cancel one another out (Newheiser et al. 2014; Shutts et al. 2016).

Somewhat surprisingly, the effect-sizes for implicit racial attitudes (such as White versus Black) are only small-to-medium (Greenwald et al. 2015). Moreover, tests of implicit attitudes have only weak test-retest reliability (Schimmack 2021). This need not undermine their importance, since they are quite stable and reliable at the group-level (Payne et al. 2017), and because repeated or pervasive small effects can multiply to produce large impacts on individuals and on society (Greenwald et al. 2015). But the effect-sizes in question are significantly smaller than are generally found for instantly-acquired minimal-group effects, where effect-sizes tend to be medium or large. (As a result, membership of inter-racial minimal groups has been shown to mitigate or erase the effects of implicit racial biases; Van Bavel & Cunningham 2009; Pietraszewski et al. 2014.) This may be because membership of a small cooperative group is an unambiguous cue to in-group (or "tribal") membership, whereas in contemporary societies, race is only one of many overlapping "identities," with people belonging to many different cross-cutting in-groups.

While positive evaluation of one's in-group may not, by default, come paired with denigration of the out-group, even disregard can have significant negative effects. It is well-established that people mentalize less often and less deeply for members of out-groups, and that empathy for out-group suffering is significantly reduced as a result (Cikara et al. 2011; Bruneau et al. 2017). Indeed, even five-year-old children spontaneously mentalize less when describing the behavior of animated shapes that they have been told represent members of the opposite gender or people who live in a geographically distant location (McLoughlin & Over 2017). In consequence, people may acquiesce when out-group members are harmed or suffer, and fail to protest or intervene.

Attitudes to an out-group can shift rapidly from disregard to outright hostility when people perceive the out-group to be in competition with, or aggressive towards, their own group (Cikara & Van Bavel 2014; Chang et al. 2016). And inter-group hostility can result in inter-group violence in such circumstances, as when White people in the United Kingdom attack a Muslim cultural center, or when ethnic Christians in Holland attack Jewish soccer fans. And in parts of the world with a history of inter-group hostility, violence can become much more systematic, leading to civil warfare and even genocide, as happened in Rwanda in 1994, or in what was then Yugoslavia through the 1990s following the collapse of the Soviet empire.

Given our extensive evolutionary history involving regular inter-tribal killings and warfare, reviewed briefly in Section 7, it makes sense that humans should have evolved a psychology that makes them

capable of systematic inter-group killing. This is not to say that humans possess an innate drive to kill outsiders, of course, in the way that they have a drive to eat or sleep. But it means that given sufficient cues of out-group competition and (especially) threat, humans are apt to shift from neutral evaluations of out-group members to negative ones (sometimes strongly negative). They also become capable of shutting down any empathy that they might otherwise feel for out-group members, and then (given the right cultural context), engaging in lethal violence.

Notably, rates of violent inter-tribal killing approximate those among chimpanzees, which mostly result from group attacks on isolated individuals (Wrangham & Glowacki 2012). As is now well-known, male chimpanzees periodically patrol the borders of their territory, walking silently and in single file, and will attack and kill isolated individuals from other troops (especially males).[14] This behavior is adaptive, since it rarely involves injury to the attackers and gradually leads to increased dominance of one group over another, thereby securing greater access to resources. There is good reason to think that the capacity for similar behavior among humans is inherited from our primate ancestry. For Gómez et al. (2016) survey the rates of intra-species lethal violence among mammals generally, mapping those onto the mammalian family tree. They show that the all-in rates of lethal violence among hunter-gatherers closely match what would be predicted from our position on the mammalian and primate family trees.

Violent attitudes towards out-groups are in turn facilitated by the use of dehumanizing language (Kteily & Landry 2022). Certainly inter-group conflict is often *accompanied* by such language, as when Hutu radio broadcasts in Rwanda described Tutsis as "cockroaches," or when both Israelis and Palestinians rate each other as more similar to an animal than to themselves (Bruneau & Kteily 2017). This is, arguably, a strategy for further reducing any empathy one might have for the out-group—very likely a successful one, utilizing the kind of top-down impact of language and belief on affective valuation discussed in Section 6. Although it seems most likely that the strategy is culturally evolved and transmitted rather than innate, it is noteworthy that it has deep roots. For there is anecdotal evidence of hunter-gatherers using dehumanizing language when talking about members of other tribes in cases where there is a history of inter-tribal violence (Wrangham & Glowaki 2012).

Human nature really does seem to comprise both "good angels" and "bad angels," then. We are innately

[14] Notice that this kind of lethal violence directed at out-groups is cold and calculated. In contrast, most intra-group violence is reactive and anger-driven. These differing forms of violence appear to be underlain by distinct adaptive brain networks in both humans and other animals (Sarkar & Wrangham 2023).

pro-social towards members of our in-group, but in the right conditions are capable of transitioning smoothly from indifference to murderous hostility towards out-group members. The effects of the latter are readily apparent in many current conflicts around the globe.

9 Summary & Conclusion

Nativism appears to be richly true of the human mind. Not only are there numerous domain-specific learning systems inherited from our animal ancestors, but we possess a number of other specialized adaptations that are either unique to our species, or involve significant changes to those we share with other great apes. It is unclear how much of the human capacity for language is innately channeled, beyond an expectation of hierarchical and moveable phrases, together with powerful motivations to identify and attend to language and to communicate with others. Likewise, it is unclear how much of the human mentalizing faculty is unique to us. But it seems likely that at least a conceptual primitive like TELL is innate and human-specific, as are some simple attribution rules for identifying when agents are engaged in a communicative exchange. And humans are also—uniquely—powerfully motivated to attend to and represent the mental states and dispositions of other members of their social group independent of context.

Much of what is unique to human nature relates to our existence as cultural creatures. We seem to possess a number of innate biases for cultural learning and cultural teaching, as well as for identifying and generalizing about out-group members (which meant generalizing about members of other tribes throughout most of our evolutionary history). Moreover, mechanisms for evaluative learning inherited from our animal ancestors appear to have been radically transformed to enable large top-down influences of belief and expectation on value-acquisition, thereby enabling rapid acquisition of the values and norms of one's culture. Innate pro-social motivations towards in-group members have been added, too, happening after our lineage split off from the other great apes, introduced alongside a much more ancient capacity for inter-group violence. Overall, it seems that much of what is characteristic of human nature can be explained in terms of changes to innate affective and motivational systems.

The main conclusion of this Element is that humans are deeply cultural creatures, adapted for cultural learning and cultural living. But we are also tribal animals, adapted for living in small homogenous communities. This may have major implications for human welfare. One is that inter-group suspicion, hatred, and/or violence are not only endemic but deep-rooted. And although the material gains that have come from science and industrialization are undeniable, it is perhaps no accident that contemporary populations in the developed world suffer from an epidemic of loneliness and alienation. For most of us

now live largely alone or in small families without significant social support, and lack the sense of belonging characteristic of life in small-scale communities. We are tribal creatures struggling to adapt to a cosmopolitan world.

References

Allaby, R., Stevens, C., Lucas, L., Maeda, O., & Fuller, D. (2017). Geographic mosaics and changing rates of cereal domestication. *Philosophical Transactions of the Royal Society B*, 372, 20160429.

Allen, M.W. & Jones, T.L. (eds.) (2014). *Violence and Warfare Among Hunter-Gatherers*. Left Coast Press.

Amodio, D. & Cikara, M. (2021). The social neuroscience of prejudice. *Annual Review of Psychology*, 72, 439–469.

Apperly, I. & Butterfill, S. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review*, 116, 953–970.

Aristotle. (c.350 BCE). *On the Soul* (*De Anima*). Many translations and editions available.

Arnon, I., Carmel, L., Claidière, N., Fitch, W.T., Goldin-Meadow, S., Kirby, S., Okanoya, K, Raviv, L., Wolters, L., & Fisher, S. (2025). What enables human language? A biocultural framework. *Science*, 390, eadq8303.

Ashar, Y., Chang, L.J., & Wager, T. (2017). Brain mechanisms of the placebo effect: An affective appraisal account. *Annual Review of Clinical Psychology*, 13, 73–98.

Ashar, Y., Gordon, A., Schubiner, H., Uipi, C., Knight, K., Anderson, Z., Carlisle, J., Polisky, L., Geuter, S., Flood, T., Kragel, P., Dimidjan, S., Lumley, M., & Wager, T. D. (2022). Effect of pain reprocessing therapy vs placebo and usual care for patients with chronic back pain: A randomized clinical trial. *JAMA Psychiatry*, 79, 13–23.

Atran, S. (1998). Folk biology and the anthropology of science: Cognitive universals and cultural particulars. *Behavioral and Brain Sciences*, 21, 547–609.

Atran, S. (2002). *In Gods We Trust.* Oxford University Press.

Baillargeon, R., Scott, R., & Bian, L. (2016). Psychological reasoning in infancy. *Annual Review of Psychology*, 67, 110–118.

Baillargeon, R., Southgate, V., & Buttelmann, D. (2018). Interpreting failed replications of early false-belief findings: Methodological and theoretical considerations. *Cognitive Development*, 46, 112–124.

Balci, F., Freestone, D., & Gallistel, C.R. (2009). Risk assessment in man and mouse. *Proceedings of the National Academy of Sciences*, 106, 2459–2463.

Baldwin, D. (1991). Infants' contribution to the achievement of joint reference. *Child Development*, 62, 874–890.

Bardi, L., Regolin, L., & Simion, F. (2011). Biological motion preferences in humans at birth: Role of dynamical and configural properties. *Developmental Science*, 14, 353–359.

Barlev, M. & Shtulman, A. (2021). Minds, bodies, spirits, and gods: Does widespread belief in disembodied being imply that we are inherent dualists? *Psychological Review*, 128, 1007–1021.

Baron-Cohen, S. (1995). *Mindblindness*. MIT Press.

Baron-Cohen, S., Leslie, A., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, 21, 37–

46.

Barrett, H.C. & Broesch, J. (2012). Prepared social learning about dangerous animals in children. *Evolution and Human Behavior*, 33, 499–508.

Barrett, L.F. (2017). *How Emotions are Made*. Mariner Books.

Benitez, J., Leshin, R., & Rhodes, M. (2022). The influence of linguistic form and causal explanations on the development of social essentialism. *Cognition*, 229, 105246.

Bian, L. & Baillargeon, R. (2022). When are similar individuals a group? Early reasoning about similarity and in-group support. *Psychological Science*, 33, 752–764.

Bloom, P. (2000). *How Children Learn the Meaning of Words*. MIT Press.

Bloom, P. (2004). *Descartes' Baby*. Basic Books.

Boehm, C. (2001). *Hierarchy in the Forest*. Harvard University Press.

Boyd, R. & Richerson, P. (1985). *Culture and the Evolutionary Process*. University of Chicago Press.

Boyd, R. & Richerson, P. (1992). Punishment allows for the evolution of cooperation (or anything else) in sizable groups. *Ethology and Sociobiology*, 13, 171–195.

Boyd, R. (2018). *A Different Kind of Animal*. Princeton University Press.

Boyd, R., & Richerson, P. (2022). Large-scale cooperation in small-scale foraging societies. *Evolutionary Anthropology: Issues, News, and Reviews*, 31, 175–198.

Boyer, P. (2001). *Religion Explained*. Basic Books.

Boyette, A. & Hewlett, B. (2018). Teaching in hunter-gatherers. *Review of Philosophy and Psychology*, 9, 771–797.

Brehm, J. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, 52, #384.

Brentari, D. & Goldin-Meadow, S. (2017). Language emergence. *Annual Review of Linguistics*, 3, 363–388.

Brown, D. (2001). *Human Universals*. McGraw-Hill.

Bruneau, E. & Kteily, N. (2017). The enemy as animal: Symmetric dehumanization during asymmetric warfare. *PLoS one*, 12, e0181422.

Bruneau, E., Cikara, M., & Saxe, R. (2017). Parochial empathy predicts reduced altruism and endorsement of passive harm. *Social Psychological and Personality Science*, 8, 934–942.

Buckner, C. (2023). *From Deep Learning to Rational Machines*. Oxford University Press.

Burkart, J., Hrdy, S., & Van Schaik, C. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology*, 18, 175–186.

Burnside, K., Severdija, V., & Poulin-Dubois, D. (2020). Infants attribute false beliefs to a toy crane. *Developmental Science*, 23, e12887.

Buttelmann, D., Zmyj, N., Daum, M., & Carpenter, M. (2013). Selective imitation of in-group over out-group members in 14-month-old infants. *Child Development*, 84, 422–428.

Buttelmann, F. & Kovács, A. (2019). 14-month-olds anticipate others' actions based on their belief about an object's identity. *Infancy*, 24, 738–751.

Butterfill, S. & Apperly, I. (2013). How to construct a minimal theory of mind. *Mind & Language*, 28, 606–637.

Byrne, R. & Whiten, A. (eds.) (1988). *Machiavellian Intelligence*. Oxford University Press.

Call, J. & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, 12, 187–192.

Campanella, S., Chrysochoos, A., & Bruyer, R. (2001). Categorical perception of facial gender information: Behavioral evidence and the face-space metaphor. *Visual Cognition*, 8, 237–262.

Cantlon, J. & Piantadosi, S. (2024). Uniquely human intelligence arose from expanded information capacity. *Nature Reviews Psychology*, 3, 275–293

Carruthers, P. (1992). *Human Knowledge and Human Nature*. Oxford University Press.

Carruthers, P. (2006). *The Architecture of the Mind*. Oxford University Press.

Carruthers, P. (2011). *The Opacity of Mind*. Oxford University Press.

Carruthers, P. (2013). Mindreading in infancy. *Mind & Language*, 28, 141–172.

Carruthers, P. (2015). Two systems for mindreading? *Review of Philosophy and Psychology*, 7, 141–162.

Carruthers, P. (2017). Mindreading in adults: Evaluating two-systems views. *Synthese*, 194, 673–688.

Carruthers, P. (2020). How mindreading might mislead cognitive science. *Journal of Consciousness Studies*, 27 (7-8), 195–219.

Carruthers, P. (2024). *Human Motives*. Oxford University Press.

Carruthers, P. (2025). A content-general adaptation for tribal value-acquisition. *Evolution and Human Behavior*, 46, 106791.

Cesana-Arlotti, N., Kovács, Á., & Téglás, E. (2020). Infants recruit logic to learn about the social world. *Nature Communications*, 11, 5999.

Cesana-Arlotti, N., Martín, A., Téglás, E., Vorobyova, L., Cetnarski, R., & Bonatti, L. (2018). Precursors of logical reasoning in preverbal infants. *Science*, 359, 1263–1266.

Chang, L., Krosch, A., & Cikara, M. (2016). Effects of intergroup threat on mind, brain, and behavior. *Current Opinion in Psychology*, 11, 69–73.

Charlesworth, T., Kurdi, B., & Banaji, M. (2020). Children's implicit attitude acquisition: Evaluative statements succeed, repeated pairings fail. *Developmental Science*, 23, e12911.

Chomsky, N. (1957). *Syntactic Structures*. Mouton Press.

Chomsky, N. (1965). *Aspects of a Theory of Syntax*. MIT Press.

Chudek, M. & Henrich, J. (2011). Culture-gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in Cognitive Sciences*, 15, 218–226.

Chudek, M., McNamara, R., Birch, S., Bloom, P., & Henrich, J. (2018). Do minds switch bodies? Dualist interpretations across ages and societies. *Religion, Brain & Behavior*, 8, 354–368.

Cikara, M. & Van Bavel, J. (2014). The neuroscience of intergroup relations: An integrative review. *Perspectives on Psychological Science*, 9, 245–274.

Cikara, M., Bruneau, E., & Saxe, R. (2011). Us and them: Intergroup failures of empathy. *Current Directions in Psychological Science*, 20, 149–153.

Cohen, E., Burdett, E., Knight, N., & Barrett, J. (2011). Cross-cultural similarities and differences in person-body

reasoning: Experimental evidence from the United Kingdom and Brazilian Amazon. *Cognitive Science*, 35, 1282–1304.

Cornelissen, G., Dewitte, S., & Warlop, L. (2011). Are social value orientations expressed automatically? Decision making in the dictator game. *Personality and Social Psychology Bulletin*, 37, 1080–1090.

Csibra, G. & Gergely, G. (2011). Natural pedagogy as an evolutionary adaptation. *Philosophical Transactions of the Royal Society B*, 366, 1149–1157.

Cuijpers, P., Driessen, E., Hollon, S., van Oppen, P., Barth, J., & Andersson, G. (2012). The efficacy of non-directive supportive therapy for adult depression: A meta-analysis. *Clinical Psychology Review*, 32, 280–291.

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection*. John Murray.

de Araujo, I., Rolls, E., Velazco, M., Margot, C., & Cayeux, I. (2005) Cognitive modulation of olfactory processing. *Neuron*, 46, 671–679.

Derex, M., Bonnefon, J., Boyd, R., & Mesoudi, A. (2019). Causal understanding is not necessary for the improvement of culturally evolving technology. *Nature Human Behavior*, 3, 446–452.

Descartes, R. (1641). *Meditations on First Philosophy*. Many translations and editions now available.

Dörrenberg, S., Rakoczy, H., & Liszkowski, U. (2018). How (not) to measure infant Theory of Mind: Testing the replicability and validity of four non-verbal measures. *Cognitive Development*, 46, 12–30.

Dunham, Y. (2018). Mere membership. *Trends in Cognitive Sciences*, 22, 780–793.

Dunham, Y., Baron, A.S., & Banaji, M. (2008). The development of implicit intergroup cognition. *Trends in Cognitive Sciences*, 12, 248–253.

Dunham, Y., Baron, A.S., & Carey, S. (2011). Consequences of "minimal" group affiliations in children. *Child Development*, 82, 793–811.

Dyble, M., Salali, G., Chaudhary, N., Page, A., Smith, D., Thompson, J., Vinicius, L., Mace, R., & Migliano, A. (2015). Sex equality can explain the unique social structure of hunter-gatherer bands. *Science*, 384, 796–798.

Egan, L., Bloom, P., & Santos, L. (2010). Choice-induced preferences in the absence of choice: Evidence from a blind two choice paradigm with young children and capuchin monkeys. *Journal of Experimental Social Psychology*, 46, 204–207.

Egan, L., Santos, L., & Bloom, P. (2007). The origins of cognitive dissonance: Evidence from children and monkeys. *Psychological science*, 18, 978–983.

Elekes, F., Varga, M., & Király, I. (2016). Evidence for spontaneous level-2 perspective taking in adults. *Consciousness and Cognition*, 41, 93–103.

Elekes, F., Varga, M., & Király, I. (2017). Level-2 perspectives computed quickly and spontaneously: Evidence from eight- to 9.5-year-old children. *British Journal of Developmental Psychology*, 35, 609–622.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking Innateness*. MIT Press.

Enisman, M., Shpitzer, H., & Kleiman, T. (2021). Choice changes preferences, not merely reflects them: A meta-

analysis of the artifact-free free-choice paradigm. *Journal of Personality and Social Psychology*, 120, #16.

Fedorenko, E., Ivanova, A., & Regev, T. (2024). The language network as a natural kind within the broader landscape of the human brain. *Nature Reviews Neuroscience*, 25, 289–312.

Ferrigno, S., Huang, Y., & Cantlon, J. (2021). Reasoning through the disjunctive syllogism in monkeys. *Psychological Science*, 32, 292–300.

Ferry, A., Hespos, S., & Waxman, S. (2010). Categorization in 3- and 4-month-old infants: An advantage of words over tones. *Child Development*, 81, 472–479.

Fitch W.T., Mathur, N., de Boer, B., & Ghazanfar, A. (2016). Monkey vocal tracts are speech-ready. *Science Advances*, 2, e1600723.

Fitch, W.T. (2018). The biology and evolution of speech: A comparative analysis. *Annual Review of Linguistics*, 4, 255–279.

Flood, J. (2019). *The Original Australians* (2nd ed.). Allen & Unwin.

Forgács, B., Gervain, J., Parise, E., Csibra, G., Gergely, G., Baross, J., & Király, I. (2020). Electrophysiological investigation of infants' understanding of understanding. *Developmental Cognitive Neuroscience*, 43, 100783.

Forgács, B., Parise, E., Csibra, G., Gergely, G., Jacquey, L., & Gervain, J. (2019). Fourteen-month-old infants track the language comprehension of communicative partners. *Developmental Science*, 22, e12751.

Forgács, B., Tauzin, T., Gergely, G., & Gervain, J. (2022). The newborn brain is sensitive to the communicative function of language. *Nature Scientific Reports*, 12, 1220.

Forstmann, M. & Burgmer, P. (2015). Adults are intuitive mind-body dualists. *Journal of Experimental Psychology: General*, 144, 222–235.

Fuller, D. & Lucas, L. (2025). Contrasting pathways of domestication and agriculture around Southwest Asia. *Archaeological and Anthropological Sciences*, 17, 74.

Gallardon, G., Eichner, C., Sherwood, C., Hopkins, W., Anwander, A., & Friederici, A. (2023). Morphological evolution of language-relevant brain areas. *PLoS Biology*, 21, e3002266.

Gallistel, C.R. (1990). *The Organization of Learning*. MIT Press.

Gallistel, C.R. (2000). The replacement of general-purpose learning models with adaptively specialized learning modules. In M. Gazzaniga (ed.), *The Cognitive Neurosciences 2nd Edition*. MIT Press.

Gat, A. (2015). Proving communal warfare among hunter-gatherers: The quasi-Rousseauan error. *Evolutionary Anthropology*, 24, 111–126.

Geers, A., Kosbab, K., Helfer, S., Weiland, P., & Wellman, J. (2007). Further evidence for individual differences in placebo responding: An interactionist perspective. *Journal of Psychosomatic Research*, 62, 563–570.

Gelman, S. & Roberts, S. (2017). How language shapes the cultural inheritance of categories. *Proceedings of the National Academy of Sciences*, 114, 7900–7907.

Gelman, S. (2003). *The Essential Child*. Oxford University Press.

Gelman, S., Sánchez Tapia, I., & Leslie, S-J. (2016). Memory for generic and quantified sentences in Spanish-speaking children and adults. *Journal of Child Language*, 43, 1231–1244.

Gibbons, M., Versace, E., Crump, A., Baran, B., & Chittka, L. (2022). Motivational trade-offs and modulation of nociception in bumblebees. *Proceedings of the National Academy of Sciences*, 119, e2205821119.

Gibson, E., Futrell, R., Piantadosi, S., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23, 389–407.

Gil-White, F. (2001). Are ethnic groups biological "species" to the human brain? Essentialism in our cognition of some social categories. *Current Anthropology*, 42, 515–553.

Gleitman, L. & Trueswell, J. (2020). Easy words: Reference resolution in a malevolent referent world. *Topics in Cognitive Science*, 12, 22–47.

Goldin-Meadow, S. (2020). Discovering the biases children bring to language learning. *Child Development Perspectives*, 14, 195–201.

Goldman, A. (1989). Interpretation psychologized. *Mind and Language*, 4, 161–185.

Goldman, A. (2006). *Simulating Minds*. Oxford University Press.

Golovanova, L., Doronichev, V., Doronicheva, E., Sapega, V., & Shackley, M. (2021). Long-distance contacts and social networks of the Upper Palaeolithic humans in the North-Western Caucasus (on data from Mezmaiskaya Cave, Russia). *Journal of Archaeological Science: Reports*, 39, #103118.

Gómez, J., Verdú, M., González-Megías, A., & Méndez, M. (2016). The phylogenetic roots of human lethal violence. *Nature*, 538, 233–237.

Gopnik, A. & Meltzoff, A. (1997). *Words, Thoughts, and Theories*. MIT Press.

Grabenhorst, F., Rolls, E., & Bilderbeck, A. (2008). How cognition modulates affective responses to taste and flavor: Top-down influences on the orbitofrontal and pregenual cingulate cortices. *Cerebral Cortex*, 18, 1549–1559.

Greenwald, A., Banaji, M., & Nosek, B. (2015). Societally small effects of the implicit association test can have societally large effects. *Journal of Personality and Social Psychology*, 108, 553–561.

Gregg, A., Seibt, B., & Banaji, M. (2006). Easier done than undone: Asymmetry in the malleability of implicit preferences. *Journal of Personality and Social Psychology*, 90, 1–20.

Griffiths, P. (2002). What is innateness? *The Monist*, 85, 70–85.

Griffiths, P., Machery, E., & Linquist, S. (2009). The vernacular concept of innateness. *Mind & Language*, 24, 605–630.

Grueneisen, S. & Warneken, F. (2022). The development of prosocial behavior: From sympathy to strategy. *Current Opinion in Psychology*, 43, 323–328

Hackel, L., Zaki, J., & Van Bavel, J. (2017). Social identity shapes social valuation: Evidence from prosocial behavior and vicarious reward. *Social Cognitive and Affective Neuroscience*, 12, 1219–1228.

Hamlin, J.K. & Wynn, K. (2011). Young infants prefer prosocial to antisocial others. *Cognitive Development*, 26, 30–39.

Hamlin, J.K. (2013). Failed attempts to help and harm: Intention versus outcome in preverbal infants' social evaluations. *Cognition*, 128, 451V474.

Hamlin, J.K., Wynn, K., & Bloom, P. (2007). Social evaluation by preverbal infants." *Nature*, 450, 557–559.

Hare, B., Call, J., Agnetta, B. & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behavior*, 59, 771–785.

Harris, J.A., Boyd, R., & Wood, B. (2021). The role of causal knowledge in the evolution of traditional technology. *Current Biology*, 31, 1798–1803.

Hart, B. & Risley, T. (1995). *Meaningful Differences in the Everyday Experience of Young American Children*. Brookes Press.

Hauser, M., Chomsky, N., & Fitch, W.T. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569–1579.

Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., Kato, S., Hori, Y., Nagai, Y., Iijima, A., Someya, T., & Hasegawa, I. (2020). Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Reports*, 30, 4433–4444.

Henrich, J. & Gil-White, F. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22, 165–196.

Henrich, J. & Muthukrishna, M. (2024). What makes us smart? *Topics in Cognitive Science*, 16, 322–342.

Henrich, J. (2004). Demography and cultural evolution: How adaptive cultural processes produce maladaptive losses: The Tasmanian case. *American Antiquity*, 69, 197–214.

Henrich, J. (2016). *The Secret of Our Success*. Princeton University Press.

Henrich, J., Blasi, D., Curtin, C., Davis, H., Hong, Z., Kelly, D., & Kroupin, I. (2023). A cultural species and its cognitive phenotypes: Implications for philosophy. *Review of Philosophy and Psychology*, 14, 349–386.

Hepach, R., Vaish, A., & Tomasello, M. (2012). Young children are intrinsically motivated to see others helped. *Psychological Science*, 23, 967–972.

Hermer, L. & Spelke, E. (1996). Modularity and development: The case of spatial reorientation. *Cognition*, 61, 195–232.

Hewlett, B. & Roulette, C. (2016). Teaching in hunter-gatherer infancy. *Royal Society Open Science*, 3, #150403.

Heyes, C. (2018). *Cognitive Gadgets*. Harvard University Press.

Hickok, G. & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8, 393–402.

Hill, K. & Hurtado, A.M. (2009). Cooperative breeding in South American hunter-gatherers. *Proceedings of the Royal Society B*, 276, 3863–3870.

Hill, K., Walker, R., Božičević, M., Eder, J., Headland, T., Hewlett, B., Hurtado, A.M., Marlowe, F., Wiessner, P., & Wood, B. (2011). Co-residence patterns in hunter-gatherer societies show unique human social structure. *Science*, 331, 1286–1289.

Hill, K., Wood, B., Baggio, J., Hurtado, A., & Boyd, R. (2014). Hunter-gatherer inter-band interaction rates: Implications for cumulative culture. *PloS One*, 9, e102806.

Hobbes, T. (1651). *Leviathan*. Many editions now available.

House, B., Kanngiesser, P., Barrett, H.C., Broesch, T., Cebioglu, S., Crittendon, A., Erut, A., Lew-Levy, S., Sebastian-Enesco, C. Marcus Smith, A., Yilmaz, S., & Silk, J. (2020). Universal norm psychology leads to

societal diversity in prosocial behavior and development. *Nature Human Behavior*, 4, 36–44.

House, B., Silk, J., Henrich, J., Barrett, H.C., Scelza, B., Boyette, A., Hewlett, B., McElreath, R., & Laurence, S. (2013). Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences*, 110, 14586–14591.

Hrdy, S. (2009). *Mothers and Others*. Harvard University Press.

Hume, D. (1739). *A Treatise of Human Nature*. Many editions now available.

Izard, C. (2007). Basic emotions, natural kinds, emotion schemas, and a new paradigm. *Perspectives on Psychological Science*, 2, 260–280.

Jara-Ettinger, J., Gweon, H., Shulz, L., & Tenenbaum, J. (2016). Computational principles underlying common-sense psychology. *Trends in Cognitive Sciences*, 20, 589–604.

Jin, K-s. & Baillargeon, R. (2017). Infants possess an abstract expectation of group support. *Proceedings of the National Academy of Sciences*, 114, 8199–8204.

Jones, R. (1995). Tasmanian archaeology: Establishing the sequence. *Annual Review of Anthropology*, 24, 423–446.

Kahneman, D. (2011). *Thinking, Fast and Slow*. Farrar, Straus, and Giroux.

Kammermeier, M. & Paulus, M. (2018). Do action-based tasks evidence false-belief understanding in young children? *Cognitive Development*, 46, 31–39.

Kampis, D. & Kovács, Á. (2021). Seeing the world from others' perspective: 14-month-olds show altercentric modulation effects by others' beliefs. *Open Mind*, 5, 189–207.

Kampis, D., Kármán, P., Csibra, G., Southgate, V., & Hernik, M. (2021). A two-lab direct replication attempt of Southgate, Senju and Csibra (2007). *Royal Society Open Science*, 8, 210190.

Kanakogi, Y., Inoue, Y., Matsuda, G., Butler, G., Hiraki, K., & Myowa-Yamakoshi, M. (2017). Preverbal infants affirm third-party interventions that protect victims from aggressors. *Nature Human Behavior*, 1, 0037.

Kano, F., Krupenye, C., Hirata, S., Tomonaga, M., & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences*, 116, 20904–20909.

Kaplan, H., Hill, K., Lancaster, J., & Hurtado, A. (2000). A theory of human life history evolution: Diet, intelligence, and longevity. *Evolutionary Anthropology*, *9*, 156–185.

Kern, A., Kramm, C., Witt, C., & Barth, J. (2020). The influence of personality traits on the placebo/nocebo response: A systematic review. *Journal of Psychosomatic Research*, 128, #109866.

Kerner, G., Neehus, A., Philippot, Q., Bohlen, J., Rinchai, D., Kerrouche, N., Puel, A., Zhang, S., Boisson-Dupuis, S., Abel, L., & Casanova, J. (2023). Genetic adaptation to pathogens and increased risk of inflammatory disorders in post-Neolithic Europe. *Cell genomics*, 3, #100248.

Khan, A., Faucett, J., Lichtenberg, P., Kirsch, I., & Brown, W.A. (2012). A systematic review of the comparative efficacy of treatments and controls for depression. *PLoS ONE*, 7, e41778.

Khatsenovich, A., Shelepaev, R., Rybin, E., Shelepov, Y., Marchenko, D., Odsuren, D., Gunchinsuren, B., & Olsen, J. (2020). Long distance transport and use of mica in the Initial Upper Paleolithic of Central Asia: An example from the Kharganyn Gol 5 site (northern Mongolia). *Journal of Archaeological Science: Reports*, 31, #102307.

Khefets, A. & Gallistel, C.R. (2012). Mice take calculated risks. *Proceedings of the National Academy of Sciences*, 109, 8776–8779.

Király, I., Oláh, K., & Kovács, Á. (2023). Can 18-month-olds revise attributed beliefs? *Open Mind*, 7, 435–444.

Király, I., Oláh, K., Csibra, G., & Kovács, Á. (2018). Retrospective attribution of false beliefs in 3-year-old children. *Proceedings of the National Academy of Sciences*, 115, 11477–11482.

Kline, M. & Boyd, R. (2010). Population size predicts technological complexity in Oceania. *Proceedings of the Royal Society B*, 277, 2559–2564.

Knobe, J. & Samuels, R. (2013). Thinking like a scientist: Innateness as a case study. *Cognition*, 126, 72–86.

Kovács, Á., Téglás, E., & Csibra, G. (2021). Can infants adopt underspecified contents into attributed beliefs? Representational prerequisites of theory of mind. *Cognition*, 213, 104640.

Kretzer, S., Lawrence, A., Pollard, R., Ma, X., Chen, P., Amasi-Hartoonian, N., Pariante, C., Vallée, C., Meany, M., & Dazzan, P. (2024). The dynamic interplay between puberty and structural brain development as a predictor of mental health difficulties in adolescence: A systematic review. *Biological psychiatry*, 96, 585–603.

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, 354, 110–114.

Kteily, N. & Landry, A. (2022). Dehumanization: Trends, insights, and challenges. *Trends in Cognitive Sciences*, 26, 222–240.

Laurence, S. & Margolis, E. (2024). *The Building Blocks of Thought*. Oxford University Press.

Leddon, E. & Lidz, J. (2006). Reconstruction effects in child language. In D. Bamman, T. Magnitkaia, & C. Zaller (eds.), *Proceedings of the 30th Annual Boston University Conference on Language Development*, 328–39. Cascadilla Press.

LeDoux, J. (2012). Rethinking the emotional brain. *Neuron*, 73, 653–676.

Leibniz, G. (1704). *New Essays Concerning Human Understanding*. Many editions now available.

Leslie, A. & Thaiss, L. (1992). Domain specificity in conceptual development: Evidence from autism. *Cognition*, 43, 225–251.

Leslie, S-J. & Gelman, S. (2012). Quantified statements are recalled as generics: Evidence from preschool children and adults. *Cognitive Psychology*, 64, 186–214.

Lew-Levy, S., Reckin, R., Lavi, N., Cristóbal-Azkarate, J., & Ellis-Davies, K. (2017). How do hunter-gatherer children learn new subsistence skills? A meta-ethnographic review. *Human Nature*, 28, 367–394.

Li, M., Tan, H-E., Lu, Z., Tsang, K., Chung, A., & Zuker, C. (2022) Gut-brain circuits for fat preference. *Nature*, 610, 722–730.

Liberman, Z., Woodward, A., & Kinzler, K. (2017). Preverbal infants infer third-party social relationships based on language. *Cognitive Science*, 41, 622–634.

Liberman, Z., Woodward, A., Sullivan, K., & Kinzler, K. (2016). Early emerging system for reasoning about the social nature of food. *Proceedings of the National Academy of Sciences*, 113, 9480–9485.

Lidz, J. & Gagliardi, A. (2015). Universal grammar and statistical learning. *Annual Review of Linguistics*, 1, 333–

353.

Liebenberg, L. (2013). *The Origin of Science*. Available from: https://cybertracker.org

Lieberman, M., Ochsner, K., Gilbert, D., & Schacter, D. (2001). Do amnesics exhibit cognitive dissonance reduction? The role of explicit memory and attention in attitude change. *Psychological Science*, 12, 135–140.

Lindström, B., Jangard, S., Selbing, I., & Olsson, A. (2018). The role of a "common is moral" heuristic in the stability and change of moral norms. *Journal of Experimental Psychology: General*, 147, #228.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2008). Twelve-month-olds communicate helpfully and appropriately for knowledgeable and ignorant partners. *Cognition*, 108, 732–739.

Liu, Y-Z., Wang, Y-X., & Jiang, C-L. (2017). Inflammation: The common pathway in stress-related diseases. *Frontiers in Human Neuroscience*, 11, #316.

Locke. J. (1690). *An Essay Concerning Human Understanding*. Many editions now available.

Low, J. & Watts, J. (2013). Attributing false beliefs about object identity reveals a signature blind spot in humans' efficient mind-reading system. *Psychological Science*, 24, 305–311.

Lucca, K., Yuen, F., Wang, Y., Alessandroni, N., Allison, O., Alvarez, M., ... & Hamlin, J.K. (2025). Infants' social evaluation of helpers and hinderers: A large-scale, multi-lab, coordinated replication study. *Developmental Science*, 28, e13581.

Mahajan, N. & Wynn, K. (2012). Origins of "us" versus "them": Prelinguistic infants prefer similar others. *Cognition*, 124, 227–233.

Marcus, G. (2003). *The Birth of the Mind*. Basic Books.

Margoni, F., Surian, L., & Baillargeon, R. (2024). The violation-of-expectation paradigm: A conceptual overview. *Psychological Review*, 131, 716–748.

Marlowe, F. (2005). Hunter-gatherers and human evolution. *Evolutionary Anthropology*, 14, 54–67.

Marno, H., Farroni, T., Dos Santos, Y., Ekramnia, M., Nespor, M., & Mehler, J. (2015). Can you see what I am talking about? Human speech triggers referential expectation in four-month-old infants. *Scientific Reports*, 5, 13594.

Martin, A. , Guevara Beltran, D., Koster, J., & Tracy, J. (2024). Is gender primacy universal? *Proceedings of the National Academy of Sciences*, 121, e2401919121.

Mathieson, I., Lazaridis, I., Rohland, N., Mallick, S., Patterson, N., Roodenberg, S.A., Harney, E., Stewardson, K., Fernandes, D., Novak, M., & Sirak, K. (2015). Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*, 528, 499–503.

McCabe, C., Rolls, E., Bilderbeck, A., & McGlone, F. (2008). Cognitive influences on the affective representation of touch and the sight of touch in the human brain. *Social Cognitive and Affective Neuroscience*, 3, 97–108.

McCloskey, M. (1983). Naïve theories of motion. In D. Gentner & A. Stevens (eds.), *Mental Models*, Erlbaum.

McGinn, C. (1991). *The Problem of Consciousness*. Blackwell Press.

McLoughlin, N. & Over, H. (2017). Young children are more likely to spontaneously attribute mental states to members of their own group. *Psychological Science*, 28, 1503–1509.

Medin, D. & Atran, S. (2004). The native mind: Biological categorization and reasoning in development and across cultures. *Psychological Review*, 111, 960–983.

Meeuwis, S., van Middendorp, H., van Laarhoven, A., van Leijenhorst, C., Pacheco-Lopez, G., Lavrijsen, A., Veldhuijzen, D., & Evers, A. (2020). Placebo and nocebo effects for itch and itch-related immune outcomes: A systematic review of animal and human studies. *Neuroscience and Biobehavioral Reviews*, 113, 325–337.

Mercier, H. & Sperber, D. (2017). *The Enigma of Reason*. Harvard University Press.

Mobbs, D., Headley, D., Ding, W., & Dayan, P. (2020). Space, time, and fear: Survival computations along defensive circuits. *Trends in Cognitive Sciences*, 24, 228–241.

Newheiser, A., Dunham, Y., Merrill, A., Hoosain, L., & Olson, K. (2014). Preference for high status predicts implicit outgroup bias among children from low-status groups. *Developmental Psychology*, 50, 1081–1090.

Nichols, S. & Stich, S. (2003). *Mindreading*. Oxford University Press.

Nielsen, M. & Tomaselli, K. (2010). Over-imitation in Kalahari bushman children and the origins of human cultural cognition. *Psychological Science*, 21, 729–736.

Noyes, A. & Keil, F. (2019). Generics designate kinds but not always essences. *Proceedings of the National Academy of Sciences*, 116, 20354–20359.

Ogilvie, R. & Carruthers, P. (2016). Opening up vision: The case against encapsulation. *Review of Philosophy and Psychology*, 7, 721–742.

Onishi, K. & Baillargeon, R. (2005). Do 15-month-olds understand false beliefs? *Science*, 308, 255–258.

Oyama, S. (2000). *The Ontogeny of Information*. Second Edition. Duke University Press.

Panksepp, J. & Watt, D. (2011). What is basic about basic emotions? Lasting lessons from affective neuroscience. *Emotion Review*, 3, 387–396.

Payne, B., Vuletich, H., & Lundberg, K. (2017). The bias of crowds: How implicit bias bridges personal and systemic prejudice. *Psychological Inquiry*, 28, 233–248.

Peoples, H., Duda, P., & Marlowe, F. (2016). Hunter-gatherers and the origins of religion. *Human Nature*, 27, 261–282.

Petrie, K. & Rief, W. (2019). Psychobiological mechanisms of placebo and nocebo effects: Pathways to improve treatments and reduce side effects. *Annual Review of Psychology*, 70, 599–625.

Pietraszewski, D., Cosmides, L., & Tooby, J. (2014). The content of our cooperation, not the color of our skin: An alliance detection system regulates categorization by coalition and race, but not sex. *PLoS One*, 9, e88534.

Pinker, S. (2002). *The Blank Slate*. Viking.

Pinker, S. (2011). *The Better Angels of our Nature*. Viking.

Plassmann, H. & Weber, B. (2015). Individual differences in marketing placebo effects: Evidence from brain imaging and behavioral experiments. *Journal of Marketing Research*, 52, 493–510.

Plassmann, H., O'Doherty, J., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences*, 105, 1050–1054.

Plato (c.380 BCE). *Meno*. Many translations and editions now available.

Povinelli, D. (2000). *Folk Physics for Apes*. Oxford University Press.

Powell, L. & Spelke, E. (2013). Preverbal infants expect members of social groups to act alike. *Proceedings of the National Academy of Sciences*, 110, E3965–E3971.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, 1, 515–526.

Prinz, J. (2012). *Beyond Human Nature*. Allen Lane.

Pun, A., Birch, S., & Baron, A. (2021). The power of allies: Infants' expectations of obligations during intergroup conflict. *Cognition*, 211, #104630.

Pyke, G., Pulliam, H., & Charnov, E. (1977). Optimal foraging: A selective review of theory and tests. *Quarterly Review of Biology*, 52, 137–154.

Quesque, F., & many others. (2024). Defining key concepts for mental state attribution. *Nature Communications: Psychology*, 2, #29.

Quirk, G. & Mueller, D. (2008). Neural mechanisms of extinction learning and retrieval. *Neuropsychopharmacology*, 33, 56–72.

Rand, D. (2016). Cooperation, fast and slow: Meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, 27, 1192–1206.

Rhodes, M. & Baron, A. (2019). The development of social categorization. *Annual Review of Developmental Psychology*, 1, 359–386.

Rhodes, M., Gelman, S., & Leslie, S.J. (2025). How generic language shapes the development of social thought. *Trends in Cognitive Sciences*, 29, 122–132.

Rhodes, M., Hetherington, C., Brink, K., & Wellman, H. (2015). Infants' use of social partnerships to predict behavior. *Developmental Science*, 18, 909–916.

Ritchie, J.B. (2021). What's wrong with the minimal conception of innateness in cognitive science? *Synthese*, 199, S159–S176.

Roazzi, M., Nyhof, M., & Johnson, C. (2013). Mind, soul and spirit: Conceptions of immaterial identity in different cultures. *International Journal for the Psychology of Religion*, 23, 75–86.

Rolls, E. (1999). *The Brain and Emotion*. Oxford University Press

Rosenberg, K. (2021). The evolution of human infancy: Why it helps to be helpless. *Annual Review of Anthropology*, 50, 423–440.

Rousseau, J-J. (1762). *Émile, ou, De L'education*. Many editions now available.

Rumbaugh, D., Beran, M., & Savage-Rumbaugh, E. (2003). Language. In D. Maestripieri (ed.), *Primate Psychology*. Harvard University Press.

Russell, J.A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. *Psychological Bulletin*, 115, 102–141.

Samson, D., Apperly, I., Braithwaite, J., Andrews, B., & Bodley Scott, S. (2010). Seeing it their way: Evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, 36, 1255–1266.

Samuels, R. (2002). Nativism in cognitive science. *Mind & Language*, 17, 233–265.

Sarkar, A. & Wrangham, R. (2023). Evolutionary and neuroendocrine foundations of human aggression. *Trends in Cognitive Sciences*, 27, 468–493.

Scarantino, A. & Griffiths, P. (2011). Don't give up on basic emotions. *Emotion Review*, 3, 444–454.

Schimmack, U. (2021). The Implicit Association Test: A method in search of a construct. *Perspectives on Psychological Science*, 16, 396–-414.

Schmidt, L., Skvortsova, V., Kullen, C., Weber, B., & Plassmann, H. (2017). How context alters value: The brain's valuation and affective regulation system link price cues to experienced taste pleasantness. *Nature Scientific Reports*, 7, #8098.

Schulz, J., Fischbacher, U., Thöni, C., & Utikal, V. (2014). Affect and fairness: Dictator games under cognitive load. *Journal of Economic Psychology*, 41, 77–87.

Schulze, C. & Buttelmann, D. (2022). Infants differentiate between successful and failed communication in a false-belief context. *Infant Behavior and Development*, 69, 101770.

Schuwerk, T., Priewasser, B., Sodian, B., & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: A replication attempt. *Royal Society Open Science*, 5, 172273.

Schuwerk, T., Kampis, D., & many others (2022). Action anticipation based on an agent's epistemic state in toddlers and adults. In-principle accepted Stage-1 Registered Report. https://psyarxiv.com/x4jbm/

Scott, R. & Baillargeon, R. (2017). Early false-belief understanding. *Trends in Cognitive Sciences*, 21, 237–249.

Senghas, A., Kita, S., & Özyürek, A. (2004). Children creating core properties of language: Evidence from an emerging sign language in Nicaragua. *Science*, 305, 1779–1782.

Senju, A. & Csibra, G. (2008). Gaze following in human infants depends on communicative signals. *Current Biology*, 18, 668–671.

Setoh, P., Wu, D., Baillargeon, R., & Gelman, R. (2013 Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences*, 110, 15937–15942.

Sharot, T., De Martino, B., & Dolan, R. (2009). How choice reveals and shapes expected hedonic outcome. *Journal of Neuroscience*, 29, 3760–3765.

Sharot, T., Fleming, S.M., Yu, X., Koster, R., & Dolan, R. (2012). Is choice-induced preference change long-lasting? *Psychological Science*, 10, 1123–1129.

Sharot, T., Velasquez, C., & Dolan, R. (2010). Do decisions shape preference? Evidence from blind choice. *Psychological Science*, 21, 1231–1235.

Shutts, K., Brey, E., Dornbusch, L., Slywotzky, N., & Olson, K. (2016). Children use wealth cues to evaluate others. *PloS One*, 11, e0149360.

Silver, A., Stahl, A., Loiotile, R., Smith-Flores, A., & Feigenson, L. (2020). When not choosing leads to not liking: Choice-induced preference in infancy. *Psychological Science*, 31, 1422–1429.

Spelke, E. (2022). *What Babies Know*. Oxford University Press.

Spencer, S., Logel, C., & Davies, P. (2016). Stereotype threat. *Annual Review of Psychology*, 67, 415–437.

Spengler, R. (2025). *Nature's Greatest Success: How Plants Evolved to Exploit Humanity*. University of California

Press.

Sripada, C. & Stich, S. (2006). A framework for the psychology of norms. In P. Carruthers, S. Laurence, and S. Stich (eds.), *The Innate Mind, vol. 2,* Oxford University Press.

Sripada, C. (2016). Free will and the construction of options. *Philosophical Studies*, 173, 2913–2933.

Stephens, D.W. & Krebs, J. (1986). *Foraging Theory*. Princeton University Press.

Sterelny, K. (2012). *The Evolved Apprentice*. MIT Press.

Stewart-Williams, S. & Podd, J. (2004). The placebo effect: Dissolving the expectancy versus conditioning debate. *Psychological Bulletin*, 130, 324–340.

Striedter, G. (2005). *Principles of Brain Evolution*. Sinauer Associates.

Surtees, A., Butterfill, S., & Apperly, I. (2012). Direct and indirect measures of level-2 perspective-taking in children and adults. *British Journal of Developmental Psychology*, 30, 75–86.

Sutherland, S., Cimpian, A., Leslier, S-J., & Gelman, S. (2015). Memory errors reveal a bias to spontaneously generalize by categories. *Cognitive Science*, 39, 1021–1046.

Tajfel, H. (1970). Experiments in intergroup discrimination. *Scientific American*, 223, 96–102.

Takahashi, E. & Lidz, J. (2008). Beyond statistical learning in syntax. In A. Gavarró & M. Freitas (eds.), *Proceedings of GALA 2007: Language Acquisition and Development*, 444–454. Cambridge Scholars Press.

Taneja, P., Olausson, H., Trulsson, M., Svenson, P., & Baad-Hansen, L. (2021). Defining pleasant touch stimuli: a systematic review and meta-analysis. *Psychological Research*, 85, 20–35.

Thomas, A., Saxe, R., & Spelke, E. (2022). Infants infer potential social partners by observing the interactions of their parent with unknown others. *Proceedings of the National Academy of Sciences*, 119, e2121390119.

Ting, F., He, Z., & Baillargeon, R. (2019). Toddlers and infants expect individuals to refrain from helping an ingroup victim's aggressor. *Proceedings of the National Academy of Sciences*, 116, 6025–6034.

Tucker-Drob, E., Briley, D., & Harden, K. (2013). Genetic and environmental influences on cognition across development and context. *Current Directions in Psychological Science*, 22, 349–355.

Tomasello, M. (2010). *Origins of Human Communication*. MIT press.

Ullman, T., Spelke, E., Battaglia, P., & Tenenbaum, J. (2017). Mind games: Game engines as an architecture for intuitive physics. *Trends in Cognitive Sciences*, 21, 649–665.

Vallortigara, G., Regolin, L., & Marconato, F. (2005). Visually inexperienced chicks exhibit spontaneous preference for biological motion patterns. *PLoS Biology*, 3, e208.

Van Bavel, J. & Cunningham, W. (2009). Self-categorization with a novel mixed-race group moderates automatic social and racial biases. *Personality and Social Psychology Bulletin*, 35, 321–335.

Vouloumanos, A., Martin, A., & Onishi, K. (2014). Do 6-month-olds understand that speech can communicate? *Developmental Science*, 17, 872–879.

Warneken, F. & Tomasello, M. (2008). Extrinsic rewards undermine altruistic tendencies in 20-month-olds. *Developmental Psychology*, 44, 1785–1788.

Warneken, F. & Tomasello, M. (2009). Varieties of altruism in children and chimpanzees. *Trends in Cognitive Sciences*, 13, 397–402.

Webster, G., Graber, J., Gesselman, A., Crosier, B., & Schember, T. (2014). A life history theory of father absence and menarche: A meta-analysis. *Evolutionary Psychology*, 12, 273–294.

Wehner, R. & Srinivasan, M. (1981). Searching behavior of desert ants. *Journal of Comparative Physiology*, 142, 315–338.

Weisman, K., Johnson, M.V., & Shutts, K. (2015). Young children's automatic encoding of social categories. *Developmental Science*, 18, 1036–1045.

Wellman, H. (1990). *The Child's Theory of Mind*. MIT Press.

Wellman, H., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72, 655–684.

Westra, E. & Carruthers, P. (2017). Pragmatic development explains the Theory-of-Mind Scale. *Cognition*, 158, 165–176.

Westra, E. (2017). Pragmatic development and the false belief task. *Review of Philosophy and Psychology*, 8, 235–257.

Whitehead, H., Laland, K., Rendell, L., Thorogood, R., & Whiten, A. (2019). The reach of gene-culture coevolution in animals. *Nature Communications*, 10, 2405.

Whiten, A. (2021). The burgeoning reach of animal culture. *Science*, 272, eabe6514.

Wiesmann, C., Kampis, D., Poulsen, E., Schüler, C., Duplessy, H., & Southgate, V. (2022). Cognitive dissonance from 2 years of age: Toddlers', but not infants', blind choices induce preferences. *Cognition*, 223, #105039.

Wiessner, P. (2005). Norm enforcement among the Ju/'hoansi bushmen. *Human Nature*, 16(2), 115–123.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, 13, 103–128.

von Stumm, S. & Nancarrow, A. (2024). New methods, persistent issues, and one solution: Gene-environment interaction studies of childhood development. *Intelligence*, 105, 101834.

Woo, B. & Spelke, E. (2022). Eight-month-old infants' social evaluations of agents who act on false beliefs. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 44, No. 44.

Woo, B. & Spelke, E. (2023). Toddlers' social evaluations of agents who act on false beliefs. *Developmental Science*, 26, e13314.

Woo, B., Laha, A., Chen, A., & Wolf, C. (2025). The study of early social evaluation: Contextualizing failures to replicate and looking forward. *Open Mind: Discoveries in Cognitive Science*, 9, 1308–1322.

Woo, B., Steckler, C., Le, D., & Hamlin, J.K. (2017). Social evaluation of intentional, truly accidental, and negligently accidental helpers and harmers by 10-month-old infants. *Cognition*, 168, 154–163.

Woo, B., Tan, E., Yuen, F., & Hamlin, J.K. (2023). Socially evaluative contexts facilitate mentalizing. *Trends in Cognitive Sciences*, 27, 17–29.

Wrangham, R. & Glowacki, L. (2012). Intergroup aggression in chimpanzees and war in nomadic hunter-gatherers: Evaluating the chimpanzee model. *Human Nature*, 23, 5–29.

Yamagishi, T., Matsumoto, Y., Kiyonari, T., Takagishi, H., Li, Y., Kanai, R., & Sakagami, M. (2017). Response time in economic games reflects different types of decision conflict for prosocial and proself individuals.

Proceedings of the National Academy of Sciences, 114, 6394–6399.

Yamashiro, A. & Vouloumanos, A. (2018). How do infants and adults process communicative events in real time? *Journal of Experimental Child Psychology*, 173, 268–283,

Yang, X., Yang, F., Guo, C., & Dunham, Y. (2022). Which group matters more: The relative strength of minimal vs. gender and race group memberships in children's intergroup thinking. *Acta Psychologica*, 229, #103685.

Yeomans, M., Leitch, M., Gould, N., & Mobini, S. (2008). Differential hedonic, sensory and behavioral changes associated with flavor-nutrient and flavor-flavor learning. *Physiology & Behavior*, 93, 798–806.

Acknowledgments

I am grateful to a group of students at the University of Maryland who provided very helpful discussion and feedback when I taught a graduate seminar covering these topics in Spring 2024, and to the graduate students who later read and discussed with me a draft of the first few sections of this Element. Thanks also go to Joseph Mendola, Louis Trost, and two anonymous referees for Cambridge University Press for their comments on an earlier draft of the entire Element.

Dedication

— for Seyla —

— who is no more of a blank slate than her father was —