The evolution and development of the uniquely human capacity for emotional awareness: a synthesis of comparative anatomical, cognitive, neurocomputational, and evolutionary psychological perspectives

Ryan Smith¹, Richard D. Lane², Netzin Steklis², Karen L. Weihs², Horst Dieter Steklis².

¹Laureate Institute for Brain Research, 6655 South Yale Avenue, Tulsa, OK 74136.

²University of Arizona, 1501 N. Campbell Ave,
Tucson, AZ 85724-5002, United States

Corresponding author:

Laureate Institute for Brain Research,
6655 South Yale Avenue, Tulsa, OK 74136.
Tel.: +1 602 501 4168
E-mail address: rsmith@laureateinstitute.org (Ryan Smith).

Abstract

We present an interdisciplinary framework for understanding the expanded capacity in humans (relative to other animals) to be consciously aware of emotions (emotional awareness; EA). To do so, we provide a synthesis of many different bodies of work, including those on cognitive and computational neuroscience, evolutionary psychology, and comparative anatomy. Based on active inference models within computational neuroscience, we first argue that the disproportional expansion of association cortices in humans during evolution reflects additional hierarchical levels of processing that allow learning, inference, and simulation of multimodal regularities over longer timescales – affording abstract concept learning, the ability to internally simulate the distal future outcomes of actions, and an expanded working memory capacity. This increase in general reflective capacity then allows for emotion concept learning and the ability to simulate emotions and manipulate them in working memory when deciding how to act – plausibly due to selective pressures within a hypersocial niche. We also review evidence for domain-specific adaptations affording biased attention toward socio-affective signals necessary to learn about and infer the emotions of self and others. Drawing on the construct of life history strategy within evolutionary psychology, we argue that individual differences in EA within humans can be understood as the result of tuning particular parameters to the statistics of the local environment. When early environmental/social signals are unpredictable, this reduces the use of reflective capacity when long timescale

(socioemotional) regularities are considered unreliable. We conclude by discussing the implications for mental health and hypotheses that could be tested in future research.

Keywords: emotion; emotional awareness; computational neuroscience; active inference; life history strategy; human evolution; sex differences

1. Introduction

Many species unambiguously possess neural mechanisms for generating affective (e.g., valenced autonomic, attentional, behavioral) responses to a range of stimuli relevant to survival and reproduction (LeDoux 2012). Although somewhat more controversial (see (Panksepp et al. 2017)), it is fairly well accepted that at least some subset of these species are also capable of internally representing and experiencing some aspects of affective responses, at least at a perceptual level (e.g., experiencing pleasant/unpleasant sensations, various aspects of bodily arousal, motivations to approach/avoid). Humans also generate, represent, and experience these (or at least similar) aspects of affective responses. However, humans appear to stand in stark contrast to other animals in their expanded capacity to represent and understand affective responses at a conceptual level (e.g., “my perceived increase in heart rate corresponds to fear,” “my perception of a furrowed brow signifies anger”). That is, compared to other primates, and perhaps most other species, humans seem to display a vastly increased capacity for emotional awareness (EA)(Steklis & Lane 2013) – defined here as the level of sophistication or complexity (e.g., granularity/specificity) with which an organism can internally represent, understand, and differentiate the affective responses of self and others (Kashdan et al. 2015; Smith, Killgore, et al. 2018).

While this between-species difference in EA is uncontroversial, its phylogenetic origins (i.e., the selective pressures that favored it during evolution) are poorly understood. Further, there are well-established individual differences in EA within the human species (Smith, Killgore, et al. 2018), and the developmental (learning) basis of these within-species differences is also not well-characterized. The goal of the present paper is to propose a theoretical framework – grounded in present work within ecological and evolutionary psychology, comparative neurology, and computational and cognitive neuroscience – that can theoretically account for between- and within-species differences in EA and that can guide future research on this presently under-studied topic. The framework we propose includes a broadly two-factor account of individual differences in EA, where both factors contribute to the explanation for both within- and between-species differences. These two factors are: 1) a domain-general expansion in reflective capacity, and 2) domain-specific adaptations that facilitate prepared learning about (biased attention toward or enhanced salience of cues relevant to) the emotional states of self and others.

The first factor we will refer to as “domain-general reflective capacity,” corresponding to the remarkably expanded general human capacity for launching goal-directed internal simulations over long time scales into the past and the future and that operate on very abstract, multimodal conceptual contents. We will argue that, while widely applied to different contents, this capacity can be (at least to a significant degree) accounted for by evolutionary changes in two

general cognitive functions: an expanded capacity for abstract concept learning and an expanded capacity to manipulate concepts within working memory. This expanded reflective capacity allows humans to do several very complex internal operations, such as intentionally pondering the distant past (e.g., internally simulating and gaining new insights about past experiences), intentionally simulating counterfactuals regarding what the present *would have been like* if different actions had been taken in the distant past, and imagining different possible futures when planning what to do in the present.

Arguably the most impressive use of this uniquely human reflective capacity corresponds to its application to socio-emotional cognition – with EA being one prime example. Consider what goes into a very typical social situation in which an individual must cooperate with a coworker to accomplish a work-related task. Although seemingly simple, the ability to navigate this social situation draws on each of the reflective operations mentioned above and more. For example, the individual must 1) hold in mind their own short-and long-term needs and desires (e.g., remain on good terms with that co-worker while simultaneously competing with them for an eventual desired promotion), 2) differentiate their own mental state from that of the co-worker's and imagine what it would be like to have their beliefs, emotions, goals (etc.), and 3) simulate how each of those mental states would change given different actions they each could take. Making these inferences in turn requires integrating very abstract multimodal knowledge from past experiences with that co-worker over long time scales and integrating that

knowledge with several noisy spatiotemporal cues in the present, many details of which are also irrelevant. For example, the dynamics of facial expressions, body postures, and voice tones could or could not carry useful information about that co-worker's emotional state and future behavior in that situation, and these signals can indicate different mental states in different contexts as well (Aviezer et al. 2008; Barrett et al. 2011); simultaneously, other available sensory information, such as the co-worker's eye color or choice of clothing that day may often carry no useful information (Haker et al. 2016). Thus, high-level social-cognitive abilities associated with learning and applying a theory of mind— both cognitive and affective and to both self and others – are plausibly at or near the peak of the level of cognitive complexity achieved by humans.

This qualitative difference in general reflective capacity in humans relative to other animals is also arguably the primary causal factor in facilitating the majority of aspects of modern human life, from modern technology and career options to the structure of social and political institutions, and the formation and maintenance of durable social relationships. For example, there are many decisions in modern society that must be made now based on uncertain probabilistic outcomes that *might* occur several years later (e.g., going to college to eventually get a better job), and maintaining present day long-term relationships can require compromise and short-term sacrifices for long-term stability and general well-being (e.g., deciding whether one's or one's spouse's dream job should take higher priority in deciding where to live). When

considering such decisions, it becomes apparent that there is also considerable variation *within* the human species in reflective capacity generally and the efficacy with which it is deployed. Being aware of one's own and others' emotions in particular is crucial to social interactions in serving the dual purpose of helping to meet both individual and social needs. Awareness of what one is feeling enables a person to consider what the feeling is, what it means, what triggered it, what actions to take in a particular social situation, and ultimately to consciously and deliberately regulate one's own emotional responses. Similarly, knowing what someone else is feeling enables predictions about what other people need and are likely to do. Thus, EA enables deliberate or controlled responses to changing environmental circumstances. Together, the individual and social aspects of EA enable adaptive decision-making and behavior in social contexts.

In the first few sections of this paper we will consider the evolution of the expanded human capacity for EA in relation to the generally enhanced reflective capacity described above. We consider what is known about evolutionary changes in the human brain and the plausible selective pressures that may have contributed to these changes in the course of human evolution, especially as these are related to expanded levels of emotional and cognitive processing. In this regard, we consider how cortical expansion, in particular, can be understood through the lens of leading perspectives within cognitive and computational neuroscience. We will then argue that greater reflective capacity and EA in particular in humans relative to other animals is plausibly best understood as the

result of the selective pressures associated with the hyper-social ecological niche that humans came to occupy. This required a strong form of within-group cooperation and interdependence for survival and the concomitant ability to navigate very complex social structures and relationships. As humans evolved to sculpt this hyper-social niche, fitness was enhanced by an expanded, domain-general capacity to learn about the highly abstract, context-varying, long-timescale regularities required to successfully navigate social life.

From a computational neuroscience perspective, this expanded capacity for tracking these complex social regularities implies an increasing number of hierarchically organized levels of representation within multimodal cortices; in other words, it should correspond to a hierarchically deeper internal generative model. These additional levels of computational processing allow for expansion of the two interconnected domain-general cognitive capacities necessary for internal reflection mentioned above: concept representation and working memory manipulation. When these two narrower cognitive capacities are expanded so as to allow the ability to learn about and represent sufficiently abstract regularities over sufficiently long time scales, the interaction between them forms the basic structure necessary for goal-directed reflection. Evidence from cognitive neuroscience further suggests that these higher levels of representation correspond to a (loosely) hierarchically organized set of large-scale neural networks observed within temporal, parietal, and frontal cortices – the very cortical regions showing disproportionate expansion in humans.

With these conceptual and empirical resources in place, we will then argue that – because emotion concepts themselves refer to multimodal, long-timescale regularities in self and others (e.g., allowing predictions about the causes and consequences of affective responses) – cortical expansion during human evolution, and the deeper generative models this allows for, would provide the necessary conditions for the expanded human capacity for EA. We will also illustrate how, based on certain learning dynamics within deep generative models, individual differences in early experience can account for within-species human differences in EA (Smith, Killgore, et al. 2018). Specifically, we will argue that individual differences in EA and reflective capacity can be explained, at least to a large degree, by bounded rational statistical learning processes that are applied to different types of environments accessible early in an individual's development and lead to individual trait differences in a construct called life history strategy (LHS) (Figueredo et al. 2005)). Briefly, if early experience is associated with highly volatile and unpredictable environments, an individual will learn that long timescale regularities are not reliable bases for prediction and action selection. Under this implicit belief, many “unreflective” decisions can be explained that involve a lack of thoroughly considering distant future outcomes in both general and specifically social decision-making (e.g., why choose to go to college if future career benefits are highly unpredictable? Or why consider what another person is feeling if emotions are highly unpredictable from moment to moment?).

The second factor in our proposed account corresponds to domain-specific prepared learning mechanisms that may have also been selected for to help humans navigate hyper-social environments. For example, learning about and recognizing the emotions of others plausibly requires (or is at least strongly facilitated by) the automatic salience of certain perceptual features over others. For example, a failure to automatically attend to an individual's eyes and mouth may hinder emotional learning/recognition in autistic individuals with poor emotional awareness (Moriuchi et al. 2017). We argue that “built-in” implicit expectations that certain cues are more informative/salient than others (e.g., automatically biasing attention toward socially relevant cues) may also have been selected for during human evolution and advanced the capacity for EA in a more domain-specific manner. This second factor will be covered in the latter portions of the paper. We then close by describing how our framework can unify a number of related empirical findings as well as make novel empirical predictions.

2. Evolution of the capacity for EA in humans: Comparative, computational, and cognitive neuroscientific perspectives

2.1 Comparative neuroscience and the evolution of EA

One of the more dramatic features of human evolution is the increase in brain size and reorganization of brain structures that are at the root of distinctively human cognitive, emotional, and behavioral capacities. As we will show, some of these neurobiological changes—ones that differentiate humans from other primates—are likely responsible for the expanded human capacity for reflective capacity generally and EA in particular.

Brain size and organizational changes in cortical and subcortical components have been empirically documented. In the course of human evolution, brain size has increased dramatically to a size that is about three times larger than expected for a primate of our body size (Falk 2016). This relative size increase itself has significant functional implications, but it also resulted in allometric size changes in cortical and subcortical brain components. For example, the total number of neurons in the human brain (i.e., the computational units), while higher than that for other primates, matches the number expected for a human-sized primate brain (Herculano-Houzel 2009). On the other hand, a number of volumetric changes in human cortical areas and subcortical structures depart from the expected allometric relationships to whole brain volume and are thus appropriate candidates for evolutionary specializations (Rilling 2006). In addition to volumetric evolutionary changes in brain areas, subtler changes, such as in neural connectivity or neurochemistry, must also be considered in assessing evolved human brain specializations (for a review, see (Reyes & Sherwood 2015)). In the following, we will review what is known about specializations in

subcortical and cortical structures and then proceed with a proposal for the selective pressures that account for the evolution of these neural specializations in relation to EA and general reflective capacity.

Before proceeding, a brief note about the meaning of ‘specialization’ is in order. We have just described one way in which we can think of neurobiological specializations, namely, changes in the size or connectivity of neural structures that depart from allometric expectations. Such changes are specializations in the sense that they evolved in the hominin lineage after the split from a common ancestor with African apes and thus they are good candidate changes that reflect adaptations to new challenges encountered during the course of hominin evolution (e.g., navigating a hypersocial niche, see below). A second equally important way in which we will be talking about specialization is in regard to the domain-specific vs. domain-general functions of specific neural systems or structures. For example, components of the human neocortical association areas may have expanded disproportionately (departing from allometric expectations) and thus likely represent evolutionary specializations for an enhanced or new function, but this function may be domain-specific or a domain-general one. For example, the function could be one dedicated to serve in the social domain (domain-specific) or serve in a variety of social and non-social domains (domain-general). When data allow, we will make this distinction.

One last point concerns the evolution of neurobiological mechanisms generally. New functions (both domain-specific and domain-general) generally evolve by co-opting and repurposing structures and functions that originally evolved for other purposes. For example, the neuropeptide oxytocin, which in early vertebrates played an anxiolytic role, was co-opted and repurposed in mammals in support of parental behavior and social bonding more generally, especially in primates (Chang et al. 2013). Likewise, it appears that, with some important exceptions, most of the neural circuitry serving social functions in primates is not exclusively dedicated to those functions but rather these neural structures were co-opted to also serve additional functions (reviewed in (Chang et al. 2013; Wittmann et al. 2018)).

With respect to subcortical (and related “limbic”) regions, humans appear to show several structural differences compared to apes. For example, compared to what would be expected for an ape brain scaled up to human size, there is a greater number of neurons in the lateral amygdala and anterior thalamus, greater density of specific neuron types (e.g., spindle neurons) in the anterior insula and anterior cingulate, lower than expected volume in the striatum, and greater than expected volume in the anterior insula, orbital/ventromedial prefrontal cortex, hippocampus, lateral amygdala, anterior thalamus and septal nuclei (reviewed in (Lew & Semendeferi 2017)). Given current evidence for the functions of these regions (e.g., reviewed in (Smith & Lane 2015; Smith, Killgore, et al. 2018)), it is plausible that many of these differences contribute to differences in affective

response generation processes (e.g., variation in the patterns and complexity of valenced autonomic and fast skeletomotor responses that can be generated in socio-affective contexts (Steklis & Lane 2013)). In this regard, it is of interest that human amygdala volume and cortical thickness of some of the above mentioned structures (e.g., anterior cingulate, frontal cortex) that are closely connected to the amygdala are correlated with social network size (Bickart et al. 2011).

The anterior insula, in particular, through its expanded size and apparently uniquely strong connectivity to other limbic and thalamic areas, may play a key role in EA by integrating body state and exteroceptive information (reviewed in (Craig 2009)). Humans appear to have a particularly well-developed pathway from the basal ventromedial and posterior ventromedial nucleus of the thalamus to the insula that provides interoceptive (homeostatic) information. This is integrated with information about the external environment, including social context, in the most anterior parts of the insula. In this conception, the anterior insula makes an important contribution to the representation of all emotions, but does not contribute exclusively to emotion processing (Craig 2009). Despite the insula's apparent contribution to EA, it likely does so through its close connections with other neural areas, especially the anterior cingulate cortex (Medford and Critchley, 2010), that together comprise a 'salience' network functioning to more generally direct cognitive resources toward processing information most relevant to current or anticipated metabolic demands (see below, and (Seeley et al. 2007)). The specific role of the insula in EA in humans

relative to other animals will need to be confirmed by further empirical evidence on the structural and functional characteristics that underlie subjective awareness.

The human neocortex also has undergone marked evolutionary changes as a consequence of the increase in human brain size. For example, the overall size of human neocortex is an allometric consequence of increased brain size; which is to say that, compared to other primates, human neocortical size is unexceptional (Finlay 2019). However, the proportions of functional neocortical areas have changed, indicating evolutionary reorganization of the human neocortex (Reyes & Sherwood 2015). Human association cortices within several temporal, parietal, and frontal regions show starkly disproportionate expansion (reviewed in (Buckner & Krienen 2013; Krienen & Buckner 2017). This expansion contrasts with much less notable changes in unimodal sensory and motor cortices, and it is mirrored in the relative timing of regional cortical maturation during human development.

There has been much discussion within the comparative neuroscience literature on the functional significance of these increases in cortical surface area. For example, under the plausible assumption that expanded association cortex corresponds to a greater number of processing units, “remapping factors” have been calculated (based on the proportion of cortical tissue relative to its number of inputs and outputs) that suggest human cortex possesses a disproportionately

greater capacity (i.e., 2.5 to 4 times greater) for abstract multimodal information processing relative to chimpanzees (Passingham & Smaers 2014). This increased processing power has also been linked to a range of particular cognitive/behavioral skills, including language ability, integration of conceptual information, tool use, decision-making and planning, socio-emotional abilities (e.g., self-awareness, moral reasoning, empathy, theory of mind), learning abstract rules and higher-order relations (i.e., relations between relations), spatial and numerical abilities, and the flexible ability to adapt to new environments (reviewed in (Mars et al. 2017)).

The human prefrontal cortex is of particular interest in the present context because it appears to have undergone evolutionary modifications that enhance aspects of social cognition (Amodio & Frith 2006). Compared to other primates, the human prefrontal cortex is larger and contains disproportionately more gray and white matter (Donahue et al., 2018). Within the prefrontal cortex, the frontal pole (BA 10) is disproportionately large and shows a distinctively specialized cytoarchitecture (low neuron density and increased space between neurons)—a minicolumn reorganization that reflects increased connectivity, corticocortical integration, and processing power (Semendeferi et al. 2011). These size and structural specializations are consistent with this region's established contribution to specific aspects of social cognition: the monitoring of one's own emotional state, empathy, and reflecting on the mental states of self and others (Amodio & Frith 2006). Of additional relevance are the findings that human ventromedial

prefrontal cortex volume predicts understanding of others and social network size (Lewis et al. 2011), and in monkeys (macaques), social network size corresponds to increases in gray matter within, and functional coupling between, prefrontal and temporal association cortex (Sallet et al. 2011). Thus, these studies further confirm the importance of the prefrontal cortex (and its connections to other areas) in social cognition of humans and other primates.

Clearly, a number of cortical and subcortical areas subserve aspects of primate social cognition, which collectively may comprise a ‘circuit for social cognition’ (reviewed by (Wittmann et al. 2018)). As we have mentioned, however, much of this circuitry is not dedicated exclusively to social functions. Nevertheless, our review up to this point suggests that in humans social information processing is comprised of a mix of domain-general and domain-specific mechanisms. This is consistent with studies on the neuroethology of primate social behavior which indicate that social information processing at the intermediate computational stage (e.g., as in association cortex) consists of a mix of domain-general and domain-specific mechanisms, while domain-specific mechanisms are more likely to operate at the input side of social information processing (Chang et al. 2013).

For primates generally, one of the best candidates for domain-specific mechanisms concerns the visual perception of social stimuli. Primates show an attentional bias toward faces and bodies of conspecifics. They actively seek out information from these stimuli as a kind of social information “foraging” (see

(Chang et al. 2013)). Days old human infants are selectively attracted to human faces or visual stimuli with similar geometrical properties, a skill that, while it exists in the absence of early visual experience, depends on it for its developmental refinement (Hoffman & Gauthier 2007). Cells in the superior temporal sulcus (STS) respond selectively to faces, and they appear to be particularly involved in the discrimination of individual faces (Wittmann et al. 2018). For humans and other primates the eyes are a special focus of attention (Kingstone et al. 2017), and eye gaze especially communicates social and non-social information of current interest. Parts of the STS are specialized for processing gaze direction (Wittmann et al. 2018). This circuitry may be more modular (Hoffman & Gauthier 2007) and specialized in humans as compared to apes, as only humans attend more to eye gaze than face orientation (Hare 2017). An increased specialization for gaze processing in humans is consistent with – among primates – the uniquely white sclera of the human eye that makes the tracking of human eye movement highly accurate and thus facilitates social communication (Kingstone et al. 2017). Weeks-old human infants prefer to look at eyes with a white sclera and extract emotional cues from it (reviewed in (Hare 2017)). The human sclera, along with a distinctly elongated eye shape that exposes more sclera, together amplify gaze direction and likely evolved to facilitate unique kinds of human social cooperation (see below and (Mayhew & Gómez 2015; Hare 2017)). The basolateral amygdala appears to play an especially important role in initiating the fast and subliminal processing of, and biased automatic attention toward, the human white sclera (Whalen et al. 2004)

and facial expressions of emotion generally (reviewed in (Leppänen & Nelson 2009)). Lastly, the idea that this is an evolutionary specialization for social cooperation is supported by comparative data from canids with varying eye and facial coloration patterns that link coloration patterns that facilitate gaze-following and gaze duration behavior to socially cooperative species (Ueda et al. 2014).

2.2 The evolution of hypersociality

We now want to consider more explicitly the socio-environmental selective pressures that, in the course of human evolution, resulted in the set of neuro-behavioral specializations we have described. We will focus on three related foundational elements of the uniquely human hypersocial niche: within-group hyper-cooperation, proactive prosociality, and social tolerance (Burkart et al. 2014). Comparative studies of primate cooperation and cognitive skills clearly show that humans outperform apes on tasks of social cognition (e.g., reading others' intentions) but not in reasoning about the physical world (e.g., space, quantities) (Tomasello & Herrmann 2010). A major reason even 2-year-old children do better than apes on social cognition tasks is that they are motivated to cooperate, to communicate about and understand joint goals and jointly solve tasks, a suite of traits called “shared intentionality” (Tomasello & Herrmann 2010). The shared intentionality that earmarks human hyper-cooperation is made possible by a high degree of social tolerance—especially through the reduction of

aggressive and competitive motivations—and a heightened disposition toward proactive prosociality; that is, the readiness to provide spontaneous, unsolicited help to others. Distinctive human hyper-cooperation, therefore, relies on the equally distinctive cognitive-emotional achievements in the capacity for mind-reading (or theory of mind), including empathic understanding, symbolic communication, and for mentalizing generally—to imagine scenarios of present and future actions and multiple possible outcomes (Suddendorf 2018). This package of distinctively human socio-cognitive traits is the basis for the ratcheting effect in human cultural evolution, wherein human cultures evolve quickly and adaptively to changing circumstances through the successive pooling and building of knowledge; in effect, the building of a “collective brain” that exceeds the intelligence of any individual (Henrich 2015; Laland & Rendell 2017). High-fidelity copying, or “overimitation”, shown by young children in different cultures but not in apes, is crucial to the transmission of cultural skills and conventions, the pooling of knowledge, and ultimately the speed of human cultural evolution (Nielsen 2012). The uniquely human mental and behavioral skills required for hyper-cooperation apparently also drove human brain size expansion. Among the factors responsible for achieving a modern-sized brain, most important after ecological challenges, was the challenge of cooperation within groups (and to some extent competition between groups; see (González-Forero & Gardner 2018)).

Proposals about the evolutionary pressures that triggered the construction of a hypersocial niche by ancestral hominins have considered factors such as a shift to cooperative foraging strategy (Tomasello et al. 2012) or a move to cooperative breeding (i.e., shared parenting by adult females; (Burkart et al. 2014)). The cooperative foraging strategy proposal suggests that ecological changes reduced food sources that could be obtained through individual foraging, thus favoring collaborative foraging with the spoils being shared by all group members, much as is the case in extant human foraging groups. In this scenario, the survival and reproductive interests of each group member became linked to those of others in the group, thus fostering proactive prosociality among all group members. In the cooperative breeding scenario, early hominins communally raised young, with both related and unrelated group members cooperating in aiding, provisioning, and protection of offspring. The evolution of cooperative breeding as a cause of hypersociality is backed by comparative primate data showing that the extent of allomaternal care is the best predictor of variation in proactive prosociality (Burkart et al. 2014). Further, comparative analyses indicate that the adoption of cooperative breeding in our hominin ancestors made possible brain expansion, and best accounts for the distinctive suite of human life history traits (increased fertility, shortened interbirth intervals, altricial neonates, early weaning, and longevity) through the energetic savings garnered by an extensive postnatal allomaternal care system (Isler & van Schaik 2012).

These are not mutually exclusive evolutionary scenarios in that both could have contributed at different times to the formation of hyper-cooperative groups, with cooperative breeding as the initial stimulus followed later by the hyper-cooperative mode pervading all human activity, including foraging (Burkart et al. 2014). In both scenarios, the end result was the evolution of an unusually tightly-knit human group founded in strong and durable emotional bonds between kin and non-kin whose fitness was interdependent (Brown & Cialdini 2015; Apicella & Silk 2019). Proactive prosociality evolved in these groups because of the advantages of acting as a cooperative unit and because fitness interdependence reduces the threat of prosociality being exploited by free-riders (Brown & Cialdini 2015). As well, the free-rider problem can be much reduced by evolving brains that can keep track of multiple relationships and interactions, which is the likely reason for the relationship between group size and brain size among primates (Dunbar 1998). Lastly, we note that our long history of living in such groups of cooperative, interdependent members entails that social isolation and ostracism can be tantamount to a death sentence. This helps explain the prevalence of social anxiety—the fear of being judged or socially excluded—which involves the anterior insula and related components of the salience network (Miskovic & Schmidt 2012).

A critical factor in both proposals for the evolution of human hypersocial groups is between-group competition (Bowles 2006). As our hominin ancestors became more proficient and successful culturally, shared cultural practices and conformity

to those practices created easily recognizable cultural boundaries (e.g., language, dress) that differentiated human groups who eventually competed with one another over resources. Indeed, the human past is rife with evidence of inter-group competition, hostility, and warfare (Kissel & Kim 2019), which forged a ‘group-minded’ psychology (Tomasello 2014). An important component of this psychology is the in-group vs. out-group distinction that humans make flexibly depending on who the collaborating entities are (e.g., foraging bands, tribes, or nations). Regardless, proactive prosocial behavior—pure altruism—is only extended to fellow cooperators. This may help explain the prevalence of xenophobia in human society. In short, a “caregiving system” had evolved to be deployed only among socially-bonded individuals (Brown & Cialdini 2015). The crucial role of the neuropeptide oxytocin in this caregiving system is well established (Brown and Cialdini, 2015). Notably, oxytocin promotes prosociality within-groups, while promoting defensive aggression toward competing out-groups (De Dreu et al. 2010).

Although this account of the evolution of hypersociality is somewhat speculative, it is consistent with the observation that prosociality (esp. increased tolerance, reduced aggression, and increased affiliative tendencies) is a common and desired outcome of animal domestication. In other words, the evolutionary changes in human prosociality and social tolerance are strikingly similar to those observed in domesticated animals, suggesting that in becoming hypersocial humans have undergone a process of self-domestication (reviewed in (Hare

2017)). This proposal is particularly strongly supported by genetic comparisons between domesticated dogs, wolves, and humans with Williams-Beuren syndrome (WBS), a congenital disorder characterized by hypersociality, which show that the same genetic locus associated with WBS also underlies hypersociality in domesticated dogs (vonHoldt et al. 2017). Of special note, too, is the finding from the well-known silver fox domestication experiments of increased serotonin concentrations in several neural structures that likely account for their friendlier disposition (Trut et al. 2009). Compared to apes, humans, too, have increased serotonergic innervation in the basolateral nuclei of the human amygdala, which are closely connected to orbitofrontal and temporal association areas (Lew et al. 2019). Given the amygdala's role in emotional arousal and serotonin's role in its suppression, this increased innervation likely contributes to the higher level of social tolerance required for hypercooperation.

We can conclude from our review of the comparative neuroscience and evolutionary evidence that the evolution of EA likely involved both co-opted domain general neurobiological mechanisms (e.g., association areas for mentalizing and general reflective capacities allowing abstract concept learning, expanded working memory capacity, and longer timescale internal simulation abilities) along with specialized, domain-specific ones (e.g., face recognition and eye gaze detection). In the next section, we will provide an account, from a cognitive and computational neuroscience perspective, of how our evolved

neural machinery serves the expanded levels of cognitive and emotional processing required to successfully navigate a hypersocial niche.

2.3 Cognitive neuroscience

A large body of recent work in cognitive neuroscience has established the existence of multiple large-scale networks within human association cortices, which display consistent gradients that may correspond to hierarchical relationships (Yeo et al. 2011; Margulies et al. 2016; Zhou et al. 2018). Each network's frontal, parietal, and temporal hubs wire preferentially to one another, leading to highly correlated activity at rest (i.e., comparable to localized unimodal sensory and motor cortices) – although they may also contain separable inter-digitated sub-networks (Hermundstad et al. 2013; Braga & Buckner 2017). They also correspond to observed activation patterns across many task-based neuroimaging analyses (Smith et al. 2009).

Importantly, these networks appear to both cooperate and compete with one another in the service of highly multi-modal, domain-general information processing functions. Based on a growing body of work (Doucet et al. 2011; Yeo et al. 2011; Power et al. 2011; Barrett & Satpute 2013), the “dorsal attention network” (DAN), the “salience network” (SN; also termed the “ventral attention network”), the “executive control network” (ECN; also termed the “frontoparietal network”), the “default mode network” (DMN; sometimes also referred to as the

“mentalizing network”), and the “limbic network” (LN) each serve separable functions. The DMN and ECN sit high in the processing hierarchy, subserving domain-general conceptualization/simulation and goal-directed control functions (such as working memory manipulation), respectively. The LN is involved in part in representing and regulating valenced visceromotor responses, while the SN is thought to use visceral information to direct cognitive resources toward selective processing of information relevant to maintaining homeostatic and allostatic control. Finally, the DAN is a lower-level network primarily involved in directing visual attention. There also appear to be (at least partial) homologues of these networks in macaques (Felleman & Van Essen 1991; Corbetta & Shulman 2002; Vincent et al. 2007, 2008; Buckner et al. 2008; Margulies et al. 2009; Binder et al. 2009; Patel et al. 2015).

Finally, a range of recent theoretical/integrative work has highlighted how interactions between these networks can allow for complex cognitive-emotional processes necessary for EA and social cognition more broadly (Barrett 2017; Smith, Killgore, et al. 2018). Specifically, the DMN may represent abstract, multimodal relationships, including those associated with identifying abstract concepts such as those associated with mental states such as emotions, goals, beliefs and values, and therefore allows for conceptualizing both one’s own cognitive and emotional states as well those observed in others; this network is also linked to temporally extended internal simulation (Binder et al. 2009). The SN and LN may both play roles in representing valenced bodily percepts during

emotions, and the ECN may facilitate goal-directed attention to emotion (and other mental state) concepts as well as holding and manipulating those concepts in working memory (Smith, Lane, et al. 2017; Smith, Lane, Alkozei, et al. 2018).

2.4 Computational neuroscience

In recent years a large body of work – both empirical (e.g., (Iglesias et al. 2013; Schwartenbeck et al. 2015; Diaconescu et al. 2017)) and theoretical (e.g., (Friston 2005, 2010; Bastos et al. 2012; Clark 2015; Badcock et al. 2017; Friston, Parr, et al. 2017)) – has begun to converge on the idea that the brain is an organ that evolved to approximate hierarchical Bayesian inference, and to use this inference process to optimally control the body. In exteroceptive perception, this amounts to inferring the most probable causes of sensory input, based on an internal generative model that 1) specifies the a priori probability of different possible causes (“priors”), 2) provides a probabilistic mapping from those causes to the sensory inputs they would produce (a so-called “likelihood” function), and 3) provides estimates of the reliability (“precision”) of different sensory inputs and priors in different contexts, which arbitrates the degree to which inputs are weighted relative to prior expectations during inference. Perception then amounts to inverting this type of generative model – that is, inferring the most probable causes given both prior expectations and sensory input (leading to a “posterior” probability estimate).

Within biologically plausible “hierarchical predictive coding” schemes that describe how the brain may implement this type of generative model in perception and conceptualization (Bastos et al. 2012; Parr & Friston 2018), it is proposed that each level of cortical processing contains columns containing populations of (superficial pyramidal) neurons that convey prediction-error signals laterally and upward to higher cortical levels, as well as populations of (deep pyramidal) neurons that convey prediction signals (prior expectations) laterally and downward to lower cortical levels. Essentially, by converging on a multi-level set of perceptual estimates that minimize precision-weighted prediction-error signals across levels, this process can approximate Bayes’ optimal inference about the causes of sensory input. Learning from exteroceptive input in these models corresponds to changes in synaptic strength values representing model parameters (e.g., prior expectations, precision estimates, and the form of the likelihood function) based on biologically plausible Hebbian update rules that respond to the consistent statistical patterns in sensory input over time.

In interoception and proprioception, the scheme described above has been extended to what have been termed “active inference” models, where prediction signals in these domains also act as visceromotor and skeletomotor commands (respectively) when these prediction signals are assigned high precision-estimates (Adams et al. 2013; Barrett & Simmons 2015; Pezzulo et al. 2015; Seth & Friston 2016; Stephan et al. 2016; Smith, Thayer, et al. 2017). In these

models, after being progressively unpacked through lower and lower cortical and subcortical levels, such (initially highly abstract) predictions dynamically alter the set points of homeostatic and skeletomotor reflex arcs over time, leading predicted trajectories to be fulfilled by bodily and behavioral responses under closed-loop control. To implement decision-making within this form of behavioral control, the underlying mathematical framework (Friston et al. 2016; Friston, FitzGerald, et al. 2017) also describes a process in which the organism attempts to infer the sequence of actions (termed a “policy”) that is predicted to both minimize the deviation between expected and preferred outcomes (i.e., its value, approximated as an expected free energy) while also maximizing information gain (e.g., improving the accuracy of the model itself). Notably, this decision-making component of active inference is most often modeled in terms of discrete (categorical) states, whereas predictive coding and lower-level motor control operate over continuous states. Thus, these will interact during conceptualization processes, in which continuous sensorimotor quantities (e.g., brightness, loudness, size, muscle tension, etc.) must be mapped onto categorical concepts (e.g., food, threat, car, etc.) (Friston, Parr, et al. 2017).

Two aspects of this perspective are especially relevant to the present manuscript. First, one straightforward manner in which evolutionary pressures can sculpt neurally instantiated generative models is by increasing hierarchical depth (see Figure 1 below), based on the spatiotemporal structure and relevant information domain of the signals an organism must learn to thrive within its

ecological niche (Linson et al. 2018). Within the increasingly complex social niches that characterize primate, hominid, and human evolution described above, fitness is very plausibly enhanced by the ability to learn and make use of more and more complex, long timescale, multimodal, and context-dependent environmental regularities (e.g., inferring the intended meaning of a spoken paragraph, learning the order, significance, and implicit norms of social rituals, learning social roles and how they are moderated by context, etc.). Importantly, this is precisely what increases in hierarchical depth (i.e., corresponding to an increased number of levels of cortical processing) can provide. As downward prediction signals appear to be conveyed by glutamatergic NMDA receptors, and these receptors have slow time constants (i.e., whereas the AMPA receptors that appear to mediate upward prediction-error signals have fast time constants; (Salin & Bullier 1995; Friston 2005)), this allows predictions at one level to involve extended, dynamic patterns of change at the level below. Higher and higher cortical levels can therefore learn about predictable patterns at longer and longer timescales ((e.g., inferring the meaning of a word, a phrase, a sentence, a paragraph, etc.; (Hasson et al. 2008, 2015; Kiebel et al. 2008; Murray et al. 2014)). In addition to temporal depth, higher levels also incorporate a wider and wider (convergent) array of inputs, leading to both greater spatial integration and the incorporation of multiple sensory channels. At these higher levels within association cortices, one can therefore learn about temporally extended covariance relationships across sensory modalities. This would allow one to learn, for example, that exteroceptive inferences about the presence of a beach

typically predict proprioceptive inferences that one is lying down and interoceptive inferences associated with pleasant low arousal (i.e., if you see a beach you would learn to lie down and expect to feel relaxed). One could think about the above-mentioned abstract predictive structure as encoding part of the content of the concept “beach.” Recent simulation work has also shown that the next level up in a hierarchy within association cortex could in turn act as a type of working memory, allowing information about being relaxed (for example) and other concepts to be combined and held active over longer time scales and inform higher-level policy selection (Parr & Friston 2017).

Importantly, association cortices that implement the domain-general large-scale neural networks described in the previous section (e.g., DMN and ECN) make up the highest levels of this type of predictive hierarchy. At these levels, the hierarchy is thought to become more “loose,” “heterarchical,” or “non-canonical” (i.e., relative to the stricter hierarchical structure observed in sensory cortices) such that long-distance connections between nodes and networks allow more “cooperative” and “competitive” interactions with no single node always playing the role of a “higher level” than another (Mesulam 1998, 2000; Smith 2017). Disproportionate expansion of these association cortices during development might therefore be better understood as expanding the local hierarchical (i.e., columnar) processing structure *within* each network node, allowing it to track and distribute its respective inferences about regularities over larger spatiotemporal scales to the rest of the brain (van den Heuvel & Sporns 2011). This would then

allow between-node/-network interactions to operate on such information for different domain-general purposes based on context, goals, etc. It is also worth briefly noting that generative models of different species also must come equipped with “structural priors” (i.e., the initial form of the generative model prior to learning, reflecting, for example, which levels of processing begin to incorporate which types of convergent input) – providing the possibility of inheriting both domain-general and domain-specific capacities (e.g., expected salience of facial signals).

The second aspect of the computational perspective that is especially relevant to the present manuscript has to do with the need for the (both phylogenetic and ontogenetic) calibration of several high-level prior expectations about the general structure of the environment. In the context of this paper, the most relevant class of such prior expectations that must be learned early in development has been variably referred to as beliefs about “transition precision” or “volatility” (Lawson et al. 2017; Mirza et al. 2019). In brief, these parameters reflect how predictable the regularities within the environment are over a given temporal scale. With respect to long temporal scales, high transition precision (low volatility) entails that environmental regularities are highly predictable such that, based on what one knows about the present, one can be highly confident in the way the distant future will go if choosing one action vs. another. This entails that reflecting on beliefs about distant future outcomes is a rational application of cognitive/metabolic resources due to its ability to reliably inform how one should

act to achieve one's needs, goals, and desires. Low transition precision (high volatility), in contrast, entails that such reflection is much less worth the cost because it will not be sufficiently informative. Crucially, learned estimates about transition precision can also be context-specific, such that reflection is “not worth it” in some situations (e.g., high stress situations involving immediate danger) but is “worth it” in others (e.g., when contemplating career choice).

Another relevant high-level prior in active inference models is “prior policy precision” – a general expectation about the reliability of achieving desired outcomes when making decisions based on current beliefs. This acts as a kind of generic “confidence” parameter, such that one will be more goal-directed and deterministic in one's choices when considering long timescales if it is high (i.e., one will expect their actions to have the eventual intended consequences); in contrast, if it is low, one will act more randomly or out of previously acquired habits that are less context-sensitive (FitzGerald et al. 2014; Pezzulo et al. 2015; Friston et al. 2016). The effects of this parameter are also dependent on the related parameter of “policy depth,” however, which reflects the number of “steps” into the future one considers in making choices. Lower policy depth therefore entails more impulsive, less forward-looking decision-making irrespective of beliefs about future predictability or policy precision (Mirza et al. 2019).

More generally, it follows from these models that if different values for these parameters are learned/inherited and assigned to high hierarchical levels of processing in an individual's brain, it will promote significant differences in that individual's predisposition to learn from and use long-timescale regularities and to subsequently engage in more or less forward-looking deliberation within cognition and action selection (Pezzulo 2012; Pezzulo et al. 2015; Smith, Thayer, et al. 2017). In the following sections we will now introduce the specific explanatory framework we propose to account for between-species and within-species differences in EA and the especially important role of a related construct in evolutionary psychology: life history strategy (LHS).

3. Between-species differences in EA

The level of EA displayed by non-human animals appears limited to experiential qualities such as valence, arousal, and motivated action; in contrast, the higher level of EA in humans further allows for recognizing affective responses using culturally varying conceptual categories such as fear, anger, jealousy, sadness, and excitement, that contain highly multimodal content involving extended trajectories of cognition and behavior (Barrett 2017). Even higher levels of EA can allow for simultaneous recognition of multiple emotions, as well as for the affective theory of mind (ToM) abilities thought to allow differentiation of the emotions of self and others. Learning to represent affective responses in these differentiated, fine-grained conceptual terms, as opposed to the coarser-grained,

perceptual level (e.g., pleasant/unpleasant, approach/avoid) discrimination abilities of other animals, can also alter and refine the more basic experiential aspects of affective responses. For example, in some cases the identification of an affective response as belonging to a specific emotion concept category may reduce the experienced level of arousal associated with that response; or conceptualizing a high arousal affective response as excitement vs. fear may alter the experienced valence of that response (Kircanski et al. 2012; Barrett 2017).

The knowledge built into acquired emotion concepts, which can also be thought of as including schemas/scripts within semantic memory (Shaver et al. 1987; Widen & Russell 2008), can further provide expectations regarding the course an affective response will take over relatively long time scales – including prototypical eliciting contexts, cognitive/physiological responses, and behavioral tendencies. For example, humans might learn to expect that an episode of fear will often be 1) preceded by thoughts or percepts involving impending danger, 2) accompanied by worry, automatic attention to potential threats, and increases in heart rate, and 3) followed by defensive behaviors such as freezing or running away. Such knowledge can provide complex and context-specific expectations, which can then be used to interpret the behavior of others, predict the affective responses of others, predict one's own future affective responses, and engage in complex goal-directed planning processes that take this information into account – abilities that can be especially adaptive within complex social situations.

3.1 Cortical expansion and reflective capacity

The content of emotion concepts therefore appears to inherently reflect particular types of long timescale, multimodal, context-dependent regularities – many of which function specifically within social contexts. For example, to thoroughly acquire the concept of sadness, one must link particular eliciting contexts (e.g., a funeral, loss of a desired job, social rejection, etc.), particular cognitive/physiological responses (e.g., rumination, low energy and motivation, pessimism, crying), and behavioral tendencies (e.g., social isolation, slowed movement). Given the nature of their content, therefore, the presence of deep generative models with many hierarchical levels appears to be a necessary condition for their acquisition. The ability to reflect on emotion concepts (e.g., holding and manipulating information about the emotions of self and others in working memory) and use this information in goal-directed internal simulation and decision-making, would similarly depend on the ability to assign high transition precision estimates to the hierarchical levels that represent them (e.g., for use in simulating the expected distal outcomes of different possible policies).

Using deep active inference models, we have recently reported computational simulations formally demonstrating the hierarchical basis of emotion concept learning (see Figure 1; (Smith, Parr, et al. 2019)), as well as both 1) the necessary interaction between emotion concept representations and higher-level

(longer-timescale) working memory processes to emulate high EA and 2) the interfering effects of low transition precision (Smith, Lane, et al. 2019). According to these models, one can think of particular network regions within expanded association cortices in humans as being “experience-expectant” (Johnson 2011; McLaughlin et al. 2017), such that genetically pre-specified axonal wiring patterns (i.e., inherited structural priors) uniquely provide these regions with the multimodal inputs allowing them to learn abstract concepts such as socio-emotional concepts if afforded the right experiences. The resulting greater capacity for EA in humans would have conferred a strong advantage within the hypersocial ecological niche of humans due to its ability to aid in predicting social behavior (and aiding local cultural transmission; (Veissière et al. 2019)), and would have required the expanded association cortices necessary for implementing the requisite high levels of internal generative models. Consistent with this idea, several studies have shown relationships between social and emotional awareness and the highest-level, most anterior regions of medial and lateral prefrontal cortex (and related DMN/ECN regions; (Frith & Frith 2006, 2012; Skerry & Saxe 2015; Smith, Lane, et al. 2017; Saxe & Houlihan 2017; Smith, Lane, Alkozei, et al. 2018; Smith, Lane, Sanova, et al. 2018)).

Figure 1. In the framework we propose in this paper, cortical expansion during human evolution corresponded in part to increases in the number of available hierarchical levels of processing within association cortices (i.e., higher levels from right to left in the bottom portion of the figure). This allowed for abstract concept learning and greater working memory capacity – which in turn allowed for (but was not specific to) the ability to learn about and reflect on the emotions of self and others (i.e., emotional awareness). The top portion of the figure depicts the relative size of association cortices (gray) in humans vs. other mammals in contrast to that of unimodal sensory processing areas (red, yellow, and blue correspond to somatosensory, auditory, and visual cortices, respectively; based on (Buckner & Krienen 2013)). The bottom portion of the figure heuristically illustrates the neural process theory proposed within the computational (active inference) framework discussed in the text, as applied to hierarchical processing of emotion-related information (based specifically on the simulations reported in (Smith, Parr, et al. 2019; Smith, Lane, et al. 2019)). Here, each additional hierarchical level integrates information from multiple modalities at the level below to learn and infer regularities over longer and longer timescales. In this case the lowest level would process interoceptive/proprioceptive/exteroceptive information relating to affective factors such as valence, arousal, motivation, and the response-eliciting context. The second level would learn regularities across those lower-level representations corresponding to emotion

concepts (e.g., fear most often generates negative valence, high arousal, avoidance motivation, and threat context), whereas the third level can maintain emotion concepts over longer timescales (i.e., working memory) and use them to inform goal-directed decision-making. In the proposed implementation, neuronal populations are arranged to reproduce known intrinsic (within cortical area) connections. Red connections are excitatory, blue connections are inhibitory, and green connections are modulatory (i.e., involve a multiplication or weighting). Cyan units correspond to state representations and (future) sensory input under each policy, while red units indicate their weighted averages. Pink units correspond to different types of prediction errors that are used to update beliefs and evaluate the value (expected free energy) of each policy within subcortical circuitry (e.g., basal ganglia, with midbrain dopaminergic neurons encoding policy precision estimates). This (neural) network formulation entails that synaptic connection strengths encode beliefs about both the way that states are related to sensory input, and the predictability of future states given present states (i.e., transition precision). Only exemplar connections are shown to avoid visual clutter. Furthermore, we have just shown neuronal populations encoding hidden states under two policies over three time points (i.e., two transitions), whereas individuals with greater policy depth would engage additional units encoding predictions about more distal future time points. The implementation of the major parameters discussed within the text are underlined in the figure. For more information regarding the mathematics and processes illustrated in this figure, see (KJ Friston, Lin, et al., 2017; KJ Friston, Parr, et al., 2017).

To be clear, we are not asserting that all emotions serve inherently social functions (e.g., recognizing fear in one's self could be adaptive for purely intra-personal reasons). However, we are suggesting that EA is an example of a number of social-cognitive functions that, because of the advantages they confer, could have driven cortical expansion allowing for domain-general abilities to learn and reflect on abstract concepts, and that emotion concept acquisition may have then conveyed more specific benefits within particular socio-cultural contexts. While the highest hierarchical levels of processing within the human brain may have been selected for due to pressures in a hypersocial niche, their domain generality could then also be applied to other non-social types of abstract and temporally extended cognition.

In support of our claim about EA in particular, a large body of work now suggests that subcortical circuits do not exist that map to emotion concepts in a 1-to-1 manner (Barrett 2017), that different emotion concepts are learned in different cultures (Russell 1991), and that self-reported emotional experiences correspond to the activation of domain-general cortical systems (Wager et al. 2015). Further, emotion recognition, emotional experience, emotion-focused attention, and emotion-focused working memory all appear to activate association cortex regions that have disproportionately expanded during human evolution. This includes regions of the DMN, ECN, SN, and LN (Gusnard et al. 2001; Peelen et al. 2010; Smith et al. 2014; Smith, Lane, et al. 2017; Smith, Lane, Alkozei, et al. 2018). Therefore, a range of empirical findings supports the idea that these domain-general multimodal networks – which plausibly implement the highest levels within the brain’s generative model – are those required for EA.

3.2 Links to life history strategy and other concepts in comparative neuroscience and biology

As briefly mentioned above, given the domain-generality of the explanation described above, it can also provide broad links to other psychological abilities discussed in relation to human evolution (reviewed in (Mars et al. 2017). Most straightforwardly, the broad explanation for the ability to learn emotion concepts above is very easily extended to learning other mental concepts that contribute to

human theory of mind abilities. As a second example, the human ability to learn second-order relations (and related abstract rules) can also be captured in terms of deeper generative models. Specifically, if one level learns each of the single relations, a higher level in the hierarchy would be required to learn the higher-order relations between them. Thus, expanded cortical levels between humans and other primates can straightforwardly account for this ability difference as well (at least in broad strokes). As a third example, the ability to simulate the outcomes of specific actions at extended timescales would also facilitate expanded tool-use abilities during human evolution (e.g., the ability to simulate what would happen if a spear were thrown at an animal).

Of special present interest is the construct of life history strategy (LHS) within evolutionary biology and psychology (Figueredo et al. 2005). As we will argue, the pressures driving the type of LHS in humans may help explain why humans constructed a hypersocial niche, which in turn selected for EA in our proposal. One broad way of characterizing LHS is as a description of the degree to which, both within and between species, an organism's behavior reflects an implicit assumption of environmental predictability or stability over long timescales. From an organism's point of view, the most salient environmental information is represented by the unavoidable extrinsic mortality faced by members of a community or species, and thus it is the principal determinant of LHS evolution (Charnov 1993). All LHSs reflect fundamental tradeoffs in the lifetime allocation of finite energy resources to either somatic growth and development or to

reproduction. It follows that populations or species in environments with, say, high unavoidable extrinsic mortality will evolve a LHS that is relatively ‘fast’ –that is, one with less investment in somatic growth, whose benefits are unlikely to be fully capitalized upon in the future, and more investment in reproduction early and often. A fast LHS is therefore one in which behavior reflects little assumption of long-term survival, and is therefore associated with bio-behavioral traits such as relatively earlier sexual maturation, shorter life span, and a larger number of offspring with little investment in each of those offspring. Each of these traits makes sense under the assumption that only proximal outcomes are predictable (e.g., why invest in only one offspring if it could die at any moment?). It follows from the work reviewed above that species with a fast LHS would have relatively little need for tracking long timescale regularities – corresponding to less cortical tissue in association areas and less reflective capacity relative to species with ‘slower’ LHS (Herculano-Houzel 2016). In contrast, a ‘slow’ LHS represents a fit to environments with relatively low levels of unavoidable extrinsic mortality, in which biology and behavior reflect strong commitments to long-term investments, such as in somatic growth and maintenance. This strategy consists of a relatively slower maturational rate with later sexual maturity, greater longevity, and smaller number of offspring with intense parental investment in each of those offspring. Again, this makes sense given information that entails that distal future outcomes are sufficiently reliable to “bet on” during action selection (i.e., so long as early death is not reliably predicted). Based on the perspective we provide above, species with a slow LHS would have relatively greater need for tracking long

timescale regularities – corresponding to greater amounts of cortical tissue in association areas and greater reflective capacity. This is consistent with the finding that for mammals the number of cortical neurons is highly correlated with slow LHS characteristics (e.g., slow maturation, longevity; (Herculano-Houzel 2016)).

Among primates, humans have a slow life history with unusual characteristics, such as slow maturation and longevity, but also characterized by a higher reproductive rate due to a shortened interbirth interval. These distinctive features are best accounted for principally by our extraordinary brain size and specifically our number of cortical neurons (Herculano-Houzel 2016). The main stimulus for human brain evolution, beginning around two million years ago with the appearance of the genus *Homo* in Africa, was the challenge posed by pulsatile climatic changes (e.g., severe cold) and its effects on local habitats and food resources that likely triggered migration to northern latitudes (Bailey & Geary 2009). Larger brains and increased body size, as evident in *Homo erectus*, necessitated a slower LHS, while also enhancing the socio-cognitive abilities for successfully addressing ecological and social challenges. Combined with a higher reproductive output (i.e., shorter interbirth interval), the larger brain and slowed development required our human ancestors to adopt cooperative breeding, as described earlier. Slow life history traits promote cultural learning and transmission, trans-generational economic investment (e.g., ‘grand-mothering’), and long-term social bonds among kin and non-kin. This slow LHS—

also described as a ‘high K’ strategy—is linked to a package of “prosocial” psychological traits (e.g., low risk-taking, higher executive functioning and emotional intelligence, see (Figueredo et al. 2006)) that manifest as high levels of social tolerance, cooperation and altruism, and adherence to social rules. In short, the evolution of a slow LHS allowed humans to construct the uniquely hypersocial niche we described earlier that, in turn, proved effective in facing the continuing socio-ecological challenges as humans expanded outside Africa into a diversity of climates and habitats. Lastly, human geographic expansion led to the evolution of plasticity in life history speed (or, degree of K), allowing populations or individuals within populations to adjust life history speed in accordance with salient local environmental clues received during development that are reliable indicators of environmental unpredictability (e.g., behavioral responses to a problem are not predictably successful or effective) and harshness (e.g., resource scarcity, pathogen prevalence; (Ellis et al. 2009)). Both environmental unpredictability and harshness, as stressors, concurrently affect rates of morbidity and mortality which, in turn, determines individual LHS (Brumbach et al. 2009).

While serving as a further explanatory variable for the creation of a hyper-social niche in humans (which in turn explains higher EA within humans in our proposed account), the developmental calibration of LHS within the human species also provides further explanatory power in understanding developmental contributions to EA. In the following section we relate within-species differences

in LHS to individual differences in EA. Before doing so, however, it is worth emphasizing that, even at a between-species level, this account makes strong predictions, many of which have not been empirically examined to our knowledge. For example, this explanation specifically predicts correlations between degree of complexity in social structure, degree of tool-use ability, degree of reflective capacity, and relative differences in cortical expansion within regions sub-serving multimodal processing.

4. Within-species differences in EA

4.1 Developmental calibration and reflective tendencies

The account described above can also provide insights regarding the origin of differences in EA within the human species (Lane et al. 2015; Smith, Killgore, et al. 2018). There are several empirical measures of individual differences in EA, including the Levels of Emotional Awareness Scale (LEAS; (Lane et al. 1990)), the Toronto Alexithymia Scale (TAS-20; (Bagby et al. 1994)), and others (Kashdan et al. 2015). Consistent with our account above, individual differences within EA using such measures have been associated primarily with large-scale association networks. This includes observed associations between individual differences in EA and functional connectivity within the DMN (Smith, Alkozei, et al. 2017), cortical thickness within the LN (Smith, Bajaj, et al. 2018), and functional activation within regions of the DMN, SN, and ECN (Lane et al. 1998;

Frewen et al. 2008; McRae et al. 2008; Tavares et al. 2011; Smith, Lane, et al. 2017; Smith, Lane, Alkozei, et al. 2018). There is also evidence that EA differences correspond to differences in experience during development (Smith, Quinlan, et al. 2019), and that EA can be improved with training (Farnam et al. 2014; Burger et al. 2016; Neumann et al. 2017; Lumley et al. 2017). Of particular note are findings that early life stress (e.g., neglect, emotional abuse) is associated with increased basal cortisol levels that may mediate an earlier maturation of amygdala-prefrontal connectivity (Gee et al. 2013; Callaghan & Tottenham 2016) or reduce the connectivity between these brain areas (Burghy et al. 2012) – altering the regulatory influence of the prefrontal cortex on the amygdala, which may lower emotional regulatory capacities (e.g., increasing impulsive behavior).

The early developmental/learning-related influences on EA suggest a calibration process. Here we suggest that this calibration process can be understood within the framework we proposed above. Specifically, while brain evolution provides the necessary conditions for EA, it does not provide sufficient conditions.

Through the lens of predictive coding and active inference models, developing high EA would further require that the relevant statistical regularities are in fact present and reliable within an individual's experience to learn from; otherwise, although one has the capacity to learn such regularities, the necessary signal would be absent (for explicit computational simulations supporting this, see (Smith, Parr, et al. 2019; Smith, Lane, et al. 2019)). In other words, one needs to

be exposed to the right patterns of sensory input (i.e., have the right set of experiences with the right low-level covariance relations) in order to arrive at a generative model with predictions corresponding to high EA. For example, this is consistent with work showing that children who grow up in orphanages, and lack sufficient personal interaction with adults that can model social/emotion concept use, display deficits in related socio-emotional and cognitive abilities later in life (Colvert et al. 2008).

Another more general aspect of early calibration has to do with learned regularities regarding the reliability of different levels of representation. Specifically, if a child's early experiences were highly unpredictable over longer timescales (e.g., low socioeconomic conditions, inconsistent parenting, etc.), then this calibration process would lead a child to learn priors favoring the assignment of low precision estimates to high-level generative model predictions – and specifically to priors for low transition precision. It follows that future regularities of this type in later life experience would tend to be ignored, corresponding to a generally reduced capacity for learning such regularities and the cognitive tendency to avoid temporarily extended reflection. Such an individual would therefore tend to behave relatively unreflectively in the context of affective responses, and would have difficulty conceptualizing these responses (i.e., they would have trouble understanding their own emotions and how to regulate them). This may explain previously observed links between a range of mental and physical health conditions and both EA (reviewed in (Lane et al.

2015)) and childhood trauma (Heim & Nemeroff 2001; Heim et al. 2008), both of which may correlate with high levels of early stress and learned uncertainty. Recently, active inference models have afforded explicit simulation of emotion concept learning and confirmed that a generative model has difficulty learning and inferring emotion concepts if beliefs about transition precision are too low – because integrating the necessary information to identify such concepts over time requires the assumption that an emotional pattern is stable over time (Smith, Parr, et al. 2019). Related simulation work has also reproduced interactions between emotion concept representation and working memory in a deep hierarchical model that were sufficient to reproduce high EA behavior – and that simulated the way low transition precision interferes with emotion-focused working memory (Smith, Lane, et al. 2019).

There is also work linking early calibration of abstract prior expectations to stress-induced epigenetic influences on the set points for neuromodulatory tone (e.g., serotonin, norepinephrine, dopamine, acetylcholine, etc.), which are thought to influence multiple types of reliability estimates (Clark et al. 2018)). This work has suggested, for example, that reduced sensitivity of the amygdala and hippocampus to circulating cortisol levels may represent neural correlates of maladaptive priors within mood disorders – as these regions have descending influences on neuromodulatory tone (McEwen 2000; Herman et al. 2002; Radley et al. 2011; Braithwaite et al. 2015). Early stress/unpredictability appears to epigenetically alter the expression of glucocorticoid receptors, which are present

in the amygdala and hippocampus (Morimoto et al. 1996), and which have been linked to the onset of mood disorders (Nantharat et al. 2015; Smart et al. 2015; Palma-Gudiel et al. 2016). Because high-level prior expectations are often most easily adjusted in the absence of previously learned expectations, very early childhood may represent the most sensitive or flexible period for calibration – an idea supported by the greater levels of epigenetic variability observed at particular early developmental periods (Heim & Binder 2012) and by work demonstrating the influence of maternal behavior on resulting epigenetic profiles (Weaver et al. 2004). It therefore appears plausible to suggest that unpredictable early environments may lead to high-level prior expectations for large amounts of environmental unpredictability – leading to low transition precision estimates (which may be mediated by the norepinephrine system; (Lawson et al. 2017)) and perhaps to low policy depth and less confidence in desired/expected outcomes of chosen actions (i.e., prior policy precision; likely mediated by the dopamine system; (Schwartenbeck et al. 2015)) – leading to less deliberation, reflection, and self-focused awareness, as well as less context-sensitive behavior.

However, an important overall take-away point of individual variation in life history speed, emotional awareness and reflective capacity is that the developmental life history adjustments (or calibrations) and their correlated neuropsychological traits that we have been describing are adaptive in nature in the sense that they lead to the best likely fitness outcome in the predicted future adult environment, whether harsh or benign (Ellis & Del Giudice 2014); however,

this can become maladaptive when the learned statistics of the environment in childhood do not match those in adulthood.

There are many untested predictions based on this within-species account as well, which may be important for guiding future research. For example, our explanation predicts that individuals with higher EA will also tend to have a slower LHS. It also entails that higher EA should be associated with the correlates of slow LHS, such as being more sensitive to long-term consequences, being less impulsive, etc. As we will discuss in the subsequent section, it also makes interesting predictions about the relationship between observed sex differences in EA and sex differences in parental investment patterns associated with LHS.

4.2 Sex differences, emotional awareness, and life history

There is a notable female advantage in both emotion recognition ability and in EA more broadly (e.g., see (Barrett et al. 2000; Ciarrochi et al. 2005; Kret & De Gelder 2012)). Interestingly, a recent study showed that differences in EA – an ability that depends on concept learning – mediates the relationship between sex and emotion recognition ability (Wright et al. 2017). This suggests that early learning may also play an important role in fostering these sex differences, as is consistent with studies showing that boys and girls tend to be treated in ways that encourage different norms for attending to and expressing emotions. For

example, across multiple cultures, parents tend to encourage control of negative emotions more in boys than in girls, and encourage expression of positive emotions more in girls than in boys, (Diener & Lucas 2004; Chaplin et al. 2005). Parents also tend to discuss emotions less with their sons than with their daughters during preschool years (Fivush et al. 2000).

Given that this appears to occur across cultures, the explanation for why women tend to learn greater EA than men may still have (at least) an indirect evolutionary contribution. A related finding that requires explanation is that variability in EA tends to be much lower in women than in men (Wright et al. 2017). Here we highlight that, just as there is greater parental investment (i.e., the tendency to devote more time and resources to individual offspring) in those with slower LHS, there are also evolutionary constraints that promote greater parental investment by women than by men, as well as greater variability in parental investment by men than by women. As mothers can only reproduce once every 9-10 months, whereas men can in principle impregnate many women within a period of days, parental investment theory makes the empirically confirmed prediction that women will be much more sexually discriminating to ensure offspring quality, given their much higher obligate parental investment (Figueredo et al. 2005). From an evolutionary and life history standpoint, women in effect have an overall slower life history than men (e.g., fewer gametes, fewer offspring, greater obligate parental investment, longer lifespan; (Figueredo et al. 2006)), which is consistent with their higher average levels of EA.

Likewise, given the unavoidable paternity uncertainty, women and their genetic kin will be much more likely to invest in raising their offspring than men. Given this asymmetry, women may have much less variability in their experience attending to the subtle psychological signals that carry information about the well-being of their children – providing stable opportunities to learn higher EA. Given that women who are better at this skill may also confer greater parenting quality, this may also create selective pressures in favor of this ability in women. In addition, women's likely long history of alloparenting (cooperative breeding) would have contributed to the selection for reduced variation in their EA skills and theory of mind capacity generally (Hrdy 1999). In contrast, while some men in a context-dependent way (e.g., when paternity confidence is high) invest heavily in their children, many others do not. This suggests less selective pressure toward EA-related abilities in men. It also predicts greater variability in such abilities in men, depending on their level of investment (e.g., time spent directly interacting with their children).

Outside the parenting context, there is also much evidence that women are in general more empathic and prosocial than men (Singer et al. 2006), a difference that is linked to the higher sensitivity in women of the dopaminergic reward system to prosocial rewards (Soutschek et al. 2017). Women also have higher synaptic concentrations of dopamine in the striatum and a stronger striatal response to prosocial decisions (reviewed in (Borland et al. 2019)). There are, of

course, also contexts where some men would benefit from reading social signals among coordinated groups of adults, however, which, given its likely dependence on cultural practices (e.g., fraternal interest groups in warring societies), could increase this variability even further. While we can't rule out experiential contributions to these gender differences, they nevertheless could help explain the male-female difference in EA. Thus, there may be genetic adaptations promoting these gender differences, but variable constraints on the range of experiences that are typically encountered by men and women may allow for different learning opportunities.

Based on these considerations, it will also be important for future research to specifically test the predicted relationship between parental investment, other life history characteristics, and EA. If this account is correct, then men with higher parental investment behavior should also tend to possess higher EA.

5. Discussion

The account we have provided represents a synthesis of many different bodies of work, including those on cognitive and computational neuroscience, socioemotional cognition, cognitive science, evolutionary psychology, and comparative anatomy, among others. The intersection of such fields represents a promising area for providing new insights into multi-level explanations for

understanding uniquely human capacities – and offers empirical directions for testing those explanations. We consider a few such directions below.

One important direction may be further study of the neuroscientific basis of LHS within cognitive and computational neuroscience. With respect to large-scale neural networks, higher reflective capacity under our account should be associated with specific adaptive interactions between the ECN (application of top-down maintenance/manipulation of represented information within working memory) and the DMN (concept representation and internal stimulation based on long-term memory). For example, in the context of a task in which internal simulation was necessary in goal-directed decision-making (Spreng et al. 2010), we would expect that slower LHS would be associated with stronger coupling between the ECN and the DMN than in those with faster LHS and that decision times would also be faster in those with faster LHS. This would parallel previous work demonstrating increased activation in both ECN and DMN regions when holding one's own emotions in working memory (and showing that greater DMN activation corresponded to higher emotional awareness; see (Smith, Lane, Alkozei, et al. 2018; Smith, Lane, Sanova, et al. 2018)).

With respect to computational neuroscience, it will also be important to test the prediction that individuals with low EA and fast LHS both show behavioral evidence for beliefs favoring low transition precision (a.k.a. high environmental volatility estimates) with respect to long timescale predictions. In previous work

(Lawson et al. 2017), transition precision estimates have been linked to central noradrenergic modulation, which can also be tested via neuroimaging and indirect measures of locus coeruleus activation (e.g., pupillary responses). Interestingly, this work has found estimates of higher volatility in individuals with autism, which is characterized by a reduced capacity for social cognition.

Yet another phenomenon worth further study is the role of genetic vs. environmental influences on gender differences in EA, LHS, and other measures of reflective social cognition. Twin studies have yielded high heritability estimates (60-70%) for a single life history factor, called the K –factor, and its associated life history variables (Figueredo et al. 2004), which indicates that a good portion of the variance among individuals in life history speed is due to genetic differences. In other words, one's genotype determines the range of plasticity in LHS adjustment to local conditions sampled during development. This could be taken to suggest that individual differences in EA should also be heritable. However, on our account differences in LHS can also facilitate/hinder opportunities to learn from emotional signals – therefore suggesting an account in which a genetic predisposition facilitates emotion concept learning and emotion-focused reflection. Future heritability and developmental studies will need to test these different possible explanations. Also open to further study is the relationship between individual differences in EA and differences in neurochemistry (e.g., serotonergic or oxytocinergic systems), especially given the effects of early life stress on these systems and EA (it is also noteworthy that

women appear more sensitive to early life stress than men, as evidenced by higher basal cortisol levels in childhood (Burghy et al. 2012)). However, we want to reiterate an important conclusion drawn from our evolutionary life history perspective, namely, that higher vs. lower EA is neither better nor worse, but rather each adaptively fits the individual's developmental niche.

The degree to which gender differences are due to socialization vs. due to gender-specific selective pressures during evolution is also important to test in future work. It could be evaluated, for example, by identifying cultures in which gender roles differ significantly from those in Western cultures and testing whether EA varies with culturally defined gender roles.

Finally, while we have suggested that selective pressures due to hypersociality were the primary driver of expansions in reflective capacity that facilitated EA, this will also have to be borne out in future research. On our account, the adaptations were largely domain-general, and then were applied and further refined for different planning and relational tracking purposes (e.g., tool use, learning higher-order relations, multistep logical reasoning, etc.). But it should be clear that the exact magnitude of contribution of each of these adaptive cognitive advances is not clear, and the degree to which domain-general reflective capacities vs. domain-specific prepared learning mechanisms discussed each contribute to EA is also an open question.

In closing, it is important to emphasize the limitations of the account we propose. First, we do not claim that it is a complete account. While we believe that interactions between concept learning and working memory capacity, when expanded to include the more abstract, long timescale regularities that vastly surpass that of other animals, can provide a parsimonious account of many examples of goal-directed reflection, there are other phenomena that also arguably fall under the rubric of “reflective” and that show expanded capacity in humans relative to other animals. One example could be more bottom-up driven types of internal simulation, associated with mind wandering or rumination, in which there is increased activation of the DMN but reduced activation of ECN (Buckner et al. 2008). This type of internal simulation does depend on representing longtime scale regularities, but it is not goal-driven and relates less directly to cognitive control in decision-making. That said, this type of automatic simulation process has been linked to a type of Bayesian model reduction process within computational models that can facilitate subsequent insight and creative problem-solving (Friston, Lin, et al. 2017) – which is not an element explicitly discussed in our account. Surely, there are also other phenomena as well that some might consider as part of expanded reflective capacity in humans and that may not be fully accounted for in our proposal. With these limitations in mind, however, we believe our integrative account has broad and unifying explanatory power that ultimately arises from relatively few basic building blocks. It holds the promise of a more unified account of the unique human capacity for EA spanning across many fields of research, and holds the promise of inspiring

more interdisciplinary, collaborative research projects to further advance knowledge in this area.

References

- Adams, R., Shipp, S. & Friston, K. (2013): Predictions not commands: active inference in the motor system. – *Brain Structure and Function*, 218 (3): 611–643 <https://doi.org/10.1007/s00429-012-0475-5>.
- Amodio, D.M. & Frith, C.D. (2006): Meeting of minds: the medial frontal cortex and social cognition. – *Nature Reviews Neuroscience*, 7 (4): 268–277.
- Apicella, C.L. & Silk, J.B. (2019): The evolution of human cooperation. – *Current Biology*, 29 (11): R447–R450 <https://doi.org/10.1016/j.cub.2019.03.036>.
- Aviezer, H., Hassin, R., Ryan, J., Grady, C., Susskind, J., Anderson, A., Moscovitch, M. & Bentin, S. (2008): Angry, Disgusted, or Afraid? – *Psychological Science*, 19 (7): 724–732 <https://doi.org/10.1111/j.1467-9280.2008.02148.x>.
- Badcock, P.B., Davey, C.G., Whittle, S., Allen, N.B. & Friston, K.J. (2017): The Depressed Brain: An Evolutionary Systems Theory. – *Trends in Cognitive Sciences*, 21 (3): 182–194 <https://doi.org/10.1016/J.TICS.2017.01.005>.
- Bagby, R., Parker, J. & Taylor, G. (1994): The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. – *Journal of psychosomatic research*, 38 (1): 23–32.
- Bailey, D.H. & Geary, D.C. (2009): Hominid Brain Evolution. – *Human Nature*, 20

(1): 67–79 <https://doi.org/10.1007/s12110-008-9054-0>.

Barrett, L. (2017): How emotions are made: The secret life of the brain. –

(Houghton Mifflin Harcourt) New York.

Barrett, L., Lane, R., Sechrest, L. & Schwartz, G. (2000): Sex Differences in

Emotional Awareness. – *Personality and Social Psychology Bulletin*, 26 (9):

1027–1035 <https://doi.org/10.1177/01461672002611001>.

Barrett, L., Mesquita, B. & Gendron, M. (2011): Context in Emotion Perception. –

Current Directions in Psychological Science, 20 (5): 286–290

<https://doi.org/10.1177/0963721411422522>.

Barrett, L. & Satpute, A. (2013): Large-scale brain networks in affective and

social neuroscience: towards an integrative functional architecture of the

brain. – *Current opinion in neurobiology*, 23 (3): 361–72

<https://doi.org/10.1016/j.conb.2012.12.012>.

Barrett, L. & Simmons, W. (2015): Interoceptive predictions in the brain. – *Nature*

reviews. Neuroscience, 16 (7): 419–29 <https://doi.org/10.1038/nrn3950>.

Bastos, A., Usrey, W., Adams, R., Mangun, G., Fries, P. & Friston, K. (2012):

Canonical microcircuits for predictive coding. – *Neuron*, 76 (4): 695–711

<https://doi.org/10.1016/j.neuron.2012.10.038>.

Bickart, K.C., Wright, C.I., Dautoff, R.J., Dickerson, B.C. & Barrett, L.F. (2011):

Amygdala volume and social network size in humans. – *Nature*

neuroscience, 14 (2): 163–4 <https://doi.org/10.1038/nn.2724>.

Binder, J., Desai, R., Graves, W. & Conant, L. (2009): Where Is the Semantic

System? A Critical Review and Meta-Analysis of 120 Functional

Neuroimaging Studies. – *Cerebral Cortex*, 19 (12): 2767–2796

<https://doi.org/10.1093/cercor/bhp055>.

Borland, J.M., Aiani, L.M., Norvelle, A., Grantham, K.N., O’Laughlin, K.,

Terranova, J.I., Frantz, K.J. & Albers, H.E. (2019): Sex-dependent regulation of social reward by oxytocin receptors in the ventral tegmental area. –

Neuropsychopharmacology, 44 (4): 785–792 <https://doi.org/10.1038/s41386-018-0262-y>.

Bowles, S. (2006): Group competition, reproductive leveling, and the evolution of human altruism. – *Science*, 314 (5805): 1569–72

<https://doi.org/10.1126/science.1134829>.

Braga, R. & Buckner, R. (2017): Parallel Interdigitated Distributed Networks

within the Individual Estimated by Intrinsic Functional Connectivity. –

Neuron, 95 (2): 457-471.e5 <https://doi.org/10.1016/j.neuron.2017.06.038>.

Braithwaite, E., Kundakovic, M., Ramchandani, P., Murphy, S. & Champagne, F.

(2015): Maternal prenatal depressive symptoms predict infant *NR3C1* 1F and *BDNF* IV DNA methylation. – *Epigenetics*, 10 (5): 408–417

<https://doi.org/10.1080/15592294.2015.1039221>.

Brown, S.L. & Cialdini, R. (2015): Functional Motives and Functional

Consequences of Prosocial Behavior. – In: SCHROEDER, D.A. & GRAZIANO,

W.G.: *The Oxford Handbook of Prosocial Behavior*. – pp. 346–361 (Oxford

University Press) <https://doi.org/10.1093/oxfordhb/9780195399813.013.025>.

Brumbach, B.H., Figueredo, A.J. & Ellis, B.J. (2009): Effects of Harsh and

Unpredictable Environments in Adolescence on Development of Life History

- Strategies: A Longitudinal Test of an Evolutionary Model. – *Human nature* (Hawthorne, N.Y.), 20 (1): 25 <https://doi.org/10.1007/S12110-009-9059-3>.
- Buckner, R., Andrews-Hanna, J.R.J. & Schacter, D.L.D. (2008): The Brain's Default Network. – *Annals of the New York Academy of Sciences*, 1124 (1) <https://doi.org/10.1196/annals.1440.011>.
- Buckner, R. & Krienen, F. (2013): The evolution of distributed association networks in the human brain. – *Trends in Cognitive Sciences*, 17 (12): 648–665 <https://doi.org/10.1016/j.tics.2013.09.017>.
- Burger, A., Lumley, M., Carty, J., Latsch, D., Thakur, E., Hyde-Nolan, M., Hijazi, A. & Schubiner, H. (2016): The effects of a novel psychological attribution and emotional awareness and expression therapy for chronic musculoskeletal pain: A preliminary, uncontrolled trial. – *Journal of psychosomatic research*, 81: 1–8 <https://doi.org/10.1016/j.jpsychores.2015.12.003>.
- Burghy, C.A., Stodola, D.E., Ruttle, P.L., Molloy, E.K., Armstrong, J.M., Oler, J.A., Fox, M.E., Hayes, A.S., Kalin, N.H., Essex, M.J., Davidson, R.J. & Birn, R.M. (2012): Developmental pathways to amygdala-prefrontal function and internalizing symptoms in adolescence. – *Nature Neuroscience*, 15 (12): 1736–1741 <https://doi.org/10.1038/nn.3257>.
- Burkart, J.M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A., Huber, J., Isler, K., Kosonen, Z.K., Martins, E., Meulman, E.J., Richiger, R., Rueth, K., Spillmann, B., Wiesendanger, S. & van Schaik, C.P. (2014): The evolutionary origin of human hyper-cooperation. – *Nature Communications*,

5 (1): 4747 <https://doi.org/10.1038/ncomms5747>.

Callaghan, B.L. & Tottenham, N. (2016): The Stress Acceleration Hypothesis: Effects of early-life adversity on emotion circuits and behavior. – *Current opinion in behavioral sciences*, 7: 76–81
<https://doi.org/10.1016/j.cobeha.2015.11.018>.

Chang, S.W.C., Brent, L.J.N., Adams, G.K., Klein, J.T., Pearson, J.M., Watson, K.K. & Platt, M.L. (2013): Neuroethology of primate social behavior. – *Proceedings of the National Academy of Sciences of the United States of America*, 110 Suppl 2 (Suppl 2): 10387–94
<https://doi.org/10.1073/pnas.1301213110>.

Chaplin, T., Cole, P. & Zahn-Waxler, C. (2005): Parental Socialization of Emotion Expression: Gender Differences and Relations to Child Adjustment. – *Emotion*, 5 (1): 80–88 <https://doi.org/10.1037/1528-3542.5.1.80>.

Charnov, E. (1993): Life history invariants: some explanations of symmetry in evolutionary ecology. – (Oxford University Press) Oxford.

Ciarrochi, J., Hynes, K. & Crittenden, N. (2005): Can men do better if they try harder: Sex and motivational effects on emotional awareness. – *Cognition & Emotion*, 19 (1): 133–141 <https://doi.org/10.1080/02699930441000102>.

Clark, A. (2015): Surfing uncertainty: Prediction, action, and the embodied mind. – (Oxford University Press) New York.

Clark, J., Watson, S. & Friston, K. (2018): What is mood? A computational perspective. – *Psychological Medicine*: 1–8
<https://doi.org/10.1017/S0033291718000430>.

- Colvert, E., Rutter, M., Kreppner, J., Beckett, C., Castle, J., Groothues, C., Hawkins, A., Stevens, S. & Sonuga-Barke, E. (2008): Do Theory of Mind and Executive Function Deficits Underlie the Adverse Outcomes Associated with Profound Early Deprivation?: Findings from the English and Romanian Adoptees Study. – *Journal of Abnormal Child Psychology*, 36 (7): 1057–1068 <https://doi.org/10.1007/s10802-008-9232-x>.
- Corbetta, M. & Shulman, G. (2002): Control of goal-directed and stimulus-driven attention in the brain. – *Nature reviews. Neuroscience*, 3 (3): 201–15 <https://doi.org/10.1038/nrn755>.
- Craig, A.D. (2009): How do you feel--now? The anterior insula and human awareness. – *Nature Reviews Neuroscience*, 10 (1): 59–70.
- Diaconescu, A., Mathys, C., Weber, L., Kasper, L., Mauer, J. & Stephan, K. (2017): Hierarchical prediction errors in midbrain and septum during social learning. – *Social Cognitive and Affective Neuroscience*, 12 (4): nsw171 <https://doi.org/10.1093/scan/nsw171>.
- Diener, M. & Lucas, R. (2004): Adults Desires for Childrens Emotions across 48 Countries: Associations with Individual and National Characteristics. – *Journal of Cross-Cultural Psychology*, 35 (5): 525–547 <https://doi.org/10.1177/0022022104268387>.
- Doucet, G., Naveau, M., Petit, L., Delcroix, N., Zago, L., Crivello, F., Jobard, G., Tzourio-Mazoyer, N., Mazoyer, B., Mellet, E. & Joliot, M. (2011): Brain activity at rest: a multiscale hierarchical functional organization. – *Journal of Neurophysiology*, 105 (6): 2753–2763 <https://doi.org/10.1152/jn.00895.2010>.

- De Dreu, C.K.W., Greer, L.L., Handgraaf, M.J.J., Shalvi, S., Van Kleef, G.A., Baas, M., Ten Velden, F.S., Van Dijk, E. & Feith, S.W.W. (2010): The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. – *Science*, 328 (5984): 1408–11
<https://doi.org/10.1126/science.1189047>.
- Dunbar, R.I.M. (1998): The social brain hypothesis. – *Evolutionary Anthropology: Issues, News, and Reviews*, 6 (5): 178–190
[https://doi.org/10.1002/\(SICI\)1520-6505\(1998\)6:5<178::AID-EVAN5>3.0.CO;2-8](https://doi.org/10.1002/(SICI)1520-6505(1998)6:5<178::AID-EVAN5>3.0.CO;2-8).
- Ellis, B.J., Figueredo, A.J., Brumbach, B.H. & Schlomer, G.L. (2009): Fundamental Dimensions of Environmental Risk. – *Human Nature*, 20 (2): 204–268 <https://doi.org/10.1007/s12110-009-9063-7>.
- Ellis, B.J. & Del Giudice, M. (2014): Beyond allostatic load: Rethinking the role of stress in regulating human development. – *Development and Psychopathology*, 26 (1): 1–20 <https://doi.org/10.1017/S0954579413000849>.
- Falk, D. (2016): Evolution of brain and culture: the neurological and cognitive journey from Australopithecus to Albert Einstein. – *Journal of anthropological sciences*, 94: 99–111 <https://doi.org/10.4436/JASS.94027>.
- Farnam, A., Somi, M., Farhang, S., Mahdavi, N. & Ali Besharat, M. (2014): The therapeutic effect of adding emotional awareness training to standard medical treatment for irritable bowel syndrome: a randomized clinical trial. – *Journal of psychiatric practice*, 20 (1): 3–11
<https://doi.org/10.1097/01.pra.0000442934.38704.3a>.

- Felleman, D. & Van Essen, D. (1991): Distributed hierarchical processing in the primate cerebral cortex. – *Cerebral cortex*, 1 (1): 1–47.
- Figueredo, A.J., Vásquez, G., Brumbach, B.H. & Schneider, S.M.R. (2004): The heritability of life history strategy: The k-factor, covitality, and personality. – *Biodemography and Social Biology*, 51 (3–4): 121–143
<https://doi.org/10.1080/19485565.2004.9989090>.
- Figueredo, A.J., Vásquez, G., Brumbach, B.H., Schneider, S.M.R., Sefcek, J.A., Tal, I.R., Hill, D., Wenner, C.J. & Jacobs, W.J. (2006): Consilience and Life History Theory: From genes to brain to reproductive strategy. – *Developmental Review*, 26 (2): 243–275
<https://doi.org/10.1016/J.DR.2006.02.002>.
- Figueredo, A.J., Vásquez, G., Brumbach, B.H., Sefcek, J.A., Kirsner, B.R. & Jacobs, W.J. (2005): The K-factor: Individual differences in life history strategy. – *Personality and Individual Differences*, 39 (8): 1349–1360
<https://doi.org/10.1016/J.PAID.2005.06.009>.
- Finlay, B.L. (2019): Human exceptionalism, our ordinary cortex and our research futures. – *Developmental Psychobiology*, 61 (3): 317–322
<https://doi.org/10.1002/dev.21838>.
- FitzGerald, T., Dolan, R. & Friston, K. (2014): Model averaging, optimal inference, and habit formation. – *Frontiers in Human Neuroscience*, 8: 457
<https://doi.org/10.3389/fnhum.2014.00457>.
- Fivush, R., Brotman, M., Buckner, J. & Goodman, S. (2000): Gender Differences in Parent–Child Emotion Narratives. – *Sex Roles*, 42 (3): 233–253

<https://doi.org/10.1023/A:1007091207068>.

- Frewen, P., Lane, R., Neufeld, R., Densmore, M., Stevens, T. & Lanius, R. (2008): Neural correlates of levels of emotional awareness during trauma script-imagery in posttraumatic stress disorder. – *Psychosomatic Medicine*, 70 (1): 27–31.
- Friston, K. (2005): A theory of cortical responses. – *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 360 (1456): 815–36 <https://doi.org/10.1098/rstb.2005.1622>.
- Friston, K. (2010): The free-energy principle: a unified brain theory? – *Nature reviews. Neuroscience*, 11 (2): 127–38 <https://doi.org/10.1038/nrn2787>.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O Doherty, J. & Pezzulo, G. (2016): Active inference and learning. – *Neuroscience and biobehavioral reviews*, 68: 862–879 <https://doi.org/10.1016/j.neubiorev.2016.06.022>.
- Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. & Pezzulo, G. (2017): Active Inference: A Process Theory. – *Neural Computation*, 29 (1): 1–49 https://doi.org/10.1162/NECO_a_00912.
- Friston, K., Lin, M., Frith, C., Pezzulo, G., Hobson, J. & Ondobaka, S. (2017): Active Inference, Curiosity and Insight. – *Neural Computation*, 29 (10): 2633–2683 https://doi.org/10.1162/neco_a_00999.
- Friston, K., Parr, T. & de Vries, B. (2017): The graphical brain: Belief propagation and active inference. – *Network Neuroscience*, 1 (4): 381–414 https://doi.org/10.1162/NETN_a_00018.

- Frith, C. & Frith, U. (2006): The neural basis of mentalizing. – *Neuron*, 50 (4): 531–534.
- Frith, C. & Frith, U. (2012): Mechanisms of social cognition. – *Annual Review of Psychology*, 63: 287–313.
- Gee, D.G., Gabard-Durnam, L.J., Flannery, J., Goff, B., Humphreys, K.L., Telzer, E.H., Hare, T.A., Bookheimer, S.Y. & Tottenham, N. (2013): Early developmental emergence of human amygdala-prefrontal connectivity after maternal deprivation. – *Proceedings of the National Academy of Sciences of the United States of America*, 110 (39): 15638–43
<https://doi.org/10.1073/pnas.1307893110>.
- González-Forero, M. & Gardner, A. (2018): Inference of ecological and social drivers of human brain-size evolution. – *Nature*, 557 (7706): 554–557
<https://doi.org/10.1038/s41586-018-0127-x>.
- Gusnard, D., Akbudak, E., Shulman, G. & Raichle, M. (2001): Medial prefrontal cortex and self-referential mental activity: relation to a default mode of brain function. – *Proceedings of the National Academy of Sciences*, 98 (7): 4259–4264.
- Haker, H., Schneebeli, M. & Stephan, K. (2016): Can Bayesian Theories of Autism Spectrum Disorder Help Improve Clinical Practice? – *Frontiers in Psychiatry*, 7: 107 <https://doi.org/10.3389/fpsyt.2016.00107>.
- Hare, B. (2017): Survival of the Friendliest: *Homo sapiens* Evolved via Selection for Prosociality. – *Annual Review of Psychology*, 68 (1): 155–186
<https://doi.org/10.1146/annurev-psych-010416-044201>.

- Hasson, U., Chen, J. & Honey, C. (2015): Hierarchical process memory: memory as an integral component of information processing. – *Trends in Cognitive Sciences*, 19 (6): 304–313 <https://doi.org/10.1016/j.tics.2015.04.006>.
- Hasson, U., Yang, E., Vallines, I., Heeger, D. & Rubin, N. (2008): A Hierarchy of Temporal Receptive Windows in Human Cortex. – *Journal of Neuroscience*, 28 (10): 2539–2550 <https://doi.org/10.1523/JNEUROSCI.5487-07.2008>.
- Heim, C. & Binder, E. (2012): Current research trends in early life stress and depression: Review of human studies on sensitive periods, gene–environment interactions, and epigenetics. – *Experimental Neurology*, 233 (1): 102–111 <https://doi.org/10.1016/j.expneurol.2011.10.032>.
- Heim, C. & Nemeroff, C. (2001): The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. – *Biological Psychiatry*, 49 (12): 1023–1039 [https://doi.org/10.1016/S0006-3223\(01\)01157-X](https://doi.org/10.1016/S0006-3223(01)01157-X).
- Heim, C., Newport, D., Mletzko, T., Miller, A. & Nemeroff, C. (2008): The link between childhood trauma and depression: Insights from HPA axis studies in humans. – *Psychoneuroendocrinology*, 33 (6): 693–710 <https://doi.org/10.1016/j.psyneuen.2008.03.008>.
- Henrich, J. (2015): The secret of our success: How culture is driving human evolution, domesticating our species, and making a smart. – (Princeton University Press) Princeton, NJ.
- Herculano-Houzel, S. (2009): The human brain in numbers: a linearly scaled-up primate brain. – *Frontiers in Human Neuroscience*, 3: 31

<https://doi.org/10.3389/neuro.09.031.2009>.

Herculano-Houzel, S. (2016): The human advantage. – (MIT Press) Boston, MA.

Herman, J.P., Tasker, J.G., Ziegler, D.R. & Cullinan, W.E. (2002): Local circuit regulation of paraventricular nucleus stress integration. – *Pharmacology Biochemistry and Behavior*, 71 (3): 457–468 [https://doi.org/10.1016/S0091-3057\(01\)00681-5](https://doi.org/10.1016/S0091-3057(01)00681-5).

Hermundstad, A., Bassett, D., Brown, K., Aminoff, E., Clewett, D., Freeman, S., Frithsen, A., Johnson, A., Tipper, C., Miller, M., Grafton, S. & Carlson, J. (2013): Structural foundations of resting-state and task-based functional connectivity in the human brain. – *Proceedings of the National Academy of Sciences of the United States of America*, 110 (15): 6169–74 <https://doi.org/10.1073/pnas.1219562110>.

van den Heuvel, M. & Sporns, O. (2011): Rich-club organization of the human connectome. – *The Journal of neuroscience*, 31 (44): 15775–86 <https://doi.org/10.1523/JNEUROSCI.3539-11.2011>.

Hoffman, K.L. & Gauthier, I. (2007): Evolutionary Specializations for Processing Faces and Objects. – In: *Evolution of Nervous Systems*. – pp. 437–445 (Elsevier) <https://doi.org/10.1016/B0-12-370878-8/00031-8>.

Hrdy, S. (1999): Mother nature: a history of mothers, infants and natural selection. – (Pantheon) New York.

Iglesias, S., Mathys, C., Brodersen, K., Kasper, L., Piccirelli, M., den Ouden, H. & Stephan, K. (2013): Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. – *Neuron*, 80 (2): 519–530

<https://doi.org/10.1016/J.NEURON.2013.09.009>.

Isler, K. & van Schaik, C.P. (2012): How Our Ancestors Broke through the Gray Ceiling. – *Current Anthropology*, 53 (S6): S453–S465

<https://doi.org/10.1086/667623>.

Johnson, M.H. (2011): Interactive Specialization: A domain-general framework for human functional brain development? – *Developmental Cognitive Neuroscience*, 1 (1): 7–21 <https://doi.org/10.1016/J.DCN.2010.07.003>.

Kashdan, T., Barrett, L. & McKnight, P. (2015): Unpacking Emotion Differentiation: Transforming Unpleasant Experience by Perceiving Distinctions in Negativity. – *Current Directions in Psychological Science*, 24 (1): 10–16 <https://doi.org/10.1177/0963721414550708>.

Kiebel, S., Daunizeau, J. & Friston, K. (2008): A Hierarchy of Time-Scales and the Brain. – *PLoS Computational Biology*, 4 (11): e1000209 <https://doi.org/10.1371/journal.pcbi.1000209>.

Kingstone, A., Laidlaw, K.E.W., Nasiopoulos, E. & Risko, E.F. (2017): Cognitive Ethology and Social Attention. – *On Human Nature*: 365–382 <https://doi.org/10.1016/B978-0-12-420190-3.00022-3>.

Kircanski, K., Lieberman, M. & Craske, M. (2012): Feelings into words: contributions of language to exposure therapy. – *Psychological science*, 23 (10): 1086–91 <https://doi.org/10.1177/0956797612443830>.

Kissel, M. & Kim, N.C. (2019): The emergence of human warfare: Current perspectives. – *American Journal of Physical Anthropology*, 168 (S67): 141–163 <https://doi.org/10.1002/ajpa.23751>.

- Kret, M. & De Gelder, B. (2012): A review on sex differences in processing emotional signals. – *Neuropsychologia*, 50 (7): 1211–21
<https://doi.org/10.1016/j.neuropsychologia.2011.12.022>.
- Krienen, F. & Buckner, R. (2017): Human association cortex: expanded, untethered, neotenous, and plastic. – In: KAAS, J.: *Evolution of Nervous Systems 2e*, Volume 4. – pp. 169–183 (Elsevier).
- Laland, K.N. & Rendell, L. (2017): Social Evolution and the Collective Brain. – *Trends in Ecology & Evolution*, 32 (9): 625–626
<https://doi.org/10.1016/j.tree.2017.06.005>.
- Lane, R., Quinlan, D., Schwartz, G., Walker, P. & Zeitlin, S. (1990): The Levels of Emotional Awareness Scale: a cognitive-developmental measure of emotion. – *Journal of personality assessment*, 55 (1–2): 124–34
<https://doi.org/10.1080/00223891.1990.9674052>.
- Lane, R., Reiman, E., Axelrod, B., Yun, L., Holmes, A. & Schwartz, G. (1998): Neural correlates of levels of emotional awareness. Evidence of an interaction between emotion and attention in the anterior cingulate cortex. – *Journal of Cognitive Neuroscience*, 10 (4): 525–535.
- Lane, R., Weihs, K., Herring, A., Hishaw, A. & Smith, R. (2015): Affective agnosia: Expansion of the alexithymia construct and a new opportunity to integrate and extend Freud's legacy. – *Neuroscience and biobehavioral reviews*, 55: 594–611 <https://doi.org/10.1016/j.neubiorev.2015.06.007>.
- Lawson, R., Mathys, C. & Rees, G. (2017): Adults with autism overestimate the volatility of the sensory environment. – *Nature Neuroscience*, 20 (9): 1293–

1299 <https://doi.org/10.1038/nn.4615>.

LeDoux, J. (2012): Rethinking the emotional brain. – *Neuron*, 73 (4): 653–76

<https://doi.org/10.1016/j.neuron.2012.02.004>.

Leppänen, J.M. & Nelson, C.A. (2009): Tuning the developing brain to social signals of emotions. – *Nature Reviews Neuroscience*, 10 (1): 37–47

<https://doi.org/10.1038/nrn2554>.

Lew, C. & Semendeferi, K. (2017): Evolutionary Specializations of the Human Limbic System. – In: KAAS, J.: *Evolution of Nervous Systems 2e*, Volume 4. – pp. 277–291 (Elsevier).

Lew, C.H., Hanson, K.L., Groeniger, K.M., Greiner, D., Cuevas, D., Hrvoj-Mihic, B., Schumann, C.M. & Semendeferi, K. (2019): Serotonergic innervation of the human amygdala and evolutionary implications. – *American Journal of Physical Anthropology*: ajpa.23896 <https://doi.org/10.1002/ajpa.23896>.

Lewis, P.A., Rezaie, R., Brown, R., Roberts, N. & Dunbar, R.I.M. (2011): Ventromedial prefrontal volume predicts understanding of others and social network size. – *NeuroImage*, 57 (4): 1624–1629
<https://doi.org/10.1016/J.NEUROIMAGE.2011.05.030>.

Linson, A., Clark, A., Ramamoorthy, S. & Friston, K. (2018): The Active Inference Approach to Ecological Perception: General Information Dynamics for Natural and Artificial Embodied Cognition. – *Frontiers in Robotics and AI*, 5: 21 <https://doi.org/10.3389/frobt.2018.00021>.

Lumley, M., Schubiner, H., Lockhart, N., Kidwell, K., Harte, S., Clauw, D. & Williams, D. (2017): Emotional awareness and expression therapy, cognitive

behavioral therapy, and education for fibromyalgia: a cluster-randomized controlled trial. – *Pain*, 158 (12): 2354–2363

<https://doi.org/10.1097/j.pain.0000000000001036>.

Margulies, D., Ghosh, S., Goulas, A., Falkiewicz, M., Huntenburg, J., Langs, G.,

Bezgin, G., Eickhoff, S., Castellanos, F., Petrides, M., Jefferies, E. &

Smallwood, J. (2016): Situating the default-mode network along a principal gradient of macroscale cortical organization. – *Proceedings of the National Academy of Sciences*, 113 (44): 12574–12579

<https://doi.org/10.1073/pnas.1608282113>.

Margulies, D., Vincent, J., Kelly, C., Lohmann, G., Uddin, L., Biswal, B., Villringer,

A., Castellanos, F., Milham, M. & Petrides, M. (2009): Precuneus shares

intrinsic functional architecture in humans and monkeys. – *Proceedings of the National Academy of Sciences*, 106 (47): 20069–20074

<https://doi.org/10.1073/pnas.0905314106>.

Mars, R., Passingham, R., Neubert, F.-X., Verhagen, L. & Sallet, J. (2017):

Evolutionary Specializations of Human Association Cortex. – In: KAAS, J.:
Evolution of Nervous Systems 2e, Volume 4. – pp. 185–205 (Elsevier)

<https://doi.org/10.1016/B978-0-12-804042-3.00118-4>.

<https://doi.org/10.1016/B978-0-12-804042-3.00118-4>.

Mayhew, J.A. & Gómez, J.-C. (2015): Gorillas with white sclera: A naturally

occurring variation in a morphological trait linked to social cognitive

functions. – *American Journal of Primatology*, 77 (8): 869–877

<https://doi.org/10.1002/ajp.22411>.

McEwen, B. (2000): Allostasis and Allostatic Load Implications for

Neuropsychopharmacology. – Neuropsychopharmacology, 22 (2): 108–124
[https://doi.org/10.1016/S0893-133X\(99\)00129-3](https://doi.org/10.1016/S0893-133X(99)00129-3).

McLaughlin, K.A., Sheridan, M.A. & Nelson, C.A. (2017): Neglect as a Violation of Species-Expectant Experience: Neurodevelopmental Consequences. – Biological psychiatry, 82 (7): 462–471
<https://doi.org/10.1016/j.biopsych.2017.02.1096>.

McRae, K., Reiman, E., Fort, C., Chen, K. & Lane, R. (2008): Association between trait emotional awareness and dorsal anterior cingulate activity during emotion is arousal-dependent. – NeuroImage, 41 (2): 648–55
<https://doi.org/10.1016/j.neuroimage.2008.02.030>.

Mesulam, M. (1998): From sensation to cognition. – Brain, 121: 1013–1052.

Mesulam, M. (2000): Behavioral neuroanatomy: large-scale networks, association cortex, frontal syndromes, the limbic system, and hemispheric specializations. – In: MESULAM, M.: Principles of behavioral and cognitive neurology. 2nd Edition. – p. (Oxford University Press) New York.

Mirza, M., Adams, R., Parr, T. & Friston, K. (2019): Impulsivity and Active Inference. – Journal of Cognitive Neuroscience, 31 (2): 202–220
https://doi.org/10.1162/jocn_a_01352.

Miskovic, V. & Schmidt, L.A. (2012): Social fearfulness in the human brain. – Neuroscience & Biobehavioral Reviews, 36 (1): 459–478
<https://doi.org/10.1016/J.NEUBIOREV.2011.08.002>.

Morimoto, M., Morita, N., Ozawa, H., Yokoyama, K. & Kawata, M. (1996): Distribution of glucocorticoid receptor immunoreactivity and mRNA in the rat

brain: an immunohistochemical and in situ hybridization study. –

Neuroscience Research, 26 (3): 235–269 [https://doi.org/10.1016/S0168-0102\(96\)01105-4](https://doi.org/10.1016/S0168-0102(96)01105-4).

Moriuchi, J., Klin, A. & Jones, W. (2017): Mechanisms of Diminished Attention to Eyes in Autism. – American Journal of Psychiatry, 174 (1): 26–35
<https://doi.org/10.1176/appi.ajp.2016.15091222>.

Murray, J., Bernacchia, A., Freedman, D., Romo, R., Wallis, J., Cai, X., Padoa-Schioppa, C., Pasternak, T., Seo, H., Lee, D. & Wang, X.-J. (2014): A hierarchy of intrinsic timescales across primate cortex. – Nature Neuroscience, 17 (12): 1661–1663 <https://doi.org/10.1038/nn.3862>.

Nantharat, M., Wanitchanon, T., Amesbutr, M., Tammachote, R. & Praphanphoj, V. (2015): Glucocorticoid receptor gene (NR3C1) promoter is hypermethylated in Thai females with major depressive disorder. – Genetics and Molecular Research, 14 (4): 19071–19079
<https://doi.org/10.4238/2015.December.29.15>.

Neumann, D., Malec, J. & Hammond, F. (2017): Reductions in Alexithymia and Emotion Dysregulation After Training Emotional Self-Awareness Following Traumatic Brain Injury. – Journal of Head Trauma Rehabilitation: 1
<https://doi.org/10.1097/HTR.0000000000000277>.

Nielsen, M. (2012): Imitation, pretend play, and childhood: Essential elements in the evolution of human culture? – Journal of Comparative Psychology, 126 (2): 170–181 <https://doi.org/10.1037/a0025168>.

Palma-Gudiel, H., Cordova-Palomera, A., Marquès Feixa, L., Cirera Miquel, F. &

- Fañanás, L. (2016): Epigenetic signature of glucocorticoid receptor is associated with the familial component of depression: A twin-based study. – *European Psychiatry*, 33: S79 <https://doi.org/10.1016/j.eurpsy.2016.01.025>.
- Panksepp, J., Lane, R., Solms, M. & Smith, R. (2017): Reconciling cognitive and affective neuroscience perspectives on the brain basis of emotional experience. – *Neuroscience and biobehavioral reviews*, 76, part B: 187–215 <https://doi.org/10.1016/j.neubiorev.2016.09.010>.
- Parr, T. & Friston, K. (2017): Working memory, attention, and salience in active inference. – *Scientific Reports*, 7 (1): 14678 <https://doi.org/10.1038/s41598-017-15249-0>.
- Parr, T. & Friston, K. (2018): The Discrete and Continuous Brain: From Decisions to Movement—and Back Again. – *Neural Computation*: 1–29 https://doi.org/10.1162/neco_a_01102.
- Passingham, R. & Smaers, J. (2014): Is the prefrontal cortex especially enlarged in the human brain allometric relations and remapping factors. – *Brain, behavior and evolution*, 84 (2): 156–66 <https://doi.org/10.1159/000365183>.
- Patel, G., Yang, D., Jamerson, E., Snyder, L., Corbetta, M. & Ferrera, V. (2015): Functional evolution of new and expanded attention networks in humans. – *Proceedings of the National Academy of Sciences*, 112 (30): 9454–9459 <https://doi.org/10.1073/pnas.1420395112>.
- Peelen, M., Atkinson, A. & Vuilleumier, P. (2010): Supramodal representations of perceived emotions in the human brain. – *The Journal of neuroscience*, 51 (30): 10127–34 <https://doi.org/10.1523/JNEUROSCI.2161-10.2010>.

- Pezzulo, G. (2012): An Active Inference view of cognitive control. – *Frontiers in Psychology*, 3: 478 <https://doi.org/10.3389/fpsyg.2012.00478>.
- Pezzulo, G., Rigoli, F. & Friston, K. (2015): Active Inference, homeostatic regulation and adaptive behavioural control. – *Progress in neurobiology*, 134: 17–35 <https://doi.org/10.1016/j.pneurobio.2015.09.001>.
- Power, J., Cohen, A., Nelson, S., Wig, G., Barnes, K., Church, J., Vogel, A., Laumann, T., Miezin, F., Schlaggar, B. & Petersen, S. (2011): Functional Network Organization of the Human Brain. – *Neuron*, 72 (4): 665–678 <https://doi.org/10.1016/j.neuron.2011.09.006>.
- Radley, J.J., Kabbaj, M., Jacobson, L., Heydendael, W., Yehuda, R. & Herman, J.P. (2011): Stress risk factors and stress-related pathology: Neuroplasticity, epigenetics and endophenotypes. – *Stress*, 14 (5): 481–497 <https://doi.org/10.3109/10253890.2011.604751>.
- Reyes, L.D. & Sherwood, C.C. (2015): Neuroscience and Human Brain Evolution. – In: – pp. 11–37 (Springer, Cham) https://doi.org/10.1007/978-3-319-08500-5_2.
- Rilling, J.K. (2006): Human and nonhuman primate brains: Are they allometrically scaled versions of the same design? – *Evolutionary Anthropology: Issues, News, and Reviews*, 15 (2): 65–77 <https://doi.org/10.1002/evan.20095>.
- Russell, J. (1991): Culture and the categorization of emotions. – *Psychological Bulletin*, 110 (3): 426–450 <https://doi.org/10.1037/0033-2909.110.3.426>.
- Salin, P.. & Bullier, J. (1995): Corticocortical connections in the visual system: structure and function. – *Physiological Reviews*, 75 (1): 107–154

<https://doi.org/10.1152/physrev.1995.75.1.107>.

Sallet, J., Mars, R., Noonan, M., Andersson, J., O'Reilly, J., Jbabdi, S., Croxson, P., Jenkinson, M., Miller, K. & Rushworth, M. (2011): Social Network Size Affects Neural Circuits in Macaques. – *Science*, 334 (6056): 697–700
<https://doi.org/10.1126/science.1210027>.

Saxe, R. & Houlihan, S. (2017): Formalizing emotion concepts within a Bayesian model of theory of mind. – *Current Opinion in Psychology*, 17: 15–21
<https://doi.org/10.1016/J.COPSYC.2017.04.019>.

Schwartenbeck, P., FitzGerald, T., Mathys, C., Dolan, R. & Friston, K. (2015): The Dopaminergic Midbrain Encodes the Expected Certainty about Desired Outcomes. – *Cerebral Cortex*, 25 (10): 3434–3445
<https://doi.org/10.1093/cercor/bhu159>.

Seeley, W., Menon, V., Schatzberg, A., Keller, J., Glover, G., Kenna, H., Reiss, A. & Greicius, M. (2007): Dissociable intrinsic connectivity networks for salience processing and executive control. – *The Journal of neuroscience*, 27 (9): 2349–56 <https://doi.org/10.1523/JNEUROSCI.5587-06.2007>.

Semendeferi, K., Teffer, K., Buxhoeveden, D.P., Park, M.S., Bludau, S., Amunts, K., Travis, K. & Buckwalter, J. (2011): Spatial Organization of Neurons in the Frontal Pole Sets Humans Apart from Great Apes. – *Cerebral Cortex*, 21 (7): 1485–1497 <https://doi.org/10.1093/cercor/bhq191>.

Seth, A. & Friston, K. (2016): Active interoceptive inference and the emotional brain. – *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 371 (1708).

- Shaver, P., Schwartz, J., Kirson, D. & O'Connor, C. (1987): Emotion knowledge: further exploration of a prototype approach. – *Journal of personality and social psychology*, 52 (6): 1061–86.
- Singer, T., Seymour, B., O'Doherty, J.P., Stephan, K.E., Dolan, R.J. & Frith, C.D. (2006): Empathic neural responses are modulated by the perceived fairness of others. – *Nature*, 439 (7075): 466–9 <https://doi.org/10.1038/nature04271>.
- Skerry, A. & Saxe, R. (2015): Neural Representations of Emotion Are Organized around Abstract Event Features. – *Current Biology*, 25 (15): 1945–1954 <https://doi.org/10.1016/j.cub.2015.06.009>.
- Smart, C., Strathdee, G., Watson, S., Murgatroyd, C. & McAllister-Williams, R.H. (2015): Early life trauma, depression and the glucocorticoid receptor gene – an epigenetic perspective. – *Psychological Medicine*, 45 (16): 3393–3410 <https://doi.org/10.1017/S0033291715001555>.
- Smith, R. (2017): A neuro-cognitive defense of the unified self. – *Consciousness and Cognition*, 48: 21–39 <https://doi.org/10.1016/j.concog.2016.10.007>.
- Smith, R., Alkozei, A., Bao, J., Smith, C., Lane, R. & Killgore, W. (2017): Resting state functional connectivity correlates of emotional awareness. – *NeuroImage*, 159: 99–106 <https://doi.org/10.1016/j.neuroimage.2017.07.044>.
- Smith, R., Bajaj, S., Dailey, N., Alkozei, A., Smith, C., Sanova, A., Lane, R. & Killgore, W. (2018): Greater cortical thickness within the limbic visceromotor network predicts higher levels of trait emotional awareness. – *Consciousness and Cognition*, 57: 54–61.

- Smith, R., Fass, H. & Lane, R. (2014): Role of medial prefrontal cortex in representing one's own subjective emotional responses: A preliminary study. – *Consciousness and Cognition*, 29: 117–130
<https://doi.org/10.1016/j.concog.2014.08.002>.
- Smith, R., Killgore, W. & Lane, R. (2018): The structure of emotional experience and its relation to trait emotional awareness: A theoretical review. – *Emotion*, 18 (5): 670–692 <https://doi.org/10.1037/emo0000376>.
- Smith, R. & Lane, R. (2015): The neural basis of one's own conscious and unconscious emotional states. – *Neuroscience and biobehavioral reviews*, 57: 1–29 <https://doi.org/10.1016/j.neubiorev.2015.08.003>.
- Smith, R., Lane, R., Alkozei, A., Bao, J., Smith, C., Sanova, A., Nettles, M. & Killgore, W. (2017): Maintaining the feelings of others in working memory is associated with activation of the left anterior insula and left frontal-parietal control network. – *Social Cognitive and Affective Neuroscience*, 12 (5): 848–860 <https://doi.org/10.1093/scan/nsx011>.
- Smith, R., Lane, R., Alkozei, A., Bao, J., Smith, C., Sanova, A., Nettles, M. & Killgore, W. (2018): The role of medial prefrontal cortex in the working memory maintenance of one's own emotional responses. – *Scientific reports*, 8 <https://doi.org/10.1038/s41598-018-21896-8>.
- Smith, R., Lane, R., Parr, T. & Friston, K. (2019): Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance. – *Neuroscience and biobehavioral reviews* <https://doi.org/10.1101/681288>.

- Smith, R., Lane, R., Sanova, A., Alkozei, A., Smith, C. & Killgore, W.W.D. (2018): Common and Unique Neural Systems Underlying the Working Memory Maintenance of Emotional vs. Bodily Reactions to Affective Stimuli: The Moderating Role of Trait Emotional Awareness. – *Frontiers in Human Neuroscience*, 12: 370 <https://doi.org/10.3389/fnhum.2018.00370>.
- Smith, R., Parr, T. & Friston, K.J. (2019): Simulating emotions: An active inference model of emotional state inference and emotion concept learning. – *bioRxiv*: 640813 <https://doi.org/10.1101/640813>.
- Smith, R., Quinlan, D., Schwartz, G.E., Sanova, A., Alkozei, A. & Lane, R.D. (2019): Developmental Contributions to Emotional Awareness. – *Journal of Personality Assessment*, 101 (2): 150–158 <https://doi.org/10.1080/00223891.2017.1411917>.
- Smith, R., Thayer, J., Khalsa, S. & Lane, R. (2017): The hierarchical basis of neurovisceral integration. – *Neuroscience and biobehavioral reviews*, 75: 274–296 <https://doi.org/10.1016/j.neubiorev.2017.02.003>.
- Smith, S., Fox, P.T., Miller, K., Glahn, D., Fox, P.M., Mackay, C., Filippini, N., Watkins, K., Toro, R., Laird, A. & Beckmann, C. (2009): Correspondence of the brain's functional architecture during activation and rest. – *Proceedings of the National Academy of Sciences of the United States of America*, 106 (31): 13040–5 <https://doi.org/10.1073/pnas.0905267106>.
- Soutschek, A., Burke, C.J., Raja Beharelle, A., Schreiber, R., Weber, S.C., Karipidis, I.I., ten Velden, J., Weber, B., Haker, H., Kalenscher, T. & Tobler, P.N. (2017): The dopaminergic reward system underpins gender differences

in social preferences. – *Nature Human Behaviour*, 1 (11): 819–827

<https://doi.org/10.1038/s41562-017-0226-y>.

Spreng, R.N., Stevens, W.D., Chamberlain, J.P., Gilmore, A.W. & Schacter, D.L.

(2010): Default network activity, coupled with the frontoparietal control

network, supports goal-directed cognition. – *NeuroImage*, 53 (1): 303–17

<https://doi.org/10.1016/j.neuroimage.2010.06.016>.

Steklis, H.D. & Lane, R.D. (2013): The Unique Human Capacity for Emotional

Awareness: Psychological, Neuroanatomical, Comparative and Evolutionary

Perspectives. – In: – pp. 165–205 (Springer, Tokyo)

https://doi.org/10.1007/978-4-431-54123-3_8.

Stephan, K., Manjaly, Z., Mathys, C., Weber, L., Paliwal, S., Gard, T.,

Tittgemeyer, M., Fleming, S., Haker, H., Seth, A. & Petzschner, F. (2016):

Allostatic Self-efficacy: A Metacognitive Theory of Dyshomeostasis-Induced

Fatigue and Depression. – *Frontiers in human neuroscience*, 10: 550

<https://doi.org/10.3389/fnhum.2016.00550>.

Suddendorf, T. (2018): Two key features created the human mind: Inside our

heads. – *Scientific American*, 139 (3): 42–47.

Tavares, P., Barnard, P. & Lawrence, A. (2011): Emotional complexity and the

neural representation of emotion in motion. – *Social cognitive and affective*

neuroscience, 6 (1): 98–108 <https://doi.org/10.1093/scan/nsq021>.

Tomasello, M. (2014): The ultra-social animal. – *European Journal of Social*

Psychology, 44 (3): 187–194 <https://doi.org/10.1002/ejsp.2015>.

Tomasello, M. & Herrmann, E. (2010): Ape and Human Cognition. – *Current*

Directions in Psychological Science, 19 (1): 3–8

<https://doi.org/10.1177/0963721409359300>.

Tomasello, M., Melis, A.P., Tennie, C., Wyman, E. & Herrmann, E. (2012): Two Key Steps in the Evolution of Human Cooperation. – *Current Anthropology*, 53 (6): 673–692 <https://doi.org/10.1086/668207>.

Trut, L., Oskina, I. & Kharlamova, A. (2009): Animal evolution during domestication: the domesticated fox as a model. – *BioEssays*, 31 (3): 349–60 <https://doi.org/10.1002/bies.200800070>.

Ueda, S., Kumagai, G., Otaki, Y., Yamaguchi, S. & Kohshima, S. (2014): A Comparison of Facial Color Pattern and Gazing Behavior in Canid Species Suggests Gaze Communication in Gray Wolves (*Canis lupus*). – *PLoS ONE*, 9 (6): e98217 <https://doi.org/10.1371/journal.pone.0098217>.

Veissière, S.P.L., Constant, A., Ramstead, M.J.D., Friston, K.J. & Kirmayer, L.J. (2019): Thinking Through Other Minds: A Variational Approach to Cognition and Culture. – *Behavioral and Brain Sciences*: 1–97 <https://doi.org/10.1017/S0140525X19001213>.

Vincent, J., Kahn, I., Snyder, A., Raichle, M. & Buckner, R. (2008): Evidence for a Frontoparietal Control System Revealed by Intrinsic Functional Connectivity. – *Journal of Neurophysiology*, 100 (6): 3328–3342 <https://doi.org/10.1152/jn.90355.2008>.

Vincent, J., Patel, G., Fox, M., Snyder, A., Baker, J., Van Essen, D., Zempel, J., Snyder, L., Corbetta, M. & Raichle, M. (2007): Intrinsic functional architecture in the anaesthetized monkey brain. – *Nature*, 447 (7140): 83–86

<https://doi.org/10.1038/nature05758>.

vonHoldt, B.M., Shuldiner, E., Koch, I.J., Kartzinel, R.Y., Hogan, A., Brubaker, L., Wanser, S., Stahler, D., Wynne, C.D.L., Ostrander, E.A., Sinsheimer, J.S. & Udell, M.A.R. (2017): Structural variants in genes associated with human Williams-Beuren syndrome underlie stereotypical hypersociability in domestic dogs. – *Science Advances*, 3 (7): e1700398
<https://doi.org/10.1126/sciadv.1700398>.

Wager, T., Kang, J., Johnson, T., Nichols, T., Satpute, A. & Barrett, L. (2015): A Bayesian Model of Category-Specific Emotional Brain Responses. – *PLOS Computational Biology*, 11 (4): e1004066
<https://doi.org/10.1371/journal.pcbi.1004066>.

Weaver, I., Cervoni, N., Champagne, F., D'Alessio, A., Sharma, S., Seckl, J., Dymov, S., Szyf, M. & Meaney, M. (2004): Epigenetic programming by maternal behavior. – *Nature Neuroscience*, 7 (8): 847–854
<https://doi.org/10.1038/nn1276>.

Whalen, P., Kagan, J., Cook, R., Davis, F., Kim, H., Polis, S., McLaren, D., Somerville, L., McLean, A., Maxwell, J. & Johnstone, T. (2004): Human amygdala responsivity to masked fearful eye whites. – *Science*, 306 (5704): 2061 <https://doi.org/10.1126/science.1103617>.

Widen, S. & Russell, J. (2008): Children acquire emotion categories gradually. – *Cognitive Development*, 23 (2): 291–312
<https://doi.org/10.1016/j.cogdev.2008.01.002>.

Wittmann, M.K., Lockwood, P.L. & Rushworth, M.F.S. (2018): Neural

Mechanisms of Social Cognition in Primates. – Annual Review of Neuroscience, 41 (1): 99–118 <https://doi.org/10.1146/annurev-neuro-080317-061450>.

Wright, R., Riedel, R., Sechrest, L., Lane, R. & Smith, R. (2017): Sex differences in emotion recognition ability: The mediating role of trait emotional awareness. – Motivation and Emotion <https://doi.org/10.1007/s11031-017-9648-0>.

Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkari, D., Hollinshead, M., Roffman, J., Smoller, J., Zöllei, L., Polimeni, J., Fischl, B., Liu, H. & Buckner, R. (2011): The organization of the human cerebral cortex estimated by intrinsic functional connectivity. – Journal of neurophysiology, 106 (3): 1125–65 <https://doi.org/10.1152/jn.00338.2011>.

Zhou, Y., Friston, K., Zeidman, P., Chen, J., Li, S. & Razi, A. (2018): The Hierarchical Organization of the Default, Dorsal Attention and Salience Networks in Adolescents and Young Adults. – Cerebral Cortex, 28 (2): 726–737 <https://doi.org/10.1093/cercor/bhx307>.