Biological Psychology 154 (2020) 107925

Contents lists available at ScienceDirect

Biological Psychology

journal homepage: www.elsevier.com/locate/biopsycho

Review

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

Ryan Smith[a,*], Horst Dieter Steklis[b], Netzin G. Steklis[b], Karen L. Weihs[b], Richard D. Lane[b]

[a] *Laureate Institute for Brain Research, 6655 South Yale Avenue, Tulsa, OK 74136, United States*
[b] *University of Arizona, 1501 N. Campbell Ave, Tucson, AZ 85724-5002, United States*

ARTICLE INFO

Keywords:
Emotion
Emotional awareness
Computational neuroscience
Active inference
Life history strategy
Human evolution
Hypersociality
Sex differences

ABSTRACT

We offer an interdisciplinary framework for understanding the expanded capacity for emotional awareness (EA) in humans relative to other animals, synthesizing work within computational neuroscience, evolutionary psychology, and comparative anatomy. We argue that disproportionate cortical expansion during human evolution reflects additional hierarchical levels of computational processing, allowing representation of multimodal regularities over longer timescales – affording abstract concept learning, internal simulation of distal future outcomes, and expanded working memory capacity. This allows for the ability to simulate emotions, learn emotion concepts, and manipulate them in working memory when deciding how to act. We also draw on the construct of life history strategy within evolutionary psychology to argue that individual differences in EA within humans can be understood as the result of tuning particular computational parameters to the predictability of long timescale socioemotional regularities of the local environment. We conclude by discussing the implications and testable hypotheses offered by our proposed framework.

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, Lane, Solms, & Smith, 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 (LeDoux, Brown, Pine, & Hofmann, 2018) (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, 2012).

1.1. Emotional awareness

EA, as operationalized in this paper, is primarily measured linguistically. It is often evaluated on a continuum based on the granularity/specificity of the words people spontaneously use to describe their affective experiences (Kashdan, Barrett, & McKnight, 2015; Smith, Killgore, & Lane, 2018). For example, based on the Levels of Emotional Awareness Scale (LEAS; (Lane, Quinlan, Schwartz, Walker, & Zeitlin, 1990)), this continuum, from low to high, is assessed based on 1) the use of low-level somato- or viscero-sensory-motor terms (e.g., feeling sick or achy), 2) the use of action tendency or valence terms (e.g., feeling bad or feeling like punching someone), 3) the use of specific emotion concept terms (e.g., sad, angry), and 4) describing combinations of feelings (e.g., feeling both sad and afraid). The ability to distinguish one’s own emotions from those of others (i.e., a type of affective theory of mind) is also considered a hallmark of high EA. Although primarily measured in this fashion, our conception of EA, while plausibly (strongly) facilitated by language, is not seen as itself a linguistic capacity. Rather, use of specific emotion terms is considered an indirect measure that tracks the way an individual’s affective experiences (i.e., context-dependent changes in valence, arousal, action

---

[*] Corresponding author.
E-mail address: rsmith@laureateinstitute.org (R. Smith).

https://doi.org/10.1016/j.biopsycho.2020.107925
Received 24 April 2020; Received in revised form 17 June 2020; Accepted 23 June 2020

tendencies, and associated cognitions that are currently internally represented, consciously or unconsciously) are conceptually categorized and the way this alters the temporal dynamics of affective experience and its use in goal-directed cognition.

Based on a previous "three-process" model we have proposed ((Smith, Killgore, & Lane, 2018; Smith, Lane, Parr, & Friston, 2019), EA requires that affective responses are 1) generated (which may not occur in some cases of low EA; ((Smith, Kaszniak, Katsanis, Lane, & Nielsen, 2019)), 2) represented at both a sensory-motor and cognitive (concept category) level, and 3) maintained/manipulated within domain-general cognition (i.e., working memory) so as to inform verbal reporting and other goal-directed planning processes. It follows that, if an animal can only represent sensory-motor aspects of affective experience (e.g., valence, arousal, action motivations), or if it lacks the capacity to use this information to inform goal-directed planning, that animal will have low EA.

A few other points are also worth clarifying at the outset. First, we assume that affective responses are flexible and context-specific – rather than being stereotyped categorical responses with a 1-to-1 mapping with emotion concept categories – and that they reflect the metabolic, cognitive, and behavioral demands that are (most often unconsciously) predicted to be necessary for dealing with a given situation (Feldman Barrett & Finlay, 2018). As such, emotion concept learning involves learning one (e.g., culture-specific; (Russell, 1991)) way of categorizing in-context affective responses, out of many other possible conceptual systems – each of which could guide cognition and behavior in different ways (Barrett, 2017). Second, we assume that affective responses are generated and dynamically modulated based on a range of factors, including (both conscious and unconscious) interpretations of a represented situation (whether real, remembered, or imagined), interpretations of one's own internal state, and faster conditioned and unconditioned responses to salient stimuli (Smith & Lane, 2015). Finally, similar to other theories within the broader literature on conscious experience (e.g., (Prinz, 2007, 2012)), we assume that the phenomenological character of an affective response is largely at an intermediate level of representation, such as the unpleasant feeling of a racing heart, a felt motivation to avoid, etc., and that higher-level cognitive representations play an indirect role in modulating emotional phenomenology in a top-down manner (e.g., based on what one expects to feel or attend to when sad). However, both perceptual and conceptual representations can remain unconscious (only producing priming effects; (Kihlstrom, Mulvaney, Tobias, & Tobis, 2000; Smith & Lane, 2016; Winkielman & Berridge, 2004; Winkielman, Berridge, & Wilbarger, 2005; Zemack-Rugar, Bettman, & Fitzsimons, 2007)), and a further process is necessary for awareness in which locally represented information is "broadcast" and made available for use within domain-general cognition (Dehaene, 2014; Dehaene, Charles, King, & Marti, 2014).

1.2. The origin of individual differences

While the difference in EA between humans and other animals is uncontroversial, its phylogenetic origins (i.e., the selective pressures that favored it during the course of human evolution) are poorly understood. Further, there are well-established individual differences in EA within the human species (Smith, Killgore, & Lane, 2018), and the developmental 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 (e.g., biased attention toward or enhanced salience of cues relevant to) the emotional states of self and others. To be clear, by "domain-general", we refer to cognitive systems whose operations can be applied across many types and sources of information (e.g., the ability for both exteroceptive and interoceptive information to inform goal-directed planning processes). By "domain-specific", we instead refer to information processing and behavioral control mechanisms that are dedicated to only processing specific types or sources of information (e.g., selective processing of visual information about faces).

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 (see (Suddendorf & Corballis, 2007) for a comprehensive review; and (Suddendorf, Bulley, & Miloyan, 2018) for evolutionary significance). 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 is 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 co-worker 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, Mesquita, & Gendron, 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, Schneebeli, & Stephan, 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 (Heyes & Frith, 2014).

This qualitative difference in general reflective capacity in humans relative to other animals (Suddendorf & Corballis, 2007) is also arguably the primary causal factor in facilitating many 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 (Smith & Lane, 2015; Smith, Killgore, & Lane, 2018). Similarly, knowing what someone else is feeling enables predictions about what other people need and are likely to do (Smith, Killgore, Alkozei, & Lane, 2018). 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 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 humans relative to other animals is plausibly best understood as the result of the selective pressures associated with the hypersocial ecological niche that humans came to occupy (Apicella & Silk, 2019). 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 hypersocial niche, we argue that fitness was plausibly enhanced by an expanded, domain-general capacity to learn about the highly abstract, context-varying, long-timescale regularities required to successfully navigate social life.

We next argue that, 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 (i.e., an internal model that *generates* predictions about what sensory input should be observed if a given set of beliefs were correct (Friston, Rosch, Parr, Price, & Bowman, 2018; Kiebel, Daunizeau, & Friston, 2008)). 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 (Parr & Friston, 2017a; Smith, Lane et al., 2019; Smith, Parr, & Friston, 2019). 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 (Mesulam, 1998; Yeo et al., 2011; Zhou et al., 2018) – the very cortical regions showing disproportionate expansion in humans (Buckner & Krienen, 2013).

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, & Lane, 2018). Specifically, we will argue that individual differences in EA, and reflective capacity generally, can be explained, at least to a large degree, by bounded rational statistical (probabilistic) 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 harsh or unpredictable environments (Ellis, Figueredo, Brumbach, & Schlomer, 2009), an individual may learn that long-timescale regularities are not reliable bases for prediction and action selection. Under this implicit belief, many "unreflective" decisions involving a lack of thorough consideration of distant future outcomes (in both general and specifically social decision-making) can be explained (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?). This perspective entails that there is not an "optimal" level of EA. Instead, EA is tuned to the statistics of the social environment based on experience. An individual's level of EA can then only be considered "suboptimal" when they enter a new social environment with different statistical regularities, and the level of EA no longer matches those regularities.

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 hypersocial 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 emotion learning/recognition in autistic individuals with poor EA (Moriuchi, Klin, & Jones, 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.

After we have presented our proposed framework, we critically evaluate it and consider other possible explanations that could be offered, the relative empirical support for each, and how such competing views might be tested. 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 internal reflection generally, facilitating the development of EA.

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 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 domain-general. For instance, 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. We also acknowledge that in some cases this distinction may be overly coarse-grained, in that some functions may lie somewhere along a continuum between specificity and generality – by integrating inputs from an intermediate number of distal information sources (e.g., hierarchically intermediate neural circuits that integrate information from multiple exteroceptive and interoceptive channels, but do not incorporate contextual information from long-term memory; see (Smith, Thayer, Khalsa, & Lane, 2017)).

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 (Carter, 2014). 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; rather, these neural structures were co-opted to also serve additional functions (reviewed in (Chang et al., 2013; Wittmann, Lockwood, & Rushworth, 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, humans possess 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, & Lane, 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, 2012)). 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, Wright, Dautoff, Dickerson, & Barrett, 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; LeDoux et al., 2018). 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 & 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 (Barrett & Satpute, 2013; Seeley et al., 2007; Zhou et al., 2018)). 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 has also 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., 2.5–4 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, Passingham, Neubert, Verhagen, & Sallet, 2017)).

The human prefrontal cortex is of particular interest in the present context because it appears to have undergone evolutionary modifications that may be linked to aspects of social cognition (Amodio & Frith, 2006). Compared to all 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). Compared to chimpanzees, the human frontopolar cortex contains more neuropil, indicative of enhanced connectivity (Spocter et al., 2012). 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, Rezaie, Brown, Roberts, & Dunbar, 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 many other primates the eyes are a special focus of attention (Kingstone, Laidlaw, Nasiopoulos, & Risko, 2017), and eye gaze potentially communicates social and nonsocial 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 them (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 (Hare, 2017; Mayhew & Gómez, 2015)). 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, Kumagai, Otaki, Yamaguchi, & Kohshima, 2014).

2.2. The evolution of hypersociality and EA

We now want to consider more explicitly the socio-environmental selective pressures that, in the course of human evolution, could have 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) (Herrmann, Call, Hernández-Lloreda, Hare, & Tomasello, 2007; Povinelli, 2000; Tomasello & Call, 1997; 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 & Carpenter, 2007). In effect, shared intentionality amounts to a "drive to exchange thoughts with others" (Suddendorf et al., 2018) and as such brings into play theory of mind capacities (Heyes & Frith, 2014). 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. There is much empirical evidence comparing the cooperative and prosocial behaviors among nonhuman primates that warrants this description of humans as a distinctly hypersocial primate species (see overview in (Apicella & Silk, 2019)). Distinctive human hyper-cooperation, therefore, relies on a suite of cognitive-emotional capacities, particularly mind-reading (or theory of mind), including empathic understanding, imitation, symbolic communication, and for mentalizing generally—to imagine scenarios of present and future actions and multiple possible outcomes (Suddendorf, 2018). Studies comparing the imitative capacities of apes (including ones raised in proximity to humans) and young children (Subiaul, 2016) clearly show that imitation in humans, but not apes, involves faithfully copying "item-specific and motor-spatial specific rules". This type of 'high-fidelity' copying, or 'overimitation', shown by young children in different cultures, 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 package of distinctively human socio-cognitive traits is thus 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, 2017; Laland & Rendell, 2017). This package of uniquely human mental and behavioral skills required for hyper-cooperation apparently also continued to drive 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)). Within-group social competition has also been put forward as a possible driving force for brain expansion (Flinn, Geary, & Ward, 2005; Gavrilets & Vose, 2006), as selective pressures may have plausibly favored more and more intelligent strategies for achieving social success, including "Machiavellian" strategies – such as deception, manipulation, exploitation, and alliance formation – that would draw heavily on theory of mind and other social cognitive functions, including EA.

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 (i.e., shorthand for cooperative hunting-gathering) strategy (Tomasello, Melis, Tennie, Wyman, & Herrmann, 2012) or a move to cooperative breeding (i.e., allomaternal care – 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 of both sexes cooperating in aiding, provisioning, and protection of offspring. The evolution of cooperative breeding as a cause of hypersociality is supported 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 suggest that the adoption of cooperative breeding in our hominin ancestors made possible brain expansion and the evolution of the cognitive skills underlying hypersociality (Burkart, Hrdy, & Van Schaik, 2009), 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 hypercooperative groups, with cooperative breeding as the initial stimulus followed 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 nonkin whose fitness was interdependent (Apicella & Silk, 2019; Brown & Cialdini, 2015). Proactive prosociality plausibly 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). Further, 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), including communal warfare among hunter-gatherers traditionally thought to be uniformly peaceful (Gat, 2015), 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 & 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 necessarily somewhat speculative, it is consistent with several additional observations. First, the unique form of human cooperation we have described has been documented in several extant hunter-gatherer (forager) societies. The range and extent of human cooperation in these societies is distinct from that observed in other primates in that the cooperative network includes, in addition to relatives, large numbers of unrelated individuals who can be mobilized to coordinate labor or to wage war (Apicella & Silk, 2019). In other words, the size and membership of the cooperating group units can be scaled according to the cooperative task at hand, and cooperation, even without reciprocation, is supported through shared social norms acquired in childhood. Because hunter-gatherer bands are mobile, they are frequently co-resident with large numbers of unrelated individuals that form a part of an efficient social network for cooperation and cultural exchange (Migliano et al., 2016). While present day hunter-gatherers have varying degrees of regular contact with non-hunter-gatherer peoples, they nevertheless represent the best models of human social life before agriculture and hence inform about aspects of social organization that were and continue to be crucial to this way of life (Migliano et al., 2016).

Second, comparison of extant carnivores that hunt cooperatively (e.g, spotted hyena) shows that a suite of traits, including intergroup conflict, large brain size, high reproductive investment, and reduced sexual dimorphism, are key correlates of carnivore cooperation (Smith, Swanson, Reed, & Holekamp, 2012). This provides indirect support for the association of these traits with cooperative hunting (in addition to cooperative breeding) in hominins.

Lastly in support is 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), but see (Sánchez-Villagra & van Schaik, 2019), for a critique). This proposal is also 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, Oskina, & Kharlamova, 2009). Compared to apes, humans also have increased serotonergic innervation in the basolateral nuclei of the 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 may contribute to the higher level of social tolerance required for hyper-cooperation.

We can conclude from our review of both the comparative neuroscience and evolutionary evidence that the evolution of EA likely involved both 1) co-opted domain-general neurobiological mechanisms (e.g., expanded association cortices linked to mentalizing and general reflective capacities, which, as discussed further below, may have allowed for more abstract emotion concept learning, expanded working memory capacity, and longer timescale internal simulation and planning abilities), along with 2) specialized, domain-specific mechanisms (e.g., face recognition and eye gaze detection). In a hypersocial niche, the need for effective social cooperation and social competition abilities plausibly favored those who could efficiently recognize and reflect on the thoughts and feelings of self and others when planning how to act – in other words, it would have favored the capacity for EA. In the next sections, 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 that facilitated navigating 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 (Margulies et al., 2016; Yeo et al., 2011; 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 (Braga & Buckner, 2017; Hermundstad et al., 2013). 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 (Barrett & Satpute, 2013; Doucet et al., 2011; Power et al., 2011; Yeo et al., 2011), 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 (Buckner, Andrews-Hanna, & Schacter, 2008; Binder, Desai, Graves, & Conant, 2009; Corbetta & Shulman, 2002; Felleman & Van Essen, 1991; Margulies et al., 2009; Patel et al., 2015; Vincent et al., 2007; Vincent, Kahn, Snyder, Raichle, & Buckner, 2008).

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, & Lane, 2018). Specifically, the DMN (including medial prefrontal regions discussed above) may learn to represent abstract, multimodal relationships, including those associated with identifying abstract concepts (such as those associated with mental states like emotions, goals, beliefs and values), and may therefore allow 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 and semantic processing more broadly (Binder et al., 2009; Kleckner et al., 2017). The SN and LN (including insula, amygdala, and anterior cingulate regions discussed above) may both play roles in representing valenced bodily percepts during emotions, and the ECN (including lateral prefrontal regions discussed above) 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). Interactions between these networks therefore plausibly underlie the processes on which EA depends – that is, the ability to learn, represent, and internally simulate information about valenced body states and emotion concepts, and to maintain and manipulate this information in working memory.

2.4. Computational neuroscience

The field of computational neuroscience offers the promise of linking neurobiological levels of description with psychological levels of description – by specifying how neuronal connectivity and dynamics can implement algorithms that accomplish perception, learning, and decision-making. This perspective may therefore be ideal in characterizing potential links between brain evolution and increases in the capacity for EA. As such, we consider the insights this perspective may offer in detail below.

In recent years a growing body of work in this field – both empirical (e.g., (Diaconescu et al., 2017; Iglesias et al., 2013; Schwartenbeck, FitzGerald, Mathys, Dolan, & Friston, 2015)) and theoretical (e.g., (Badcock, Davey, Whittle, Allen, & Friston, 2017; Bastos et al., 2012; Clark, 2015; Friston, 2005, 2010; Friston, FitzGerald, Rigoli, Schwartenbeck, & Pezzulo, 2017; Stephan et al., 2016)) – 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 (i.e., a model that generates predictions about future sensory inputs based on current beliefs). A generative model 1) specifies the a priori probability of different possible causes of sensory input ("priors"), 2) provides a probabilistic mapping from those causes to the predicted sensory inputs they would generate (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 (e.g., perceiving bodily states and mapping them to emotion concepts; (Bastos et al., 2012; Parr & Friston, 2018; Smith, Lane et al., 2019; Smith, Parr et al., 2019)), it is proposed that each level of cortical processing contains columns with 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 sensory 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 specific types of prediction signals in these domains can also act as visceromotor and skeletomotor commands (respectively) when these prediction signals are assigned high precision-estimates (Adams, Shipp, & Friston, 2013; Barrett & Simmons, 2015; Pezzulo, Rigoli, & Friston, 2015; Seth & Friston, 2016; Smith, Thayer et al., 2017; Stephan et al., 2016). 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

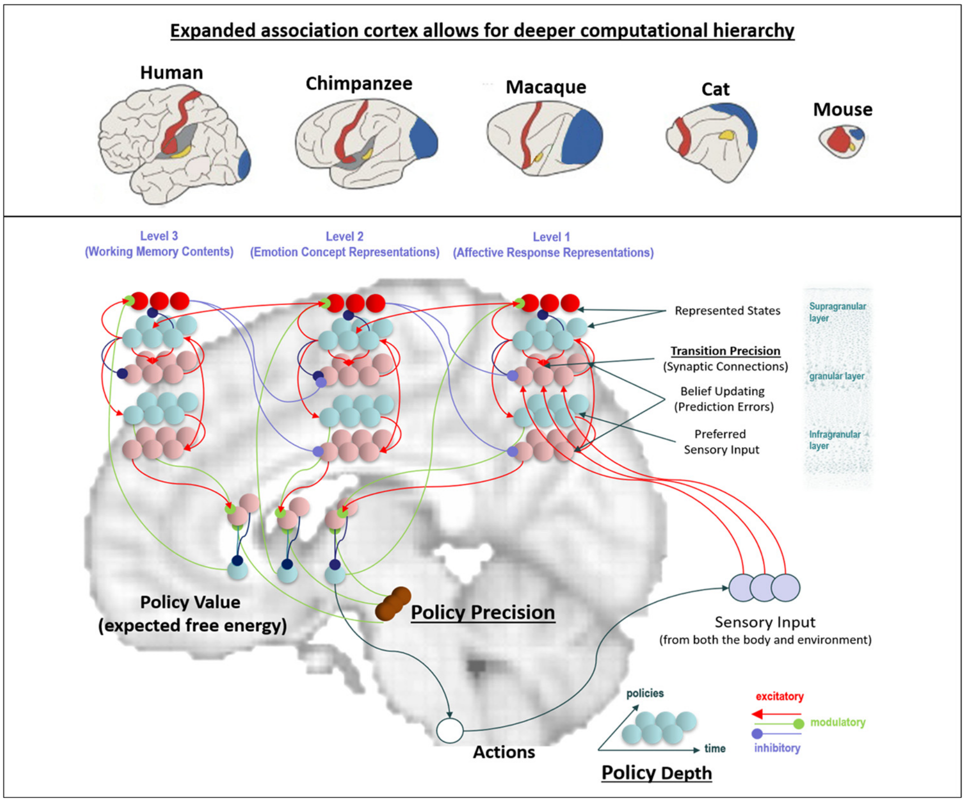

Fig. 1. In the framework we propose in this paper, cortical expansion during human evolution corresponded in part to increases in the number of available hierarchical levels of processing within association cortices (i.e., higher levels from right to left in the bottom portion of the figure). This allowed 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, Lane et al., 2019; Smith, Parr 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 separately 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 a threatening context), whereas the third level can maintain emotion concepts over longer timescales (i.e., working memory) and use them to launch internal simulations and inform goal-directed decision-making. In the proposed implementation, neuronal populations are arranged to reproduce known intrinsic (within cortical area) and extrinsic (between cortical area) connections. Red connections are excitatory, blue connections are inhibitory, and green connections are modulatory (i.e., involving a multiplication or weighting). Cyan units correspond to state representations and (predicted 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, at each hierarchical level we have only 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 (Friston, Lin et al., 2017; Friston, Parr et al., 2017; Da Costa et al., 2020) (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article).

organism attempts to infer the sequence of actions (termed a "policy") with the highest value. Here, value is scored by a mathematical quantity called "expected free energy." Briefly, policies with the lowest expected free energy (highest value) are those predicted to 1) minimize the deviation between expected and preferred outcomes (e.g., maximizing current positive emotion), while 2) also maximizing information gain (e.g., information that would increase confidence in which future actions will lead to preferred outcomes later). 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, heart rate, muscle tension, etc.) must be mapped onto categorical concepts (e.g., food, threat, car, sadness, anger, etc.) (Friston, Parr, & de Vries, 2017).

Two aspects of this perspective are especially relevant to the present thesis. First, one straightforward manner in which evolutionary pressures can sculpt neurally instantiated generative models is by increasing hierarchical depth (see Fig. 1 below), based on the spatiotemporal structure and relevant information domain of the signals an organism must learn in order to thrive within its ecological niche (Bruineberg, Rietveld, Parr, van Maanen, & Friston, 2018; Constant, Ramstead, Veissiere, Campbell, & Friston, 2018; Linson, Clark, Ramamoorthy, & Friston, 2018; Tschantz, Seth, & Buckley, 2020). 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, learning the typical causes, sensations, and behavioral responses over time during emotions, etc.). Importantly, this is precisely what increases in hierarchical depth can provide (i.e., corresponding to an increased number of levels of cortical processing). 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; (Friston, 2005; Salin & Bullier, 1995)), 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.; (Friston et al., 2018; Hasson, Yang, Vallines, Heeger, &

Rubin, 2008; Hasson, Chen, & Honey, 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 (e.g., see Smith, Thayer, Khalsa, & Lane, 2017). 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." One might similarly learn that the emotion concept of happiness (or perhaps contentment) also involves covariance between pleasantness, low arousal, and lying down on a beach. Recent simulation work has also shown that the next level up in a hierarchy within association cortex can in turn act as a type of working memory, allowing information about multiple concepts, such as happiness and excitement, to be combined and held active over longer time scales and inform higher-level decision processes (i.e., policy selection; see (Parr & Friston, 2017 ; Smith, Lane, Parr, & Friston, 2019).

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 (2000, Mesulam, 1998; 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, prepared learning; (Barrett & Broesch, 2012; Dunlap & Stephens, 2014; Peltola, Yrttiaho, & Leppänen, 2018; Theeuwes & Van der Stigchel, 2006)).

The second aspect of the computational perspective that is especially relevant to the present thesis has to do with the need for the calibration (both phylogenetic and ontogenetic) 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, Mathys, & Rees, 2017; Mirza, Adams, Parr, & Friston, 2019; Parr & Friston, 2017b). 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, or when evaluating a potential alliance partner).

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/preferred consequences); in contrast, if it is low, one will act more randomly or out of previously acquired habits that are less context-sensitive (FitzGerald, Dolan, & Friston, 2014; Friston et al., 2016; Pezzulo et al., 2015; Smith, Khalsa, & Paulus, 2019). However, the effects of this parameter are also dependent on the related parameter of "policy depth," which reflects the number of "steps" into the future one considers when 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, this 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). This will include learning about emotional regularities (e.g., the typical causes of different emotions and resulting sensations and behaviors), and holding emotional information in working memory when considering the long-term consequences of different actions. In the following sections we will now synthesize and build on the work reviewed above on comparative, cognitive, and computational neuroscience to 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, which contain highly multimodal content involving temporally 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 apparently coarser-grained, perceptual level (e.g., pleasant/unpleasant, approach/avoid) discrimination abilities of other animals (Steklis & Lane, 2012), 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 (L. Barrett, 2017; Kircanski, Lieberman, & Craske, 2012).

The knowledge built into acquired emotion concepts, which can also be thought of as including schemas/scripts within semantic memory (Shaver, Schwartz, Kirson, & O'Connor, 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 (probabilistically) 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 Fig. 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, Sheridan, & Nelson, 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 likely 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, Constant, Ramstead, Friston, & Kirmayer, 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 structural and functional relationships between social and emotional awareness and the highest-level, most anterior regions of medial and lateral prefrontal cortex (and other connected DMN/ECN regions; (Frith & Frith, 2006, 2012; Saxe & Houlihan, 2017; Skerry & Saxe, 2015; Smith, Lane, Alkozei et al., 2018; Smith, Lane et al., 2017; Smith, Lane, Sanova et al., 2018).

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 further 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 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, Akbudak, Shulman, & Raichle, 2001; Peelen, Atkinson, & Vuilleumier, 2010; Smith, Fass, & Lane, 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 proposed above is very easily extended to learning other mental state 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 (Roff, 2002) and psychology (Figueredo et al., 2005; Kuzawa & Bragg, 2012). As we will argue, the pressures driving the particular LHS in humans may help explain why humans constructed a hypersocial niche, in which EA would have served an adaptive function. 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 (i.e., implicit beliefs in higher vs. lower transition precision in a computational model) and the corresponding need to plan for only short-term or long-term outcomes (i.e., low vs. high policy depth in a computational model). From an organism's point of view, the most salient environmental information is represented by the unavoidable extrinsic mortality faced by members of a population, 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 (the benefits of which 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 (i.e., implying fewer hierarchical levels needed for adaptive computational processing and lower policy depths) – 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 a smaller number of offspring with intense parental investment in each of those offspring. Again, this makes sense given information that this 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 (i.e., the need for a greater number of hierarchical levels available for adaptive computational processing and the consideration of deeper policies) – 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)). Integrating these considerations with the cognitive and computational neuroscience results reviewed above, this therefore suggests that, all else being equal, 'faster' vs. 'slower' LHS should be associated with the capacity for lower vs. higher EA; although – as we address further below – development of EA may further require embedding within a hypersocial niche in which social interaction facilitates learning about emotions in particular.

Among primates, humans have a slow life history strategy 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 and original stimulus for human brain evolution was the challenge posed by increased climatic heterogeneity and its effects on local habitats and food resources. The increased climatic heterogeneity in the African ecosystem of our early ancestors, consisting of successive pulses of colder and drier climate, with a peak around 2.5 million years ago, is associated with the evolution of the genus *Homo* and earliest brain expansion beyond the capacities of African apes (Grove, 2017). Brain expansion, along with a suite of adaptive changes in life history characteristics, cognition and behavior (see Fig. 2) was key in allowing our hominin ancestors to successfully respond to these ecological challenges. Specifically, as shown in Fig. 2, these adaptations included enhanced cognitive capacities for flexible, domain-general problem solving and planning, changes in dietary breadth (i.e., addition of significant amounts of animal protein), expansion of tool technology, and adoption of a cooperative breeding strategy. Cooperative breeding solved the problem of the increased energetic demands of a larger brain in combination with a higher reproductive output (i.e., shorter inter-birth intervals), as described earlier. These adaptive strategies, in turn, selected for further brain expansion, a greater number of levels in a computational processing hierarchy (affording longer timescale prediction and greater policy depth), a slower life history strategy (e.g., delayed development, longer life span) and enhanced cognitive (including greater EA) and behavioral capacities that facilitated construction of a hypersocial niche.

Climatic heterogeneity likely also triggered subsequent migration to northern latitudes (Bailey & Geary, 2009). The paleontological and archeological evidence shows that between 1 million and 500,000 years B.P., *Homo erectus* and subsequent antecessor hominin species expanded geographically throughout much of Eurasia—largely as a consequence of the continued, successful hunting and gathering lifestyle—and had evolved cranial capacities within modern range, along with the slow life history features typical of modern humans (Robson & Wood, 2008; Smith, 2018). The larger brain and increased body size, as evident in *Homo erectus,* necessitated a slower LHS, while also enhancing the socio-cognitive abilities for successfully addressing both ecological and social challenges. Slow life history traits promote cultural learning and transmission, trans-generational economic investment (e.g., 'grandmothering'), 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 led 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) (for a test of this see (Figueredo, Hertler, & Peñaherrera-Aguirre, 2020)), allowing populations, or individuals within populations, to adjust life history speed in accordance with salient local environmental cues received during development that are reliable indicators of environmental unpredictability (e.g., if behavioral responses to a problem are not predictably successful or effective; associated with low prior policy precision in computational models) and harshness (e.g., resource scarcity, pathogen prevalence; (Ellis et al., 2009)) (see Fig. 3). Both environmental unpredictability and harshness, as stressors, concurrently affect rates of morbidity and mortality which, in turn, determines individual LHS (Brumbach, Figueredo, & Ellis, 2009). Slower LHS in turn indexes neural adaptations affording the capacity to learn and represent long-timescale regularities, which, when applied within a hypersocial ecological niche, in turn affords the capacity for higher EA.

While serving as a further explanatory variable for the creation of a hypersocial 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 (ontogenetic) 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 subserving 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, Weihs, Herring, Hishaw, & Smith, 2015; Smith, Killgore, & Lane, 2018). There are several empirical measures of individual differences in EA, including the Levels of Emotional Awareness Scale (LEAS; (Lane et al.,

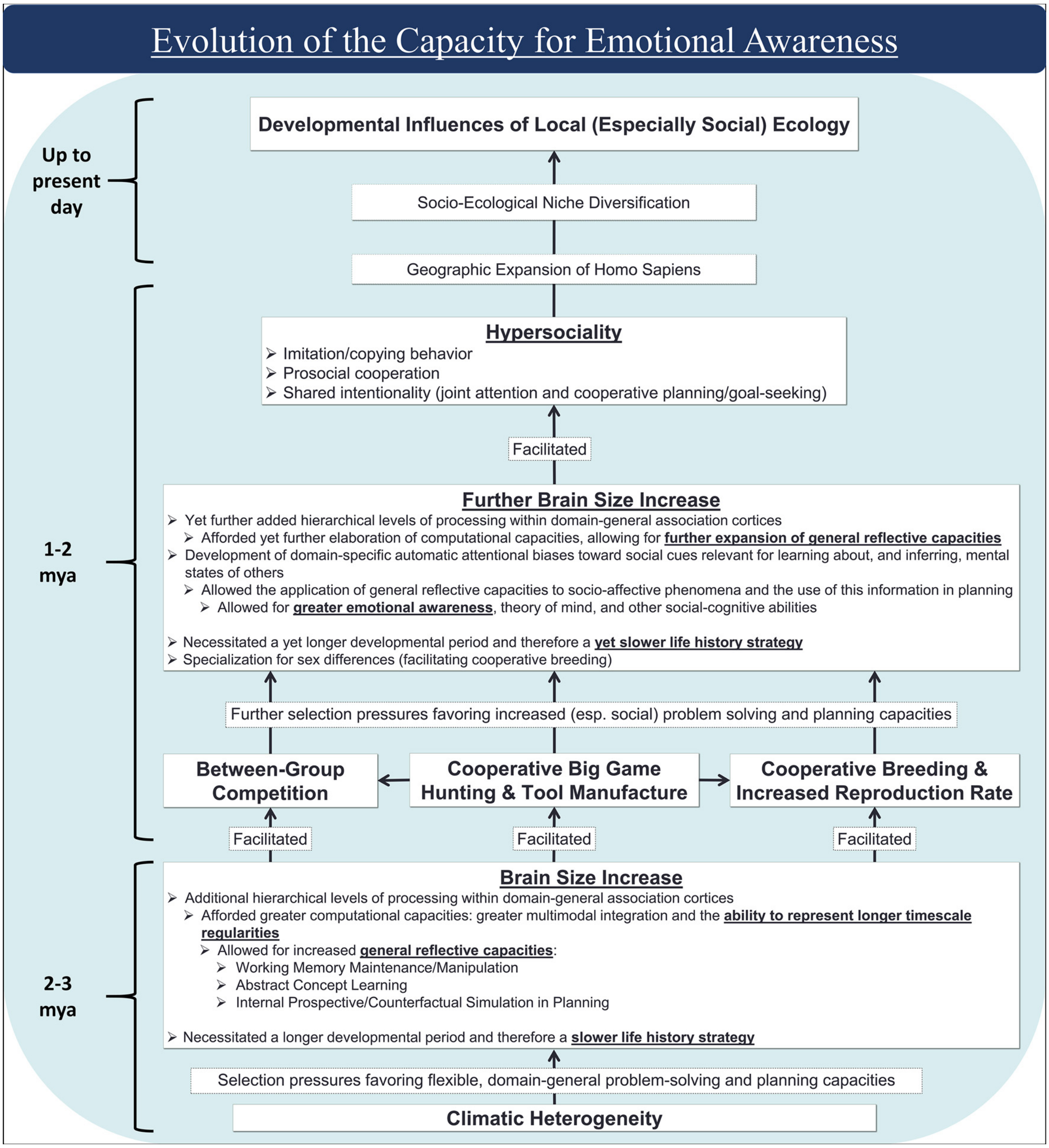

Fig. 2. Flow chart illustrating the sequence of events, and their relationships, which led to the evolution of the capacity for emotional awareness (and the general reflective capacities that allow for it) within our proposal. This figure highlights the result of a comprehensive synthesis of findings from a range of fields, and offers connections between neural, computational, and cognitive levels of description of the processes supporting reflective capacities in humans, and the influences that shaped them. Developmental influences (at the top of the diagram) are illustrated in Fig. 3. Mya = million years ago.

1990)), the Toronto Alexithymia Scale (TAS-20; (Bagby, Parker, & Taylor, 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 (Frewen et al., 2008; Lane, Reiman et al., 1998; McRae, Reiman, Fort, Chen, & Lane, 2008; Smith, Lane, Alkozei et al., 2018; Smith, Lane et al., 2017; Tavares, Barnard, & Lawrence, 2011). 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 (Burger et al., 2016; Farnam, Somi, Farhang, Mahdavi, & Ali Besharat, 2014; Lumley et al., 2017; Neumann, Malec, & Hammond, 2017). Also of 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 (Callaghan & Tottenham, 2016; Gee et al., 2013) 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 emotion regulation 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

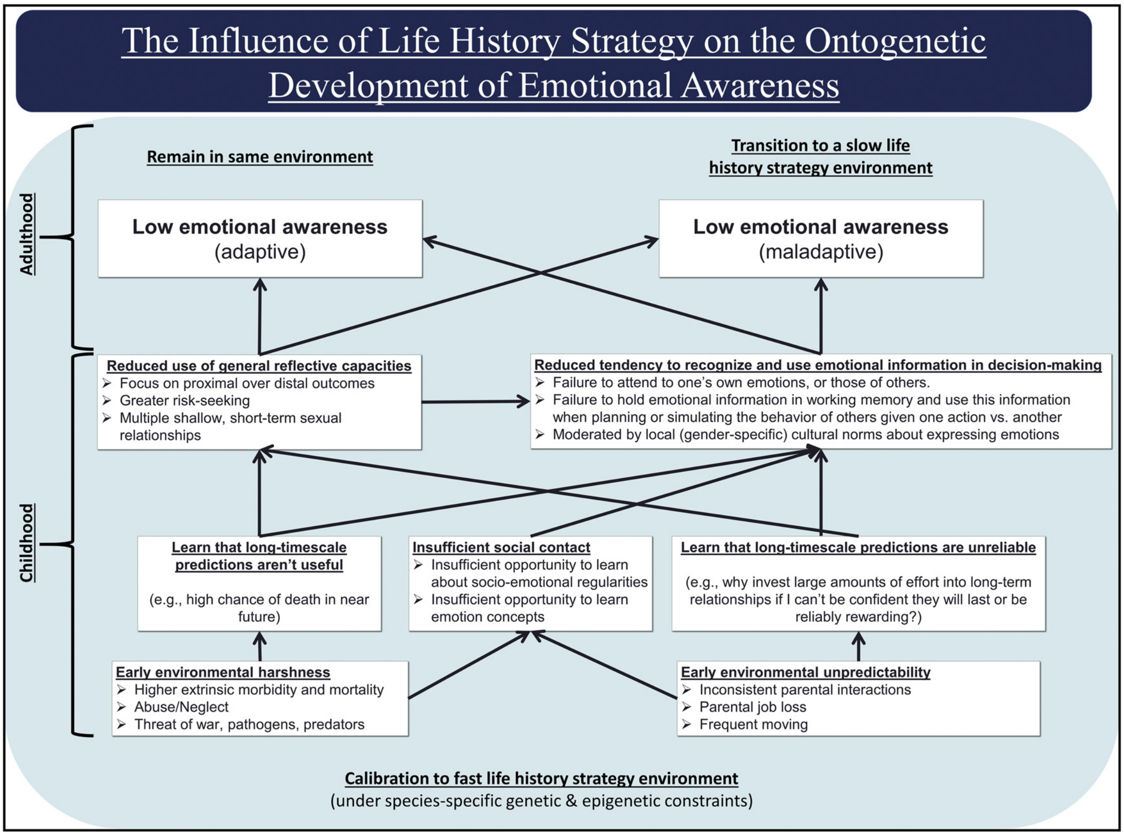

Fig. 3. Flow chart illustrating the sequence of events, and their relationships, which shape the development of emotional awareness (EA) during childhood within our proposal, and showing how EA development is constrained by life history strategy calibration (i.e., using development of low EA as an example). Also highlighted is how the level of EA, and life history strategy, develop so as to match the statistical regularities of the local environment in early childhood. Hence, low EA may be adaptive within fast life history strategy environments and result from the suite of general cognitive adaptions associated with fast life history strategy and their interaction with further social-cognitive development. But it may be maladaptive if, later in life, an individual moves into a slow life history strategy environment. In particular, this diagram illustrates factors stemming from the development of a fast life history strategy, subsequently leading to low EA; the opposite influences (e.g., high predictability, large amounts of social contact, etc.) would instead promote higher EA. Primary influences include 1) the harshness and unpredictability of the environment, where greater levels of either promote reduced reflective capacity generally, and 2) insufficient socio-emotional interactions with caretakers, preventing a child from learning about the emotions of self and others in particular.

would be absent (for explicit computational simulations supporting this, see (Smith, Lane et al., 2019; Smith, Parr et al., 2019; Smith, Schwartenbeck, Parr, & Friston, 2020)). 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 Iow-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., frequent relocation, 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 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 temporally 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, Newport, Mletzko, Miller, & Nemeroff, 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 transition precision is believed to be 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, Watson, & Friston, 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 (Braithwaite, Kundakovic, Ramchandani, Murphy, & Champagne, 2015; Herman, Tasker, Ziegler, & Cullinan, 2002; McEwen, 2000; Radley et al., 2011). Early stress/unpredictability appears to epigenetically alter the expression of glucocorticoid receptors, which are present in the amygdala and hippocampus (Morimoto, Morita, Ozawa, Yokoyama, & Kawata, 1996), and which have been linked to the onset of mood disorders (Nantharat, Wanitchanon, Amesbutr, Tammachote, & Praphanphoj, 2015; Palma-Gudiel, Cordova-Palomera, Marquès Feixa, Cirera Miquel, & Fañanás, 2016; Smart, Strathdee, Watson, Murgatroyd, & McAllister-Williams, 2015). Because high-level prior expectations are often more easily adjusted in the absence of previous learning (i.e., initial expectations may be afforded less confidence before receiving sufficient experiential support), 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 precise high-level prior expectations that the environment is unpredictable – 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 (i.e., the features characteristic of fast LHS and low EA). However, an important overall take-away point of individual variation in life history speed, EA 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 also many untested predictions of this account of within-species differences, 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 (as might be suggested by (Bréjard, Bonnet, & Pedinielli, 2012; Lane, Sechrest, & Riedel, 1998)). 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. Finally, it could also suggest that individuals who have undergone the early adversity associated with fast LHS (e.g., those with borderline personality disorder; (Zanarini & Wedig, 2014)) would benefit from therapeutic interventions aimed at improving EA and reflective emotion regulation techniques (e.g., mentalization-based therapy and dialectical behavior therapy (Feenstra, Luyten, & Bales, 2017; Swales, Heard, & Williams, 2000)), and that this effect would be mediated by improved general reflective capacities (e.g., more granular conceptualization, greater reflection before action, etc.).

4.2. Sex differences, EA, and life history

There is a notable female advantage in both emotion recognition ability and in EA more broadly (e.g., see (Barrett, Lane, Sechrest, & Schwartz, 2000; Ciarrochi, Hynes, & Crittenden, 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, Riedel, Sechrest, Lane, & Smith, 2018). 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, (Chaplin, Cole, & Zahn-Waxler, 2005; Diener & Lucas, 2004). Parents also tend to discuss emotions less with their sons than with their daughters during preschool years (Fivush, Brotman, Buckner, & Goodman, 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., 2018). 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); for a critique of this interpretation, see (Sear, 2020)), 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 be expected to contribute to the selection for reduced variation in their EA skills and theory of mind capacity generally (Hrdy, 1999). In contrast, while some men invest heavily in their children in a context-dependent way (e.g., when paternity confidence is high), 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 sex differences, they nevertheless could help explain the male-female difference in EA. Thus, there may be genetic adaptations promoting these sex differences, but variable constraints on the range of experiences that are typically encountered by men and women may also 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. Critical evaluation

Having proposed our model, we now consider its limitations and highlight other possible perspectives. One difficulty lies in specifying psychological characteristics uniquely evolved in the course of human evolution, especially in identifying such characteristics or traits as adaptations (Leavens, Bard, & Hopkins, 2019; Lewontin, 1998). The concern is that adaptationist explanations, especially in the field of evolutionary psychology, often amount to no more than “just-so stories”, which, while seemingly plausible, can be difficult to falsify empirically (see (Andrews, Gangestad, & Mathew, 2002), for a review and response to this issue). Our synthesis of data from multiple sources in order to provide a coherent proposal for the evolution of EA could likewise be seen as a “story” and is most certainly a hypothetical account, but its elements are nevertheless empirically falsifiable, as are the many novel predictions yielded by our evolutionary account (see especially Section 6, Discussion). For example, the proposals we cite and discuss on the origin and evolution of human hypersocial traits are grounded in many empirical studies comparing cognitive abilities, life history, and neuroanatomical characteristics of humans and other primates, that were guided by hypotheses and thus subject to falsification. With respect to the issue of selection pressures, while we have defended a primary role for adaptation to a hypersocial niche, this is not the only view one could take. For example, it could be that increased reflective capacity was selected for primarily due to the benefits it afforded for temporally deep planning and decision-making, the design and use of complex tools, or a number of other abilities unrelated to sociality in particular. Thus, while it is theoretically possible to construct a different evolutionary scenario from the one we offer, the plausibility and viability of an alternative proposal should depend on its better coherence with empirical evidence and greater heuristic power.

Another limitation worth emphasizing is limited sample size. Some nonhuman neurobiological data come from very few or even single datapoints (e.g., (Semendeferi et al., 2011; Sherwood et al., 2006)), which may limit broadly generalizable conclusions about human/nonhuman differences. This shortcoming can be ameliorated at least in some cases when, despite small sample sizes used in individual studies, there is agreement in results from different laboratories (e.g., the human-ape frontal pole neuropil differences reported by (Spocter et al., 2012) agree with earlier results from (Semendeferi et al., 2011)). As well, to the extent that some of the primate subjects used in the analyses were reared in artificial institutional settings with potentially limited social contact (not consistently described in the publications, but see (Spocter et al., 2012) for exception), with reduced possibilities for social/emotional learning, such conditions could have affected behavior, cognition and associated brain structure (Leavens, Bard, & Hopkins, 2010).

A second sampling issue concerns human data on life history, behavior, and cognition. Admittedly, the large majority of such data have been gathered from humans in what have been called WEIRD cultures (Henrich, Heine, & Norenzayan, 2010) – that is, those that are Western, Educated, Industrialized, Rich, and Democratic and therefore are not necessarily representative of the human species as a whole. However, the judgment of what constitutes an appropriate and sufficient cross-cultural sample to warrant a conclusion about a human universal is not at all clear nor uncontroversial (see commentary accompanying Henrich et al., 2010). For example, it can be argued that most of the increasingly interconnected world is fast becoming westernized and industrialized (i.e., more WEIRD), while at the same time small scale societies – hunter-gatherers and horticulturalists – are either extinct or already adapted to a 21st century industrialized world (Maryanski, 2010). Indeed, many aspects of WEIRD societies may be more expressive of human nature than is true of small-scale societies (Maryanski, 2010). Nevertheless, whenever feasible, data on human differences in brain and cognition, including sex differences (see section 6), should, in a hypothesis-driven way, be informed by additional cultural comparisons to increase our confidence in evolution-based claims.

Yet another issue worth emphasizing is that, while cortical expansion is often assumed to expand cognitive capacities, this assumption can be called into question. For example, children with autism, which is marked by deficits in socio-emotional cognition, show a greater number of neurons in prefrontal cortex and greater brain weight than neurotypical individuals (Courchesne et al., 2011) – suggesting that increased neural resources *per se* (e.g., perhaps without the correct computational architecture) need not expand, and could even diminish, EA.

It’s also worth noting that we have focused primarily on cortical expansion, and not addressed the potential cognitive/computational capacities that may have arisen due to similar evolutionary expansion patterns in the cerebellum (Barton & Venditti, 2014). Some computational theories of cerebellar function suggest that its reciprocal connections with cortical networks may optimize cognitive processes within association cortices (Buckner, 2013) – and so could be seen as contributing to the same expanded capacities discussed above – but this is an important area for future research (for evidence that higher EA is associated with greater functional connectivity between association cortices and the cerebellum, see (Smith, Alkozei et al., 2017); also see (Smith, Sanova, Alkozei, Lane, & Killgore, 2018)).

While the detailed proposal we offer above appears plausible, there are certainly other perspectives worth considering. These cut across (at least) the following two lines: 1) domain-specificity vs. generality, 2) the nature of selection pressures. With respect to the first, recall that the view we have defended assumes that socio-emotional cognition, and EA in particular, represents the application of an expanded general capacity for temporally deep, multimodal cognition – where emotion-focused learning within this cognitive architecture is facilitated by inherited biases toward sampling from high-fidelity sources of socio-emotional information. However, one could instead hold that social/emotion-focused cognition represents a separate domain-specific system dedicated solely to representing the mental states of self and others. While some work has supported the presence of brain regions selectively activated by empathy and mentalization processes (e.g., see (Amodio & Frith, 2006; Atique, Erb, Gharabaghi, Grodd, & Anders, 2011; Frith & Frith, 2006; Kalbe et al., 2010)), we are persuaded by meta-analytic work demonstrating that such regions carry out broader processes that support a much wider range of functions and are therefore better considered domain-general (Anderson, 2014; Kleckner et al., 2017). Interestingly, this domain-general view also underpins one prominent approach to understanding social cognition deficits in autism (Haker et al., 2016). Future research in social neuroscience will be necessary to further assess the validity of our assumptions about domain-generality.

6. Discussion

In this paper we have outlined a theory of how individual differences in EA have emerged both between and within species, involving the interplay between neural, computational, and cognitive-behavioral mechanisms as they interact with, and become adapted to, diverse socio-ecological niches. At the phylogenetic timescale (see Fig. 2), we have suggested that hominin brain expansion – driven by early adaptation to climatic heterogeneity and later adaptation to a hypersocial niche – allowed for an expanded computational capacity to learn about and represent abstract, multimodal, temporally extended regularities, which in turn afforded expanded cognitive capacities for concept learning, internal simulation, and working memory. When applied (and calibrated) to a hypersocial ecological niche, and facilitated by innate attentional biases toward socially informative cues, this promoted the emergence of higher EA (among other social-cognitive capacities) in humans. At the ontogenetic timescale (see Fig. 3), we have applied a computational neuroscience perspective to the construct of LHS in evolutionary psychology to shed light on how levels of early environmental harshness and unpredictability can calibrate neurocomputational processes to develop, or fail to develop, the innate human capacity for EA. Harsh, unpredictable early environments – involving abuse, neglect, insufficient emotional interactions with parents or siblings, among several other factors – would both 1) prevent the opportunity to learn emotion concepts, and 2) promote the expectation that distal future outcomes are bleak and unpredictable. Low EA then emerges from the combination of poor emotion understanding and limited engagement of reflective, future-oriented cognition – the adaptiveness of which later in life will depend on whether an individual remains in an environment that matches that of their early development.

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, Stevens, Chamberlain, Gilmore, & Schacter, 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 EA; see (Smith, Lane, Alkozei et al., 2018; Smith, Lane, Sanova et al., 2018)). Greater ECN activity has also been associated with higher EA when holding other's emotions in mind (Smith, Lane et al., 2017).

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; that is, unpredictability of future observations or the outcomes of actions) with respect to long-timescale predictions. In previous work (Lawson et al., 2017), volatility 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, the aforementioned 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 sex 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, Vásquez, Brumbach, & Schneider, 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, in our account, differences in LHS can also facilitate/hinder opportunities to learn from emotional signals – therefore suggesting an account in which different genetic predispositions can facilitate/hinder 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., serotoninergic 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)). Given that individual differences in neurochemistry have heritable components (e.g., genes for different neurotransmitter receptor variants), and that other related psychological-level individual differences also appear partially heritable (e.g., personality, general intelligence), this opens up yet further research avenues for investigating plausible genetic influences on individual differences in EA (Cohen, Young, Baek, Kessler, & Ranganath, 2005; Lane et al., 1990; Lane, Sechrest et al., 1998; Munafò et al., 2003; Sauce & Matzel, 2018; Smith, Lane et al., 2017; Vernon, Villani, Schermer, & Petrides, 2008; Vukasović & Bratko, 2015; Wright et al., 2018). 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 sex differences in EA are due to socialization vs. due to sex-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 accompanying hypersociality were a 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., expanded 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 each contribute to EA is also an open question.

One way to examine the contribution of domain-general mechanisms would be to test the strength of the relationship between EA and LHS as we have already suggested. Because LHS involves adaptations affecting many psychological domains, if EA and LHS were found to be highly correlated this would support the role of domain-general mechanisms. The relationship between EA and other general cognitive function could also be tested – and has been in previous work showing correlations between EA and both general intelligence and executive control network activation (Smith, Lane et al., 2017; Wright et al., 2018). One way to assess the contribution of domain-specific mechanisms could be to longitudinally examine the relationship between EA and early individual differences in social attention. For example, if higher EA in adulthood were associated with stronger attentional biases (e.g., automatic saccades) toward social cues in early childhood, this would support a strong role of domain-specific social adaptations.

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 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 long-timescale 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 the science of emotion and emotional awareness.

Author contributions

HDS and RDL conceived the initial idea of the paper; RS subsequently contributed substantial additional ideas central to the core thesis of the final manuscript, and RS and HDS wrote the majority of the manuscript; RS, HDS, and RDL edited the manuscript, RDL contributed to and improved sections of the manuscript, and NGS and KLW were involved in the manuscript's early conception and contributed substantial ideas and information that improved the development of the ideas contained in the manuscript.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

RS is supported by the William K. Warren Foundation.

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.

Anderson, M. (2014). *After phrenology: Neural reuse and the interactive brain.* Cambridge, MA: MIT Press.

Andrews, P. W., Gangestad, S. W., & Mathew, D. (2002). Adaptationism - how to carry out an exaptationist program. *Behavioral and Brain Sciences, 25*(4), 489–504. https://doi.org/10.1017/S0140525X02000092.

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.

Atique, B., Erb, M., Gharabaghi, A., Grodd, W., & Anders, S. (2011). Task-specific activity and connectivity within the mentalizing network during emotion and intention mentalizing. *NeuroImage, 55*(4), 1899–1911. https://doi.org/10.1016/j.neuroimage.2010.12.036.

Aviezer, H., Hassin, R., Ryan, J., Grady, C., Susskind, J., Anderson, A., ... 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.* New York: Houghton Mifflin Harcourt.

Barrett, H. C., & Broesch, J. (2012). Prepared social learning about dangerous animals in children. *Evolution and Human Behavior, 33*(5), 499–508. https://doi.org/10.1016/j.evolhumbehav.2012.01.003.

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), 361–372. 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), 419–429. https://doi.org/10.1038/nrn3950.

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.

Barton, R. A., & Venditti, C. (2014). Rapid evolution of the cerebellum in humans and other great apes. *Current Biology, 24*(20), 2440–2444. https://doi.org/10.1016/j.cub.2014.08.056.

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), 163–164. 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., ... 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), 1569–1572. 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. https://doi.org/10.1016/j.neuron.2017.06.038 e5.

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.

Bréjard, V., Bonnet, A., & Pedinielli, J. (2012). The role of temperament and emotional awareness in risk taking in adolescents. *L'Encéphale: Revue de Psychiatrie Clinique Biologique et Thérapeutique, 38*(1), 1–9.

Brown, S. L., & Cialdini, R. (2015). Functional motives and functional consequences of prosocial behavior. In D. A. Schroeder, & W. G. Graziano (Eds.). *The Oxford handbook of prosocial behavior* (pp. 346–361). . 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, NY), 20*(1), 25. https://doi.org/10.1007/S12110-009-9059-3.

Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L., & Friston, K. J. (2018). Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology, 455*, 161–178.

Buckner, R. (2013). The cerebellum and cognitive function: 25 years of insight from anatomy and neuroimaging. *Neuron, 80*(3), 807–815. https://doi.org/10.1016/j.neuron.2013.10.044.

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.

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.

Burger, A., Lumley, M., Carty, J., Latsch, D., Thakur, E., Hyde-Nolan, M., ... 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., ... 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., ... van Schaik, C. P. (2014). The evolutionary origin of human hyper-cooperation. *Nature Communications, 5*(1), 4747. https://doi.org/10.1038/ncomms5747.

Burkart, J. M., Hrdy, S. B., & Van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews, 18*(5), 175–186. https://doi.org/10.1002/evan.20222.

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.

Carter, C. S. (2014). Oxytocin pathways and the evolution of human behavior. *Annual Review of Psychology, 65*(1), 17–39. https://doi.org/10.1146/annurev-psych-010213-115110.

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), 10387–10394. 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: Oxford University Press.

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.* New York: Oxford University Press.

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

Cohen, M., Young, J., Baek, J.-M., Kessler, C., & Ranganath, C. (2005). Individual differences in extraversion and dopamine genetics predict neural reward responses. *Cognitive Brain Research, 25*(3), 851–861. https://doi.org/10.1016/j.cogbrainres.2005.09.018.

Colvert, E., Rutter, M., Kreppner, J., Beckett, C., Castle, J., Groothues, C., ... 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.

Constant, A., Ramstead, M. J., Veissiere, S. P., Campbell, J. O., & Friston, K. J. (2018). A variational approach to niche construction. *Journal of The Royal Society Interface, 15*(141), 20170685.

Corbetta, M., & Shulman, G. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience, 3*(3), 201–215. https://doi.org/10.1038/nrn755.

Courchesne, E., Mouton, P. R., Calhoun, M. E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M. J., ... Pierce, K. (2011). Neuron number and size in prefrontal cortex of children with autism. *JAMA - Journal of the American Medical Association, 306*(18), 2001–2010. https://doi.org/10.1001/jama.2011.1638.

Craig, A. D. (2009). How do you feel–now? The anterior insula and human awareness. *Nature Reviews Neuroscience, 10*(1), 59–70.

Da Costa, L., Parr, T., Sajid, N., Veselic, S., Neacsu, V., & Friston, K. (2020). *Active inference on discrete state-spaces a synthesis* arXiv:2001.07203.

De Dreu, C. K. W., Greer, L. L., Handgraaf, M. J. J., Shalvi, S., Van Kleef, G. A., Baas, M., ... Feith, S. W. W. (2010). The neuropeptide oxytocin regulates parochial altruism in intergroup conflict among humans. *Science, 328*(5984), 1408–1411. https://doi.org/10.1126/science.1189047.

Dehaene, S. (2014). *Consciousness and the brain.* Viking Press.

Dehaene, S., Charles, L., King, J.-R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology, 25*, 76–84. https://doi.org/10.1016/j.conb.2013.12.005.

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), Article 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., ... 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.

Donahue, C. J., Glasser, M. F., Preuss, T. M., Rilling, J. K., & Van Essen, D. C. (2018). Quantitative assessment of prefrontal cortex in humans relative to nonhuman primates. *Proceedings of the National Academy of Sciences, 115*(22), E5183–E5192.

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.

Dunlap, A. S., & Stephens, D. W. (2014). Experimental evolution of prepared learning. *Proceedings of the National Academy of Sciences of the United States of America, 111*(32), 11750–11755. https://doi.org/10.1073/pnas.1404176111.

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.

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.

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.

Feenstra, D. J., Luyten, P., & Bales, D. L. (2017). Mentalization-based treatment for borderline personality disorder in adults and adolescents: For whom, when, and how? *Bulletin of the Menninger Clinic, 81*(3), 264–280. https://doi.org/10.1521/bumc_2017_81_04.

Feldman Barrett, L., & Finlay, B. L. (2018). Concepts, goals and the control of survival-related behaviors. *Current Opinion in Behavioral Sciences, 24*(December 1), 172–179. https://doi.org/10.1016/j.cobeha.2018.10.001.

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., ... 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.

Figueredo, A. J., Hertler, S. C., & Peñaherrera-Aguirre, M. (2020). The biogeography of human diversity in life history strategy. *Evolutionary Behavioral Sciences* (In Press).

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.

Flinn, M. V., Geary, D. C., & Ward, C. V. (2005). Ecological dominance, social competition, and coalitionary arms races: Why humans evolved extraordinary intelligence. *Evolution and Human Behavior, 26*(1), 10–46. https://doi.org/10.1016/j.evolhumbehav.2004.08.005.

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 post-traumatic 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), 815–836. https://doi.org/10.1098/rstb.2005.1622.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nature Reviews. *Neuroscience, 11*(2), 127–138. https://doi.org/10.1038/nrn2787.

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., Doherty, J. O., & 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., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience and Biobehavioral Reviews, 90*, 486–501. https://doi.org/10.1016/J.NEUBIOREV.2018.04.004.

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.

Gat, A. (2015). Proving communal warfare among hunter-gatherers: The quasi-rousseauan error. *Evolutionary Anthropology: Issues, News, and Reviews, 24*(3), 111–126. https://doi.org/10.1002/evan.21446.

Gavrilets, S., & Vose, A. (2006). The dynamics of Machiavellian intelligence. *Proceedings of the National Academy of Sciences of the United States of America, 103*(45), 16823–16828. https://doi.org/10.1073/pnas.0601428103.

Gee, D. G., Gabard-Durnam, L. J., Flannery, J., Goff, B., Humphreys, K. L., Telzer, E. H., ... 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), 15638–15643. 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.

Grove, M. (2017). Environmental complexity, life history, and encephalisation in human evolution. *Biology and Philosophy, 32*(3), 395–420. https://doi.org/10.1007/s10539-017-9564-4.

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.

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. (2017). *The secret of our success. How culture is driving human evolution, domesticating our species, and making us smarter.* Princeton, NJ: Princeton University Press.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*, 61–83. https://doi.org/10.1017/S0140525X0999152X.

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.* Boston, MA: MIT Press.

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.

Hermundstad, A., Bassett, D., Brown, K., Aminoff, E., Clewett, D., Freeman, 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), 6169–6174. https://doi.org/10.1073/pnas.1219562110.

Herrmann, E., Call, J., Hernández-Lloreda, M. V., Hare, B., & Tomasello, M. (2007). Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis. *Science, 317*(5843), 1360–1366. https://doi.org/10.1126/science.1146282.

Heyes, C. M., & Frith, C. D. (2014). The cultural evolution of mind reading. *Science, 344*(6190), 1243091. https://doi.org/10.1126/science.1243091.

Hoffman, K. L., & Gauthier, I. (2007). *Evolutionary specializations for processing faces and objects. Evolution of nervous systems*437–445. https://doi.org/10.1016/B0-12-370878-8/00031-8.

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

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.

Kalbe, E., Schlegel, M., Sack, A. T., Nowak, D. A., Dafotakis, M., Bangard, C., ... Kessler, J. (2010). Dissociating cognitive from affective theory of mind: A TMS study. *Cortex, 46*(6), 769–780. https://doi.org/10.1016/j.cortex.2009.07.010.

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), Article e1000209. https://doi.org/10.1371/journal.pcbi.1000209.

Kihlstrom, J., Mulvaney, S., Tobias, B., & Tobis, I. (2000). The emotional unconscious. In E. Eich, J. Kihlstrom, G. Bower, J. Forgas, & P. Niedenthal (Eds.). *Cognition and emotion* (pp. 30–86). New York: Oxford University Press.

Kingstone, A., Laidlaw, K. E. W., Nasiopoulos, E., & Risko, E. F. (2017). Cognitive ethology and social attention. *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), 1086–1091. 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.

Kleckner, I., Zhang, J., Touroutoglou, A., Chanes, L., Xia, C., Simmons, W., ... Barrett, L. (2017). Evidence for a large-scale brain system supporting allostasis and interoception in humans. *Nature Human Behaviour, 1*, 69. https://doi.org/10.1038/s41562-017-0069.

Kret, M., & De Gelder, B. (2012). A review on sex differences in processing emotional signals. *Neuropsychologia, 50*(7), 1211–1221. https://doi.org/10.1016/j.neuropsychologia.2011.12.022.

Krienen, F., & Buckner, R. (2017). Human association cortex: Expanded, untethered, neotenous, and plastic. In J. Kaas (Vol. Ed.), *Evolution of nervous systems 2e: Vol. 4*, (pp. 169–183). Elsevier.

Kuzawa, C. W., & Bragg, J. M. (2012). Plasticity in human life history strategy: Implications for contemporary human variation and the evolution of genus Homo. *Current Anthropology, 53*(December 17), https://doi.org/10.1086/667410.

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), 124–134. https://doi.org/10.1080/00223891.1990.9674052.

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.

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., Sechrest, L., & Riedel, R. (1998). Sociodemographic correlates of alexithymia. *Comprehensive Psychiatry, 39*(6), 377–385. https://doi.org/10.1016/S0010-440X(98)90051-7.

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.

Leavens, D. A., Bard, K. A., & Hopkins, W. D. (2010). BIZARRE chimpanzees do not represent the chimpanzee. June *Behavioral and Brain Sciences, 33*, 100–101. https://doi.org/10.1017/S0140525X10000166.

Leavens, D. A., Bard, K. A., & Hopkins, W. D. (2019). The mismeasure of ape social cognition. *Animal Cognition, 22*(4), 487–504. https://doi.org/10.1007/s10071-017-1119-1.

LeDoux, J. E. (2012). Rethinking the emotional brain. *Neuron, 73*(4), 653–676. https://doi.org/10.1016/j.neuron.2012.02.004.

LeDoux, J., Brown, R., Pine, D., & Hofmann, S. (2018). *Know thyself: Well-being and subjective experience. Cerebrum : The Dana Forum on brain science*. 2018. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/30746034.

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 J. Kaas (Vol. Ed.), *Evolution of nervous systems 2e: Vol. 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, 170*(Nov 3), 351–360. 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.

Lewontin, R. (1998). The evolution of cognition: Questions we will never answer. In (2nd ed.). D. Osherson, D. Scarborough, & S. Sternberg (Vol. Eds.), *An invitation to cognitive science: Vol. 4*MIT Press.

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., ... 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., ... 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 J. Kaas (Vol. Ed.), *Evolution of nervous systems 2e: Vol. 4*, (pp. 185–205). . https://doi.org/10.1016/B978-0-12-804042-3.00118-4.

Maryanski, A. (2010). WEIRD societies may be more compatible with human nature. *Behavioral and Brain Sciences, 33*, 103–104. https://doi.org/10.1017/S0140525X10000191.

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.

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), 648–655. https://doi.org/10.1016/j.neuroimage.2008.02.030.

Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response. *Brain Structure and Function, 214*(5-6), 535–549.

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 M. Mesulam (Ed.). *Principles of behavioral and cognitive neurology*(2nd ed.). New York: Oxford University Press.

Migliano, A. B., Page, A., Gómez-Gardeñes, J., Viguier, S., Dyble, M., Thompson, J., ... Vinicius, L. (2016). High-resolution maps of hunter-gatherer social networks reveal human adaptation for cultural exchange. *BioRxiv*Article 40154. https://doi.org/10.1101/040154.

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 and 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.

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.

Munafò, M., Clark, T., Moore, L., Payne, E., Walton, R., & Flint, J. (2003). Genetic Polymorphisms and personality in healthy adults: A systematic review and meta-analysis. *Molecular Psychiatry, 8*(5), 471–484. https://doi.org/10.1038/sj.mp.4001326.

Murray, J., Bernacchia, A., Freedman, D., Romo, R., Wallis, J., Cai, X., ... 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. (2017a). Working memory, attention, and salience in active inference. *Scientific Reports, 7*(1), Article 14678. https://doi.org/10.1038/s41598-017-15249-0.

Parr, T., & Friston, K. J. (2017b). Uncertainty, epistemics and active inference. *Journal of The Royal Society Interface, 14*(136), 20170376.

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), 156–166. 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), 10127–10134. https://doi.org/10.1523/JNEUROSCI.2161-10.2010.

Peltola, M. J., Yrttiaho, S., & Leppänen, J. M. (2018). Infants' attention bias to faces as an early marker of social development. *Developmental Science, 21*(6), Article e12687. https://doi.org/10.1111/desc.12687.

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.

Povinelli, D. (2000). *Folk physics for apes: The chimpanzee's theory of how the world works.* Retrieved fromhttps://pdfs.semanticscholar.org/d268/60a37079f30b3d4b7093f1bffa5baf9d16c5.pdf.

Power, J., Cohen, A., Nelson, S., Wig, G., Barnes, K., Church, J., ... 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.

Prinz, J. (2007). Accessed, accessible, and inaccessible: Where to draw the phenomenal line. *Behavioral and Brain Sciences, 30*(5–6), 521–522. https://doi.org/10.1017/S0140525X07003020.

Prinz, J. (2012). *The conscious brain: How attention engenders experience (Philosophy of mind).* USA: Oxford University Press.

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.* 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.

Robson, S. L., & Wood, B. (2008). Hominin life history: Reconstruction and evolution. *Journal of Anatomy, 212*(4), 394–425. https://doi.org/10.1111/j.1469-7580.2008.00867.x.

Roff, D. (2002). *Life history evolution.* MA: Sinauer: Sunderland Associates.

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., ... Rushworth, M. (2011). Social network size affects neural circuits in macaques. *Science, 334*(6056), 697–700. https://doi.org/10.1126/science.1210027.

Sánchez-Villagra, M. R., & van Schaik, C. P. (2019). Evaluating the self-domestication hypothesis of human evolution. *Evolutionary Anthropology: Issues, News, and Reviews, 28*(3), 133–143. https://doi.org/10.1002/evan.21777.

Sauce, B., & Matzel, L. D. (2018). The paradox of intelligence: Heritability and malleability coexist in hidden gene-environment interplay. *Psychological Bulletin, 144*(1), 26–47. https://doi.org/10.1037/bul0000131.

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.

Sear, R. (2020). *Do human 'life history strategies' exist?* Retrieved fromOSF Preprintshttps://osf.io/hjezb.

Seeley, W., Menon, V., Schatzberg, A., Keller, J., Glover, G., Kenna, H., ... Greicius, M. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *The Journal of Neuroscience, 27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007.

Semendeferi, K., Teffer, K., Buxhoeveden, D. P., Park, M. S., Bludau, S., Amunts, 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), 1061–1086.

Sherwood, C. C., Stimpson, C. D., Raghanti, M. A., Wildman, D. E., Uddin, M., Grossman, L. I., ... Hof, P. R. (2006). Evolution of increased glia-neuron ratios in the human frontal cortex. *Proceedings of the National Academy of Sciences of the United States of America, 103*(37), 13606–13611. https://doi.org/10.1073/pnas.0605843103.

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), 466–469. 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, B. H. (2018). *Life history: Evolution of infancy, childhood, and adolescence. The international encyclopedia of anthropology*1–10. https://doi.org/10.1002/9781118924396.wbiea2104.

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., & Lane, R. (2016). Unconscious emotion: A cognitive neuroscientific perspective. *Neuroscience and Biobehavioral Reviews, 69*, 216–238. https://doi.org/10.1016/j.neubiorev.2016.08.013.

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, J. E., Swanson, E. M., Reed, D., & Holekamp, K. E. (2012). Evolution of cooperation among mammalian carnivores and its relevance to hominin evolution. *Current Anthropology, 53*(SUPPL. 6), https://doi.org/10.1086/667653.

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., Khalsa, S. S., & Paulus, M. P. (2019). An active inference approach to dissecting reasons for nonadherence to antidepressants. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging.* https://doi.org/10.1016/j.bpsc.2019.11.012 In Press.

Smith, R., Schwartenbeck, P., Parr, T., & Friston, K. J. (2020). An active inference approach to modeling structure learning: Concept learning as an example case. *Frontiers in Computational Neuroscience, 14*. https://doi.org/10.3389/fncom.2020.00041.

Smith, S., Fox, P., Miller, K., Glahn, D., Fox, P., Mackay, C., ... 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), 13040–13045. https://doi.org/10.1073/pnas.0905267106.

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., ... 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., Kaszniak, A., Katsanis, J., Lane, R., & Nielsen, L. (2019). The importance of identifying underlying process abnormalities in alexithymia: Implications of the three-process model and a single case study illustration. *Consciousness and Cognition, 68*, 33–46. https://doi.org/10.1016/J.CONCOG.2018.12.004.

Smith, R., Killgore, W., Alkozei, A., & Lane, R. (2018). A neuro-cognitive process model of emotional intelligence. *Biological Psychology, 139*, 131–151. https://doi.org/10.1016/J.BIOPSYCHO.2018.10.012.

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., Alkozei, A., Bao, J., Smith, C., Sanova, A., ... 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. D., Parr, T., & Friston, K. J. (2019). Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance. *Neuroscience & Biobehavioral Reviews, 107*, 473–491. https://doi.org/10.1101/681288.

Smith, R., Lane, R., Alkozei, A., Bao, J., Smith, C., Sanova, A., ... 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., 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. (2019). Simulating emotions: An active inference model of emotional state inference and emotion concept learning. *Frontiers in Psychology, 10*, 2844. https://doi.org/10.3389/fpsyg.2019.02844.

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., Sanova, A., Alkozei, A., Lane, R., & Killgore, W. (2018). Higher levels of trait emotional awareness are associated with more efficient global information integration throughout the brain: A graph-theoretic analysis of resting state functional connectivity. *Social Cognitive and Affective Neuroscience, 13*(7), 665–675. https://doi.org/10.1093/scan/nsy047.

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.

Soutschek, A., Burke, C. J., Raja Beharelle, A., Schreiber, R., Weber, S. C., Karipidis, I. I., ... 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.

Spocter, M. A., Hopkins, W. D., Barks, S. K., Bianchi, S., Hehmeyer, A. E., Anderson, S. M., ... Sherwood, C. C. (2012). Neuropil distribution in the cerebral cortex differs between humans and chimpanzees. *The Journal of Comparative Neurology, 520*(13), 2917–2929. https://doi.org/10.1002/cne.23074.

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), 303–317. https://doi.org/10.1016/j.neuroimage.2010.06.016.

Steklis, H. D., & Lane, R. D. (2012). The unique human capacity for emotional awareness: psychological, neuroanatomical, comparative and evolutionary perspectives. In S. Watanabe, & S. Kuczaj (Eds.). *Emotions of Animals and Humans. The Science of the Mind* (pp. 165–205). Tokyo: Springer.

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

Subiaul, F. (2016). What's special about human imitation? A comparison with enculturated apes. *Behavioral Sciences, 6*(3), 13. https://doi.org/10.3390/bs6030013.

Suddendorf, T. (2018). Two key features created the human mind: Inside our heads. *Scientific American, 139*(3), 42–47.

Suddendorf, T., & Corballis, M. (2007). The evolution of foresight: What is mental time travel, and is it unique to humans? *Behavioral and Brain Sciences, 30*(3), 299–313. https://doi.org/10.1017/S0140525X07001975.

Suddendorf, T., Bulley, A., & Miloyan, B. (2018). Prospection and natural selection. *Current Opinion in Behavioral Sciences, 24*(December 1), 26–31. https://doi.org/10.1016/j.cobeha.2018.01.019.

Swales, Heard, H. L., & Williams, J. M. G. (2000). Linehan's Dialectical Behaviour Therapy (DBT) for borderline personality disorder: Overview and adaptation. *Journal of Mental Health, 9*(1), 7–23. https://doi.org/10.1080/09638230016921.

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.

Theeuwes, J., & Van der Stigchel, S. (2006). Faces capture attention: Evidence from inhibition of return. *Visual Cognition, 13*(6), 657–665. https://doi.org/10.1080/13506280500410949.

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., & Call, J. (1997). *Primate cognition.* Retrieved fromhttps://books.google.com/books?hl = en&lr = &id = kszhBwAAQBAJ&oi = fnd&pg = PA3&dq = Tomasello, + M., + %26 + Call, + J. + (1997). + Primate + Cognition. + New + York:Oxford + University + Press.&ots = FUmxgNL0Yb&sig = s2L7iHxSNam8oP0JKDG90g5JQIc.

Tomasello, M., & Carpenter, M. (2007). Shared intentionality. *Developmental Science, 10*(1), 121–125. https://doi.org/10.1111/j.1467-7687.2007.00573.x.

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), 349–360. https://doi.org/10.1002/bies.200800070.

Tschantz, Alexander, Seth, Anil K., & Buckley, Christopher L. (2020). Learning action-oriented models through active inference. *PLoS computational biology, 16*(4), e1007805.

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), Article e98217. https://doi.org/10.1371/journal.pone.0098217.

van den Heuvel, M., & Sporns, O. (2011). Rich-club organization of the human connectome. *The Journal of Neuroscience, 31*(44), 15775–15786. https://doi.org/10.1523/JNEUROSCI.3539-11.2011.

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.

Vernon, P., Villani, V., Schermer, J., & Petrides, K. (2008). Phenotypic and genetic associations between the big five and trait emotional intelligence. *Twin Research and Human Genetics, 11*(5), 524–530. https://doi.org/10.1375/twin.11.5.524.

Vincent, J., Patel, G., Fox, M., Snyder, A., Baker, J., Van Essen, D., ... Raichle, M. (2007). Intrinsic functional architecture in the anaesthetized monkey brain. *Nature, 447*(7140), 83–86. https://doi.org/10.1038/nature05758.

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.

vonHoldt, B. M., Shuldiner, E., Koch, I. J., Kartzinel, R. Y., Hogan, A., Brubaker, L., ... 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), Article e1700398. https://doi.org/10.1126/sciadv.1700398.

Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin, 141*(4), 769–785. https://doi.org/10.1037/bul0000017.

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), Article e1004066. https://doi.org/10.1371/journal.pcbi.1004066.

Weaver, I., Cervoni, N., Champagne, F., D'Alessio, A., Sharma, S., Seckl, J., ... 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., ... 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.

Winkielman, P., & Berridge, K. (2004). Unconscious emotion. *Current Directions in Psychological Science, 13*(3), 120–123. https://doi.org/10.1111/j.0963-7214.2004.00288.x.

Winkielman, P., Berridge, K., & Wilbarger, J. (2005). Unconscious affective reactions to masked happy versus angry faces influence consumption behavior and judgments of value. *Personality & Social Psychology Bulletin, 31*(1), 121–135. https://doi.org/10.1177/0146167204271309.

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. D., & Smith, R. (2018). Sex differences in emotion recognition ability: The mediating role of trait emotional awareness. *Motivation and emotion, 42*(1), 149–160.

Yeo, B., Krienen, F., Sepulcre, J., Sabuncu, M., Lashkari, D., Hollinshead, M., ... Buckner, R. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology, 106*(3), 1125–1165. https://doi.org/10.1152/jn.00338.2011.

Zanarini, M. C., & Wedig, M. M. (2014). *Childhood adversity and the development of borderline personality disorder. Handbook of borderline personality disorder in children and adolescents*265–276. https://doi.org/10.1007/978-1-4939-0591-1_18.

Zemack-Rugar, Y., Bettman, J., & Fitzsimons, G. (2007). The effects of nonconsciously priming emotion concepts on behavior. *Journal of Personality and Social Psychology, 93*(6), 927–939. https://doi.org/10.1037/0022-3514.93.6.927.

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.