frontiers in Psychology

HYPOTHESIS AND THEORY
published: 25 March 2020
doi: 10.3389/fpsyg.2020.00417

A World Unto Itself: Human Communication as Active Inference

Jared Vasil[1], Paul B. Badcock[2,3,4], Axel Constant[5,6,7], Karl Friston[7] and Maxwell J. D. Ramstead[6,7,8]*

[1] Department of Psychology and Neuroscience, Duke University, Durham, NC, United States, [2] Centre for Youth Mental Health, The University of Melbourne, Melbourne, VIC, Australia, [3] Melbourne School of Psychological Sciences, The University of Melbourne, Melbourne, VIC, Australia, [4] Orygen, Melbourne, VIC, Australia, [5] Charles Perkins Centre, The University of Sydney, Camperdown, NSW, Australia, [6] Culture, Mind, and Brain Program, McGill University, Montreal, QC, Canada, [7] Wellcome Centre for Human Neuroimaging, University College London, London, United Kingdom, [8] Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, Montreal, QC, Canada

Recent theoretical work in developmental psychology suggests that humans are predisposed to align their mental states with those of other individuals. One way this manifests is in *cooperative communication*; that is, intentional communication aimed at aligning individuals' mental states with respect to events in their shared environment. This idea has received strong empirical support. The purpose of this paper is to extend this account by proposing an integrative model of the biobehavioral dynamics of cooperative communication. Our formulation is based on *active inference*. Active inference suggests that action-perception cycles operate to minimize uncertainty and optimize an individual's internal model of the world. We propose that humans are characterized by an evolved *adaptive prior belief* that their mental states are aligned with, or similar to, those of conspecifics (i.e., that 'we are the same sort of creature, inhabiting the same sort of niche'). The use of cooperative communication emerges as the principal means to gather evidence for this belief, allowing for the development of a shared narrative that is used to disambiguate interactants' (hidden and inferred) mental states. Thus, by using cooperative communication, individuals effectively attune to a hermeneutic niche composed, in part, of others' mental states; and, reciprocally, attune the niche to their own ends via epistemic niche construction. This means that niche construction enables features of the niche to encode precise, reliable cues about the *deontic* or *shared value* of certain action policies (e.g., the utility of using communicative constructions to disambiguate mental states, given expectations about shared prior beliefs). In turn, the alignment of mental states (prior beliefs) enables the emergence of a novel, contextualizing scale of *cultural* dynamics that encompasses the actions and mental states of the ensemble of interactants and their shared environment. The dynamics of this contextualizing layer of cultural organization feedback, across scales, to constrain the variability of the prior expectations of the individuals who constitute it. Our theory additionally builds upon the active inference literature by introducing a new set of neurobiologically plausible computational hypotheses for cooperative communication. We conclude with directions for future research.

Keywords: cooperative communication, mental state alignment, evolution, development, active inference, adaptive prior, free energy, circular causality

OPEN ACCESS

Edited by:
Andrew G. Ryder, Concordia University, Canada

Reviewed by:
Ronald Fischer, Victoria University of Wellington, New Zealand
Li Su, University of Cambridge, United Kingdom

****Correspondence:***
Jared Vasil
jared.vasil@duke.edu

Specialty section:
This article was submitted to Cultural Psychology, a section of the journal Frontiers in Psychology

Received: *24 June 2019*
Accepted: *24 February 2020*
Published: *25 March 2020*

Citation:
Vasil J, Badcock PB, Constant A, Friston K and Ramstead MJD (2020) A World Unto Itself: Human Communication as Active Inference. Front. Psychol. 11:417. doi: 10.3389/fpsyg.2020.00417

"*The point we emphasize is strong confidence in our original nature*,"

Suzuki (1970/2014, p. 35)

INTRODUCTION

An influential body of recent work on human communication describes it as *cooperative communication*. Cooperative communication is defined as intentional communication aimed at the alignment of mental states between conspecifics (reviewed in Tomasello, 2008, 2014, 2019). This is thought to be one particularly important behavioral manifestation of a broader, species-typical motivation to align mental states with those of others (Tomasello et al., 2005). Some have hypothesized that this motivation is the result of selective pressures acting on human evolution in the context of interdependent collaborative foraging (Tomasello et al., 2012; Whiten and Erdal, 2012). In scenarios where individuals in a group must forage together for resources (food, water, information, etc.), the alignment of multiple individuals' goals, intentions, and attentional processes is necessary for success (e.g., Liebenberg, 2006). This view has been useful for empirical investigation in developmental and comparative psychology (reviewed in Call, 2009; Carpenter and Liebal, 2011; MacLean, 2016).

The purpose of this narrative review is to extend the approach to cooperative communication introduced above by leveraging a recent *active inference formulation* in theoretical neuroscience and biology (Friston, 2012, 2013; Friston and Ao, 2012). This formulation of living systems provides a formal account of the dynamics of belief-guided, embodied action from first principles of biological self-organization (e.g., Friston et al., 2014; Sengupta et al., 2016). A formal account is arguably important, because it forces one to make explicit one's theoretical predictions in experimental and modeling work that investigates the usage, development, and cultural evolution of human communication (e.g., Christiansen and Kirby, 2003; McCauley and Christiansen, 2014). Furthermore, and although this is not the primary focus of this work, by proposing an active inference formulation of cooperative communication, we pave the way for a set of well specified predictions about the neurocomputational dynamics underwriting cooperative communication (Friston, 2010; e.g., Bastos et al., 2012; Adams et al., 2013; Parr and Friston, 2017, 2018). This is important, as precisely formulated neuroscientific hypotheses are largely absent from extant work on cooperative communication.

In brief, active inference is a mathematical formulation of the tendency of living systems to maintain themselves in a restricted set of states (i.e., their phenotypic states) while embedded in a fluctuating, partially observed environment (Friston, 2012, 2013). More precisely, active inference formalizes the structure of exchanges between organisms (individuals and groups) and their environment by explaining how the structure and function of organisms and their ecological niches become *attuned* to, or predictive of, each other (Bruineberg et al., 2018a). In short, active inference suggests that every organism optimizes its internal (generative) model of the world via circular or self-fulfilling action-perception cycles that minimize an upper bound on biophysical surprise (i.e., variational free-energy). In turn, the environment becomes attuned to the organisms that inhabit it (Constant et al., 2019). We will see later that this is formally equivalent to maximizing the evidence for internal or generative models of the world – and that when the world (e.g., the cultural niche) is 'shared,' then the generative models of its denizens become committed to a (reliably) shared narrative.

Following a recent hypothesis of the embodied human brain derived from active inference, called the hierarchically mechanistic mind (Badcock et al., 2019a,b), our proposal combines active inference with substantive research in psychology and allied disciplines that captures the specific evolutionary, developmental, and real-time dynamics that underlie the human capacity for cooperative communication.

A key corollary of this approach is the construct of an *adaptive prior* (Badcock et al., 2019a,b). Adaptive priors are evolutionarily endowed, heritable beliefs[1] that guide characteristic patterns of cognition and behavior in conspecifics. In other words, adaptive priors have been shaped by selection to steer action-perception cycles toward adaptive, unsurprising outcomes (Ramstead et al., 2018; Badcock et al., 2019a,b). Such priors depend upon genetic, epigenetic, and/or cultural inheritance, and often incorporate learned, *empirical priors* gleaned from experience to allow for sensitive adaptation to the local environment (Badcock et al., 2019b). Stated otherwise, adaptive priors effectively constrain the space of prior beliefs learned during ontogeny to enable adaptive action in local cultural niches (Badcock et al., 2019b).

Our proposal is as follows. We suggest that natural selection has endowed humans with an *adaptive prior for alignment*; i.e., an adaptive prior preference for action policies that generate sensory evidence that reliably indicates that their own mental states are aligned with, or similar to, those of conspecifics. This adaptive prior fosters intentional, patterned action sequences that gather *evidence* (i.e., sensory observations) for this belief; that is, that gather evidence for the hypothesis that 'we are the same kind of creature, inhabiting the same kind of niche.' The adaptive prior here functions to bias action and inference by leading agents to actively sample their sensorium in a way that, on average and over time, disambiguates conspecifics' (hidden) mental states. This sampling process is therefore *guided by*, and *generates evidence for*, the belief that our mental states are aligned. In short, we cast cooperative communication as an evidence gathering process; indeed, one that extends across temporally nested scales of analysis. The existence of this process follows from, and only from, an adaptive prior specifying the alignment of individuals' mental states[2]. Cooperative communication can thus be cast as a self-fulfilling prophecy, driven by the belief that we are alike. This belief is

[1]Crucial to our account, in the active inference formulation 'beliefs' refer to (subpersonal) Bayesian beliefs – in the sense of Bayesian belief updating or belief propagation (as opposed to propositional beliefs).

[2]On a deflationary view, this is the only solution that can exist, in terms of minimizing the surprise or free energy of coupled free energy minimizing agents. See below and Friston et al. (2015a) for a fuller discussion in the context of pattern formation.

then characteristically reinforced by the evidence generated by belief-guided communication.

Shweder and Sullivan (1993, p. 506) wrote that "cultural psychology endeavors to understand how such divergences [in the processes that underwrite consciousness] relate to acts of interpretation and to the socially constructed meaning or representation of stimulus events." The present article contributes to the project of cultural psychology and neuroscience (e.g., Han, 2015; and articles in the present collection) by explaining how a cultural milieu can shape and direct the dynamics of individual minds; and, in turn, how individual minds can shape their cultural milieu. We do this by providing an account of sociocultural cognition based on a shared adaptive prior for alignment, drawing on the active inference formulation. In turn, we argue that *how* one's cultural experience manifests in any given time and place – the particular tools one that uses in coming to grips with their world (i.e., words, gestures, and concepts) – is dependent on the history and current contingencies of one's culture and the minds, practices, and places that make it up.

The structure of the remainder of the paper is as follows. In order for readers to appreciate the broader context that underscores our proposal, we devote our second section to a review of some of the key phenomena that underwrite cooperative communication, as emphasized by other theorists to date. In the third section, we introduce relevant aspects of active inference, illustrated by examples drawn from studies of cooperative communication. In the fourth section, we leverage the background provided in the second and third sections to argue that human species-typical adaptive priors prescribe the alignment of one's mental states with those of conspecifics. This latter argument is presented in three subsections. The first subsection focuses on real-time dynamics (i.e., interaction) from the perspectives of an individual and dyad, respectively; the second focuses on ontogeny; and the third focuses on the timescale of cultural evolution. Our paper concludes with a few comments about the limitations of the current proposal of an adaptive prior for alignment. This is complemented with suggested directions for future research.

THEORETICAL BACKGROUND

The Evolutionary Origins of Cooperative Communication

Evolutionarily selected 'mutual expectations of cooperativeness' are thought to motivate the usage of cooperative communication (Tomasello, 2014). From the perspective of evolutionary biology, these expectations can be explained by considering the selective contexts that favored them. One promising candidate is so-called obligate collaborative foraging (Tomasello et al., 2012), where adaptive success in securing food and other resources is marked by a necessary dependence on cooperation with others (also, Baumard et al., 2013). For instance, in mutualistic 'stag hunt' games, a single individual is necessary to obtain a low risk, but low reward, food item (a hare), but two individuals are necessary to obtain a high risk, but high reward, food item (a stag). Here, collaboration appears as the riskier, but more rewarding, option[3]. It is riskier because, to cooperate effectively, the would-be partners must somehow align their mental states – their goals, intentions, and attention (Skyrms, 2001). Cooperative communication is thereby favored as a means to intentionally bring about the alignment of mental states. For instance, in high risk stag hunt scenarios preschool children communicated more, and more often, relative to low risk situations (Duguid et al., 2014). Such joint foraging scenarios may point toward an important and recurrent aspect of the early selective pressures that favored the motivations and skills underlying cooperative communication (Tomasello, 2008; e.g., McLoone and Smead, 2014).

Research examining the communicative behavior of extant non-human primates is crucial for understanding the evolutionarily nascent form of modern humans' communicative motivations and skills (Call and Tomasello, 2007; Mitani, 2009). Such work suggests that, generally speaking, the motivation and skills of non-human primates for intentional communication may have been gradually 'cooperativized' across human evolution (Tomasello, 2014); that is, exapted for both cooperative and competitive purposes with conspecifics. This trajectory may have begun with the usage of gestural communication geared toward simply eliciting specific responses from certain individuals (Call and Tomasello, 2007). For instance, something like ritualized great ape 'attention grabbers' – where an individual has learned that (for a certain conspecific) an action like slapping the ground loudly will likely bring about a desired state of the world (e.g., the initiation of play; Tomasello, 2008) – may have been the evolutionary precursor to certain manifestations of cooperative communication, like declarative pointing (Tomasello, 2019). Indeed, the motivational component is key (Rekers et al., 2011): human-raised non-human great apes will occasionally point for humans (though never for conspecifics). However, they only do this 'selfishly,' that is, only when they expect the gesture to cause the individual to (say) get an out-of-reach object for the ape (Bullinger et al., 2011). In contrast, with cooperative communication, the underlying motive is argued to be 'fundamentally' cooperative (Tomasello, 2019); that is, from the onset of cooperative communication in ontogeny, human infants only appear satisfied following a communicative bid when their communicative partner has aligned their mental states with their own, with respect to the infant's intended referent (reviewed in Carpenter and Liebal, 2011; for comparative considerations, see Carpenter and Call, 2013).

The Developmental Origins of Cooperative Communication

Human infants begin to use cooperative communication to align and coordinate mental states at 9–12 months of age (Carpenter et al., 1998). This window of emergence in ontogeny

[3]In the active inference formulation, below, collaboration is 'rewarding' in the sense of maximizing a shared or *prosocial utility* (Yoshida et al., 2008; Devaine et al., 2014).

is strongly maturationally constrained (Matthews et al., 2012), as evidenced by the emergence of communicative pointing at this age in every cultural setting studied (Callaghan et al., 2011; Liszkowski et al., 2012; Lieven and Stoll, 2013). One way this manifests initially is in declarative pointing gestures directed toward referents in the immediate environment. Experimental work suggests that the goal of infants' communication in such cases is to mutually align emotions, attitudes, and/or thoughts about a referent with another individual (Tomasello et al., 2007; e.g., Liszkowski et al., 2007, 2009). Consistent with this, infants become disgruntled when others ignore their communicative bids for alignment. For instance, Liszkowski et al. (2004) found that infants became unsatisfied with uncooperative adults who ignored infants' communicative bids, who did not provide an emotional response symmetrical to the infant's, and who did not shift the focus of their attention back and forth between the infant and their referent. This suggests that one aspect of the desired state of the world that motivates infants' earliest communication simply *is* alignment with other agents' mental states (Tomasello et al., 2007).

This example illustrates a signal feature of cooperative communication; namely, joint attention to a referent (Tomasello, 2008). There is substantial inconsistency in definitions of joint attention within and across psychological subdisciplines (Siposova and Carpenter, 2019). We follow the lead of Tomasello and colleagues (e.g., Tomasello, 1995) by defining *joint attention* as triadic situations in which two or more individuals possess reliable evidence that all participants are attending to the same referent, *and* that all participants know they are attending to the same referent (i.e., 'attending together'). This formulation of joint attention – in terms of reliable evidence for the mutually inferred alignment of attention (cf. mental states) – fits well with our proposal, which mandates the gathering of reliable evidence for the alignment of mental states.

The importance of joint attention for enabling cooperative communication comes from the fact that joint attention enables, and is enabled by, individuals' capacity to reliably 'ground' their communication in shared referents (Clark, 1996). Grounding creates something called common ground (Clark and Brennan, 1991). *Common ground* is the set of mental states (knowledge, beliefs, emotions, etc.) that is inferred to be reliably shared with others (Clark, 1996; Gadamer, 2004; Tomasello, 2014). The capacity to regulate communication with others by leveraging joint attention and common ground is present from the onset of cooperative communication (Tomasello et al., 2007). For instance, young infants use their shared experience with a particular person to interpret and produce utterances and pointing gestures directed toward that individual (Tomasello and Haberl, 2003; Ganea and Saylor, 2007; Saylor and Ganea, 2007; Liebal et al., 2009, 2010).

Moreover, part of regulating communication with respect to common ground is understanding, for instance, that one must try to 'fit' their communication to the inferred needs of another (Clark and Wilkes-Gibbs, 1986). As a simple example of this kind of 'perspectivizing' (Verhagen, 2007) or 'recipient design' (Schegloff, 2006) process, consider that how one chooses to talk about an artifact varies as a function of the inferred amount of cultural common ground shared with one's interlocutor. In the presence of much cultural common ground, a communicator might opt for brevity; and conversely, in the presence of less cultural common ground, one might use more precise (explicit, descriptive) language. For instance, when conversing with someone from non-Western cultural groups, one might employ the more cumbersome, longer descriptive utterance "the jolly old man in a red suit who gives presents to children" instead of the shorter proper name "Santa Claus". The upshot is that, in general, more common ground means less communication is needed to align mental states to a sufficient degree, and less common ground means more communication is required (Tomasello, 2008). In other words, the amount of information necessary to align mental states to a degree adequate to enable cooperative behavior within a given context is inversely proportional to the amount of common ground.

This turns on an important point: the optimization of relevance in cooperative communication (Sperber and Wilson, 1986). *Relevance* refers to the complexity-accuracy trade-off involved in the production and interpretation of communication; e.g., the trade-off between simplicity or compressibility, and meaningfulness or expressivity. A useful way to think about how this trade-off is finessed is in terms of communicative constructions (Goldberg, 2003). *Communicative constructions* are patterned pairings of form and meaning (e.g., word parts and order, intonation) whose synchronic use and form are the result of diachronic patterns of use and associated intergenerational transmission (e.g., processes of grammaticalization and reanalysis; Heine and Kuteva, 2002; Bybee, 2010). Cooperative communicators use communicative constructions to communicate (and thereby align their mental states).

Optimizing relevance, for a speaker, therefore means using the most minimal form that is expected to enable a listener to recover (something sufficiently similar to) the intended meaning (Kanwal et al., 2017); and for a listener, it means inferring the most parsimonious meaning that sufficiently explains the speaker's intentions (Kao et al., 2014; see Goodman and Frank, 2016). This means, as above, that individuals sharing more common ground require less form to adequately align mental states, while those sharing less common ground require relatively more form (Winters et al., 2018). Relatedly, simpler propositions generally require less form to convey, and more complex propositions require more form (Kemmer, 2003). Producing and interpreting relevant communicative constructions thus has implications across the communicative signal, which spans from (e.g.) lexical selection and word order choice to the sequencing of particular phonemes and intonation patterns (Aylett and Turk, 2004).

Importantly, how might an individual recognize another's intention to generate an act of communication intended 'for' oneself in the first place (e.g., Behne et al., 2005)? From another perspective, how might one make mutually apparent one's proximate motivation to align mental states, that one is communicating 'for' another individual? To this

end, researchers have proposed that *ostensive cues* (Sperber and Wilson, 1986), like eye contact, spatiotemporal contingency, and the communicative (e.g., vocal) signal itself, play an important role in making mutually apparent an agent's intentions to communicate information intended to align mental states (reviewed in Csibra, 2010; indeed, Tomasello, 2014, synonymously calls cooperative communication 'ostensive-inferential' communication). Ostensive cues work by 'grabbing' the attention of others to redirect it triadically (i.e., toward the intended referent) so as to comment on it (Szufnarowska et al., 2014). Thus, via their modulatory effects on the allocation of (joint) attention, ostensive cues play a critical (if indirect) role in increasing individuals' common ground and enhancing the reliability of one's inferences about this common ground (e.g., Moll et al., 2007). This has important downstream effects on subsequent behavior. For example, communicative eye contact causes preschoolers to quickly infer another's desire to collaboratively play a stag hunt game (Wyman et al., 2013; Siposova et al., 2018). Moreover, via their effects on attention, ostensive cues play an important role in guiding inductive inference and top-down categorization processes throughout ontogeny (Butler and Tomasello, 2016; Kovács et al., 2017).

Taking Stock

In sum, five key components characterizing cooperative communication were noted in this section. Discussion of these components structures much of the fourth section. First, great apes do not characteristically employ communication geared toward aligning mental states with conspecifics. Moreover, something like the motivations and skills underlying the communication of great apes likely served as a precursor to the evolution of cooperative communication in humans. Second, human communication is fueled by a motivation to align and coordinate mental states with conspecifics. This is a kind of mutual expectation of cooperativeness that is manifest most basically in processes of joint attention, which serves as a kind of 'evolutionarily endowed' common ground that gets the process of communication 'off the ground' in human ontogeny. Third, individuals using cooperative communication optimize the relevance of their communication, that is, the produced and inferred expressiveness of the communicative signal with respect to the production and processing costs of that signal. This depends on the common ground shared by interlocutors, such that, all else being equal, more common ground means less communication and less common ground demands more communication. Fourth, ostensive cues signal one's intention to communicate to another individual (and help one to disambiguate another's intention to communicate to oneself). These are cues like eye contact, contingency, and the speech signal itself.

Fifth and finally, it is useful to highlight that cooperative communication typically manifests, particularly in early ontogeny, as a circular or bidirectional flow of information (note, e.g., the double-arrowed base of the canonical 'joint attentional triangle'; Carpenter and Liebal, 2011). Thus, although we introduced cooperative communication by focusing largely on individual imperatives, it is a fundamentally collaborative process (Clark and Wilkes-Gibbs, 1986). The usage of cooperative communication is a relevance-optimized exchange of perspectives that manifests as a circular process of 'least collaborative effort' (Clark and Brennan, 1991). This characteristic circularity endows individuals with a single shared narrative constituted by their individual perspectives and roles in the collaborative exchange. **Table 1** summarizes these points along with several others introduced in the fourth section.

Implicit in the preceding discussion is the idea that it would be *surprising* – that is, highly atypical – to find an adult human without a communicative system that they could employ to align their mental states with those of others. In this sense, the usage of cooperative communication is a predictable, or expected, aspect that characterizes part of the human phenotype. A question one might ask is, How does this expectation over species-typical states (i.e., this aspect of the phenotype) persist, robustly, across time and (action in) a fluctuating niche?

TABLE 1 | Summary of key features circumscribing cooperative communication.

Scale of analysis	Characteristic dynamics and processes of cooperative communication
Real-time (mechanism)	• Ostension
	• Joint attention
	• Relevance optimization
	• Coupled, bidirectional flow of information
	• Proximate motivation to align and coordinate mental states (e.g., declarative, interrogative, and informative motives)
Development (ontogeny)	• Clearest behavioral onset at ∼9–12 months of age (i.e., cooperative pointing)
	• Gradual alignment with a conventionalized communicative system
Cultural evolution (phylogeny)	• Historical development of a communicative system (e.g., grammaticalization, syntactic reanalysis, and semantic bleaching)
	• Diversification of communicative systems across time, space, and speaker communities
Biological evolution (adaptation)	• 'Cooperativization' of non-human great ape communicative motives and skills (e.g., non-human ape attention getters as precursor to human declarative pointing)
	• Ultimate motivation to align and coordinate mental states (i.e., mutual expectations of cooperativeness)

Certain features, e.g., alignment with a communicative system, are discussed in the fourth section. This paper associates these phenomena with the corresponding scale of dynamics underwritten by the free-energy formulation and, more substantively, the hierarchically mechanistic mind (see Figure 4 in Badcock et al., 2019a).

ACTIVE INFERENCE

Active inference is a theory of belief-guided adaptive action (Friston et al., 2017a). It is a mathematical framework that models the processes by which organisms and their niche come to 'fit' or become 'attuned' to each other (for an introduction to the mathematical apparatus of active inference, see Bogacz, 2017; Buckley et al., 2017). In other words, active inference describes the manner in which organisms and their environments come to possess statistical properties that are predictable from each other (Bruineberg et al., 2018b; Constant et al., 2018). On this view, organisms come to embody statistical models of their ecological niche via perception and learning, and both cultural and natural selection (i.e., empirical and adaptive priors, respectively). Reciprocally, organisms modify their niche to fit their prior beliefs via adaptive action and niche construction. As detailed in **Figure 1**, the models in this formulation are 'generative models' that recapitulate the causal independencies between the factors that generate their sensory input (i.e., how the niche causes their sensory data; e.g., Hinton, 2007; LeCun et al., 2015). In active inference, organisms are, roughly speaking, normative models of what *ought* to be the case, given 'the kind of creature that I am' (Friston, 2011).

The main theoretical suggestion of this paper is that human individuals appear, characteristically (i.e., species-typically), to be endowed with an *adaptive prior that one's mental states are aligned with those of conspecifics*. Now, for human agents, the mental states of other agents are unobservable or 'hidden' states that need to be inferred on the basis of perceptual cues (e.g., gaze direction, posture, facial expression). In other words, mental state alignment is an inference problem: to align with others, an agent must infer the latent or hidden causes (i.e., mental states) that generate observable consequences (i.e., actions). Thus, for agents whose niche includes the mental states of other agents, the set of actions that resolve uncertainty about the niche must comprise actions that reliably disambiguate others' mental states[4]. We suggest that this is precisely the situation brought about by the presence of an adaptive prior for alignment. This adaptive prior fosters specific, patterned forms of (communicative) action and inference that are aimed at disambiguating the mental states of other agents. The characteristic result of this process is the alignment of mental states between conspecifics. The alignment process enables and maintains reliable hypotheses about shared narratives that contextualize our experience (Friston and Frith, 2015a).

Active Inference, Adaptive Priors, and Alignment

In active inference, actions are generated by hierarchically organized *policies* (beliefs about action). The policy pursued by an organism at a particular time is the one that minimizes an information-theoretic *variational free energy* term (Friston et al., 2015b; for a review of variational inference, see Blei et al., 2017). Roughly speaking, free energy quantifies the discrepancy between what an agent expects or prefers to sense and what it actually senses. This conception of free energy is closely related to prediction error (i.e., the mismatch between predicted and observed sensations; Clark, 2013). A complementary view of free energy is that it scores the (negative log) evidence for the internal model generating predictions, in the sense that sensory data that conform to predictions provide evidence for the veracity of the agent's generative model. In short, minimizing free energy is the same as soliciting sensory evidence for one's model of the world (sometimes known as 'self-evidencing'; Hohwy, 2016). On this view, we are our own existence proofs.

The free energy expected under a policy tracks the probability of that particular policy being pursued (i.e., of that specific policy being selected to guide action). Relatively less expected free energy indicates a relatively more probable policy (Friston et al., 2015b; Pezzulo et al., 2018; relatedly, Cisek, 2007). Expected free energy can be decomposed into two terms: *epistemic value* (the information gain of an observation), and *pragmatic value* (the expected log evidence of some outcome, given a generative model of how outcomes depend on action). The relative influence of each term quantifies the degree to which a particular policy generates actions that explore the niche (i.e., exploration), or actions that leverage reliable expectations about the niche to secure preferred outcomes[5] (i.e., exploitation) (Friston et al., 2015b). This is depicted in **Figure 1**[6].

Salient policies are those that have a high epistemic value or affordance (Parr and Friston, 2017). These energize actions that enable an agent to learn the statistical regularities of its environment (see caption, **Figure 1**). This, in turn, enables pragmatic imperatives to foster actions that capitalize on learned regularities (Friston et al., 2015b). For example, repeated exposure to the sensory phenomena characteristic of their culture's communicative constructions leads infants to become familiar with the statistical properties of those constructions (Romberg and Saffran, 2010). In turn, increasingly precise expectations about hidden causes may lead infants to prefer gathering information from speakers of their native language relative to speakers of a foreign language (e.g., Begus et al., 2016; Marno et al., 2016). This is predicted by the hypothesis that agents exploit their familiarity with the sensory phenomena characteristic of their culture's communicative constructions

[4]This is described anthropomorphically, as though individuals are explicitly engaged in an inference like 'I want to attune the statistics of our brains.' Clearly, this is not the case. Rather, this language is used for expository purposes. Indeed, this descriptive tendency is essentially the same as that used by, e.g., Tomasello (2019), where talk of humans' ultimate motivation to align and coordinate mental states often surfaces in place of proximate instantiations thereof. We thank L. Li (personal communication, December, 2018) for noting this ambiguity.

[5]These outcomes are *a priori* preferred because they are the least surprising ones; i.e., outcomes that 'a creature like me' would expect to encounter.

[6]A mathematically equivalent but complementary division of expected free energy is into *ambiguity* and *risk*. Ambiguity is the expected uncertainty about outcomes given some state in the future, while risk is the divergence between anticipated and preferred outcomes. Interestingly, risk is also the expected *complexity cost* found in statistics. This means that minimizing expected free energy – by selecting the right kind of policies – implicitly minimizes the complexity cost of inference. This is exactly the imperative established in the previous section; namely to find 'common ground' that minimizes communication cost. This particular perspective can be traced back to the foundational principles of universal computation, where variational free energy is often discussed in terms of minimum description or message lengths. See MacKay (1995), Wallace and Dowe (1999), and Schmidhuber (2010) for more discussion.

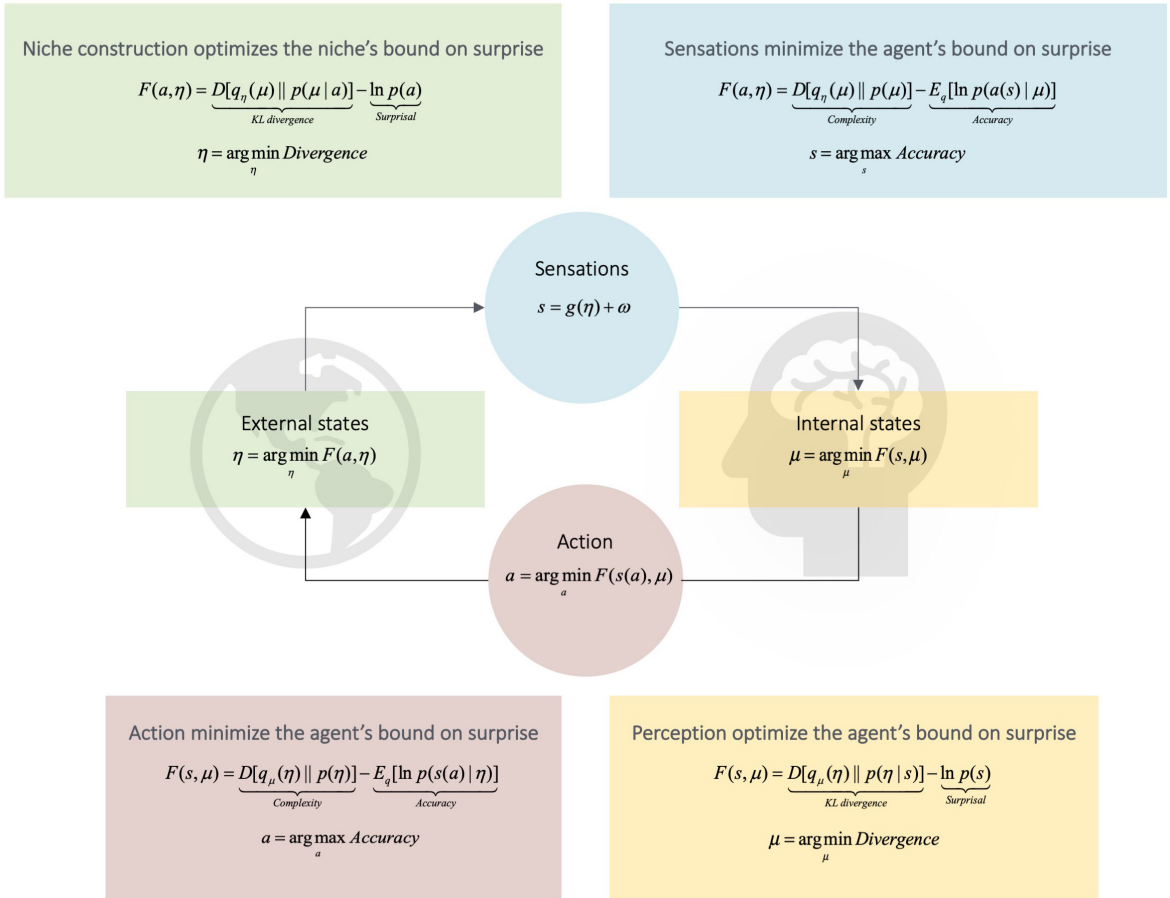

FIGURE 1 | Active inference. This Figure schematizes active inference. It depicts the coupling of an agent's internal states (the dynamics of which entail predictions or beliefs about the niche, μ) to its external states (the dynamics of the agent's niche, η). **Middle Panel:** The influence of the niche on the agent is given by the dynamics of the agent's sensations, *s*. Reciprocally, the influence of the agent upon its niche is given by the agent's action, *a*, upon the niche. This means that the niche is not directly observable from the perspective of an agent's internal states; and the agent's internal states are not directly observable from the perspective of the niche. From an agent's perspective, the niche is thus described as a set of *hidden variables*. Hidden variables must be inferred (i.e., predicted) from sensory observations. Thus, to minimize the probability of sampling surprising sensory states, the task for the agent is to attune the dynamics of internal states to those of the niche; or attune the dynamics of the niche to those of internal states. Attunement renders the agent an approximate (predictive) model of the hidden causes of its sensations. We can quantify the degree of attunement between organism and niche with a quantity called variational free energy (Bruineberg et al., 2018a; Constant et al., 2018). Free energy *F* bounds (i.e., is greater or equal to) the *surprisal –lnp(s)* associated with a sensation (Friston, 2010). Importantly, free energy is a function of two quantities to which the organism has access, namely, its sensations and predictions (for discussion, see Bruineberg et al., 2018a). **Lower Panel:** The bottom right details how perception optimizes free energy by implicitly minimizing a Kullback–Leibler (KL) divergence term *D*. The KL divergence tracks the statistical similarity of two distributions (Cover and Thomas, 1991); e.g., the similarity of prior beliefs about the state of the niche with posterior beliefs (Friston, 2012). Because the KL divergence provides an upper bound on surprisal, minimizing it renders the agent a model of the niche and thus implicitly bounds the surprise of sensory states. **Upper Panel:** These expressions define the relationship of the niche to the agent. Note the kind of 'mirror image' relationship between the equations in the (upper panel) with the equations in the lower. This relationship is a consequence of the mathematics of free energy minimization (see Bruineberg et al., 2018b; Constant et al., 2018). It means that the niche 'sees' and 'learns' about the agent (i.e., via the agent's action) in the same way an agent sees and learns about their niche (i.e., via the niche's 'action'). This insight is extended in **Figure 2**. Adapted with permission from Veissière et al. (forthcoming).

to guide attention toward sensory stimuli that is expected to be useful for disambiguating the mental states of others (**Figures 1**, **2**).

In active inference, the folk-psychological term 'attention' refers to two distinct, but closely related, phenomena; namely, epistemic value and *precision weighting* (Parr and Friston, 2017). Epistemic value, salience, or affordance is the component of policy selection just discussed; it is that component of the value of policies that tracks how much a policy reduces uncertainty about the state of the world (e.g., Friston et al., 2012). It provides a description of the folk-psychological phenomenon of actively orienting toward or 'turning one's attention' to a certain modality or part of the sensory field (e.g., in visual saccades that sample a particular location in visual space). In short, salience or epistemic affordance is an attribute of how we sample the world – in the sense that actively sampling sensory information will reduce uncertainty, in relation to our current beliefs. In contrast, precision is an attribute of the sensory data *per se*.

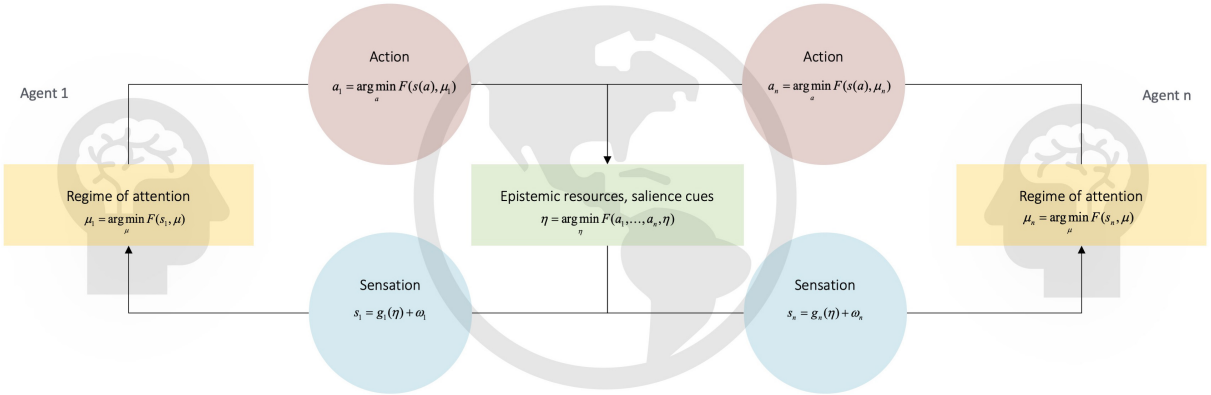

FIGURE 2 | Thinking through other minds. This Figure depicts a set of heuristic equations that describe the kind of free energy minimization hypothesized to underwrite the acquisition and production of learned cultural behaviors via the coupled dynamics sketched in the main text (full equations in **Figure 1**). In the context of human communication, coupled dynamics are energized by an adaptive prior for alignment. The adaptive prior for alignment specifies the characteristically enhanced precision of the hypothesis that 'we' exist. This prior motivates similar agents to actively couple their respective actions a_n and sensations s_n. Via the processes discussed in the main text, this statistical coupling of sensation and action enables each individual to reliably align with (i.e., infer) the hidden states μ_n of conspecific *n*. This circular process brings about a process of cultural niche construction that creates, maintains, and modifies a set of predictable epistemic (i.e., deontic) resources, η. These specify a set of high value (i.e., predictable) observation-policy mappings, which are used to disambiguate the mental states of conspecifics (Veissière et al., forthcoming). One important class of deontic resource is the set of observation-policy mappings that underwrite a system of communicative constructions (i.e., form-meaning pairings). This means that the use of communicative constructions plays a critical role in enabling agents with an adaptive prior for alignment to effectively disambiguate external states. This is because an agent's external states are constituted, in part, by the internal, mental states of another agent (and vice versa). This follows from the fact that external states cause sensation; for an agent equipped with an adaptive prior for alignment, inferring the motion of external states entails inferring other agents' hidden states. The production and observation of communicative constructions is useful because it effectively and flexibly guides 'regimes' of attention that enable species unique forms of cultural learning (see Ramstead et al., 2016; Veissière et al., forthcoming). Diachronically, communicative constructions are finessed by a community of agents via the inheritance and (intended or unintended) modification of constructions during either learning or usage. Adapted with permission from Veissière et al. (forthcoming).

Imprecise sensory data should have less effect on (Bayesian) belief updating, relative to precise information. It is therefore important to afford the right precision to each sensory sample, via precision weighting.

Precision-weighing is the related (but distinct) attentional process that determines the relative influence of bottom-up error signals and top-down expectations in the brain; e.g., a high precision on sensory signals corresponds to low confidence in top-down beliefs (Clark, 2013; Powers et al., 2016). That is, in the sense of precision-weighting, 'attention' refers to the optimization of the precision (inverse variance) of prior beliefs about the causes of sensory data, relative to the precision of those data; in other words, attentional selection is in the game of selecting the right sort of sensory information for belief updating. This precision weighting in the brain is thought to be mediated by the modulation of neuronal gain (Kanai et al., 2015). Precise (attended, ascending) error signals then serve to modulate action and direct what is learned (Feldman and Friston, 2010; Adams et al., 2013). The complement of this attentional selection is the attenuation of precision; known in psychophysics as *sensory attenuation*; i.e., attending away from or ignoring certain sensations; particularly those we cause ourselves.

Crucially, selective attention and attenuation of precision can be part of the covert (mental) actions that are entailed by a policy. In other words, when selecting the policy that minimizes expected free energy we are also committing to both overt action on the (embodied) world – through moving, blushing, speaking etc. – *and* a covert attentional set. We will now illustrate these aspects (orienting to salient stimuli and attentional selection) of active inference with two examples.

As a first example, in the case of human communication, orienting to salient sensory streams should enhance the ability to learn the causes (i.e., mental states) generating sensory evidence by making beliefs about mental states generating that stream more probable. With this in mind, note that one common motivation for infants' and young children's communication is quintessentially uncertainty resolving and 'interrogative' (Begus and Southgate, 2012; Harris et al., 2017). For instance, infants' pointing can function as a request for information about the name or function of objects (Begus and Southgate, 2012; Kovács et al., 2014). It is thus interesting that, in line with the present account, orienting to (the sensory effects of) one's communicative bids enhances learning of (e.g.) communicative constructions and object functions (Begus et al., 2014; Lucca and Wilbourn, 2018a,b; see Friston and Frith, 2015a). In short, infants evince sophisticated policies for resolving uncertainty and creating opportunities for epistemic foraging. In turn, attending to and learning the causes of the communicative stream then enables policies to exploit prior beliefs about how such sensations were caused; that is, inferring whether or not we are aligned, based on the evidence generated through our interactions (e.g., in using learned constructions to ask, explicitly, 'Do you understand?'). This brings us to our second example.

For agents who expect their predictions to be fulfilled, individuals who do not provide evidence for this expectation – despite one's attempts to actively attune mental states – should come to be treated as imprecise sources of sensory information, relative to others that fulfill their expected 'role' in the evidence gathering process; i.e., others that are afforded epistemic trust (Fonagy and Allison, 2014). In other words, in a given communicative interaction, salient policies are those that are expected to be useful with respect to the alignment of mental states; e.g., in certain instances of conversational repair (Schegloff et al., 1977). Across interactions with specific others, repeatedly experiencing surprising responses (i.e., insufficient evidence for, or evidence against, alignment) means that selective attention toward those specific others comes to be afforded low precision (i.e., ignored). Subsequently, action should lead the appearance, on average across time, of avoiding such unreliable parts of the niche (Constant et al., 2018) – much as we tend to avoid the dark when searching for something (Demirdjian et al., 2005).

We suggest that this provides an explanation of the findings by Liszkowski et al. (2004), discussed above, which reported that 12-month-old were dissatisfied with an uncooperative adult who failed to provide both look-backs between the infant and their intended referent and the same emotional response as the infant in response to the infant's communicative bids. On our view, infants were attempting a kind of fast 'error correction' by generating actions expected to minimize exposure to unexpected cues (i.e., allostatic control; see Pezzulo et al., 2015). This occurred via a rapid increase in the salience of policies that generate pointing behavior when sampling sensory data that was inconsistent with infants' prior beliefs about alignment. Moreover, only the group of infants who attempted to communicate with an uncooperative adult pointed significantly less across trials; through the lens of active inference, they had revised their expectations about the sensory effects of action, leading them to select other policies.

This second example suggests that, within and across trials of the experiment, infants appeared to climb an evidence gradient for their expectations. That is, repeated orienting to cues indicative of the (dis)alignment of prior beliefs – despite allostatic control geared toward avoiding such surprising encounters – caused infants to infer and learn that their interaction partner was unhelpful with regards to gathering evidence for their (species-typical) prior beliefs. For the infant, orienting to the sensory consequences of repeated failed attempts to elicit evidence from the adult indicative of alignment (e.g., look-backs and symmetrical emotions) had an impact on the expected free energy of policies. In particular, policies geared toward inferring the prior beliefs of the uncooperative adult came to be characterized by a relatively high expected free energy. Consequently, such policies became relatively unlikely to gain control over action; i.e., less communication with that adult.

In sum, by suggesting that humans are characterized by an adaptive prior for alignment, we effectively argue that policies expected to disambiguate others' mental states are characterized by a low expected free energy. This is by virtue of their high epistemic affordance (i.e., in a niche partly constituted by others' mental states). Consequently, these policies tend to dominate action – people tend to gesticulate and talk with others. Repeatedly leveraging this belief to guide context-sensitive patterns of action, in turn, enables agents to learn the structure and dynamics of their niche. Because the human niche includes others' mental states, beliefs about how to act to effectively infer and align with others will have high adaptive value (Constant et al., forthcoming). This means that learning likely entails refining one's set of 'communicative policies' to approximate the set of policies expected (i.e., typically used) in one's cultural milieu. In short, leveraging communicative constructions means converging on the mutually inferred, or deontic, value of policies geared toward disambiguating mental states among agents equipped with an adaptive prior for alignment.

Deontic Value: Shared Expectations About the Value of Policies

Above we assumed that the prior beliefs of conspecifics had converged on the set of constructions leveraged in their cultural niche. This assumption is important, as our argument considers the acquisition and (cultural) evolution of communicative constructions (below). Within active inference, the concept of *shared* or *deontic value* – and associated *deontic cues* – (Constant et al., 2018, 2019) may be useful for understanding the emergence of cooperative communication in ontogeny and cultural evolution.

The deontic value of a policy rests on a direct ('automatized') likelihood mapping between learned cues and associated action policies. The mapping from deontic cue to policy is 'direct' in the sense that observation of a deontic cue comes to 'automatically' elicit an associated (i.e., learned) policy[7]. Deontic cues are observations that trigger such automatic, or habitual, policy selection (Constant et al., 2019). Encultured agents learn deontic observation-policy mappings in development, through their engagement with the deontic cues that populate their local cultural niche (e.g., Chukoskie et al., 2013). By 'offloading' cognition into the environment in this way (see Clark, 2006, 2008), the direct mapping enables individuals to bypass costly updates to, and metabolic upkeep of, their beliefs about what to do (given what is inferred of the niche). This allows agents to rely directly on deontic cues to select the most appropriate policy (Constant et al., 2018). There is clearly a close relationship between deontic cues, semiotics, and signs (Sewell, 1992; Goodwin, 2000) that underwrite communication. Perhaps the most celebrated system of encultured deontic cues is language itself.

For instance, consider an individual who has learned the English construction 'let alone' (Fillmore et al., 1988); that is, a communicative construction marked by a comparative 'let alone' phrase centered between clause X and clause (fragment)

[7]Technically, expected free energy is combined with deontic value to score the likelihood of a particular policy. In the absence of deontic value, the expected free energy will select the most apt epistemic and *goal directed* policy, given beliefs about the current state of the world. Conversely, if certain cues render the deontic value of a policy sufficiently high, it will dominate policy selection – and emerge as a *habit*.

Y; e.g., 'I could barely run 1 mile let alone 4 miles.' Learning the 'let alone' construction, as one example of a more general phenomenon (see subsection "Dynamics at the Timescale of Ontogeny"), entails learning the deontic value of cues (for policies that parse spoken or written language). In short, if I hear you utter the phrase 'X' and possess prior, reliably shared knowledge of the construction 'X let alone Y,' then I can reliably expect you to follow up with 'let alone Y.' This example assumes a probabilistic (generative) model of how communicative sensations are caused (e.g., a scheme to reliably parse syntax; Levy, 2008; reviewed in Kuperberg and Jaeger, 2016). In particular, this turns on the acquisition of the deontic value of linguistic policies entailed by the hypothesis that one is witnessing a 'let alone' construction.

But how do such reliable mappings come to exist in the first place? That is, how do communicative constructions 'build up' over (neurodevelopmental or evolutionary) time? Consider a simple example: continually walking along the same path across a park each day wears down the grass along that path (Constant et al., 2018). As the grass wears down and a clear path forms, one learns to expect the associated sensory cues when revisiting the path. Because of this, the path becomes increasingly salient for both oneself and for others 'like me,' who can (like me) leverage such 'meaningful' traces left by my actions at later time points. Consequently, the cognitive processing associated with answering the question 'Where ought I to walk next' is afforded directly by physical features of the niche. This saves on the costs associated with planning as active inference (Attias, 2003; Botvinick and Toussaint, 2012; Baker and Tenenbaum, 2014; Mirza et al., 2016) – the inference is literally 'offloaded' into the environment (see equations in **Figures 1**, **2**). The niche provides a clue as to what to do, reliably, as a deontic cue.

Crucially, when this process of 'carving out' deontic *cues* in the niche is performed by an increasing number of agents, the deontic value of policies and associated cues becomes increasingly robust to perturbations. In other words, the expectations of the social niche – here, the set of form-meaning pairings constituting a communicative system – become increasingly precise with increases in the number of interactions between agents constituting that system (Constant et al., 2019). Increasingly precise, niche-based expectations mean that agents become more likely to sensitize their behavior to that cue; e.g., the dynamics of a cue become sufficiently precise so as to enable learning of that cue and its associated action policy in ontogeny (below). In multi-agent systems equipped with an adaptive prior to align mental states, learning of deontic value (i.e., inferring the most common policies undertaken by other denizens of the niche) is learning the 'shared' value of a policy – the value of a policy for people 'like me' in our community (**Figure 2**).

What might it mean to offload cognition into the environment in the fashion above, for agents equipped with an adaptive prior to attune mental states? Individuals effectively outsource solutions to the problem of 'How ought I to talk' to the niche itself. The traces left by repeatedly aligning mental states via communication may enable the niche to subsequently afford increasingly precise, shared expectations about how other agents 'like me' (should) act so as to align mental states most effectively (e.g., during evolutionarily relevant stag hunt scenarios; Grau-Moya et al., 2013). In principle, this takes pressure off inferring 'what sort of person am I in this context' (Moutoussis et al., 2014). Technically, it finesses the computational cost of belief updating from (deontically installed) priors to posterior beliefs about behaviors that are apt for the current setting. Consequently, the cue (or sequence of cues) may come to be preferred by both individuals during subsequent interactions in similar contexts (Schelling, 1960; Lewis, 1969; e.g., Clark and Wilkes-Gibbs, 1986). Formally speaking, at later instances of interaction, the expected free energy of historically selected policies – leveraged to align mental states – falls; such policies then tend to be selected to generate predictable action sequences geared toward the alignment of mental states (Friston and Frith, 2015b).

HUMAN COMMUNICATION AS ACTIVE INFERENCE

This section provides a discussion of our proposal. The species-typical motivation to align mental states with conspecifics is cast as an adaptive prior preference for alignment. This, we suggest, provides the basis for a normative framework for predicting, explaining, and modeling the behavioral, psychological, and neural underpinnings of cooperative communication. Our discussion in this section telescopes from considerations at the microscale (i.e., mechanism), to the mesoscale (i.e., ontogeny), and, finally, to the macroscale (i.e., cultural evolution).

Dynamics at the Timescale of Mechanism: The Individual in Context

A central part of the content of the prior for alignment is that the actions of agents (e.g., oneself) update the mental states (prior beliefs) of other agents. Because mental states cause action (and, hence, observations), gathering reliable evidence for this prior means that agents orient to the individual(s) toward whom their action is directed – the sensory consequences of one's action are realized by the actions of others. That is, if one expects to infer others' mental states, the only evidence available is found in the observed consequences of others' actions (**Figure 3**). Indeed, policies that direct action toward others – so as to disambiguate their mental states (e.g., attentional orienting and, later, cooperative pointing) – possess an evolutionarily unique (Seyfarth and Cheney, 2003), maturationally constrained salience from early in life (Reddy, 2003; Matthews et al., 2012). Gathering evidence for these expectations manifests in *coupled action-perception cycles* (Friston and Frith, 2015a); i.e., intentionally co-constructed loops of action-perception that induce a reliable statistical coupling between two coupled agents (reviewed in, e.g., Hasson et al., 2012; Feldman, 2015; Hasson and Frith, 2016). For expository purposes, we may say that, within the coupled action-perception cycle of human agents, evidence for the self amounts to evidence for the other; and evidence for the other is evidence for the self.

FIGURE 3 | One canonical 'loop' of the coupled action-perception cycle. This example is 'canonical' in the sense that the manifestation of the coupled action-perception cycle in a given instance may vary as a function of context and the experience of its constituent members (e.g., an infant's communicative needs with an adult are different than a pair of adults'). With this in mind, for two agents A and B expecting to reliably infer each other's mental states, the beliefs of A ('my idea') generate A's observable actions ('my behavior'). The actions of A, in turn, cause the (attended) sensory states of B. Attention directed toward agent A by agent B in turn enables the observations generated by A to entrain the hidden states of B ('your version of my idea'). This is just some hypothesis entertained by B about the causes of B's observations (i.e., about the mental states generating A's actions). To increase or maintain the reliability of B's hypothesis, B must then act on the niche ('your behavior') to test B's hypothesis about hidden causes, as it were (that is, to check for mutual understanding, for instance; e.g., Clark and Wilkes-Gibbs, 1986). B thereby causes A's attended observations and, hence, A's mental states ('my version of your idea'). This looping dynamics continues until both agents infer alignment (Friston and Frith, 2015a). Central here is that A is attending to the sensory states generated by B (and *vice versa*) because the only way to gather evidence for the adaptive prior that mental states are aligned is to attend to the sensory effects of one's actions; and evidence for hypotheses about the sensory effects of one's actions can *only* be given (in the present context) by the actions of the other agent. Working backwards, because the actions of another agent are generated by their mental states; and their mental states are entrained by (attended) sensory observations; and their sensory observations are generated by one's own actions; we thus arrive at the claim, given at the start of this section, that "A central part of the content of the prior belief prescribing the alignment of mental states among conspecifics is that the actions of agents (e.g., oneself) modulate the mental states (prior beliefs) of other agents." Adapted with permission from Friston and Frith (2015a).

Above, we discussed how mutual expectations of cooperativeness play a crucial role in getting cooperative communication off the ground in ontogeny. This just means that the (epigenetically and neurodevelopmentally) constrained, precise beliefs about the similarity of others and oneself enable nascent individuals to engage in cooperative communication. In particular, such couplings are only possible because both agents possess reliable expectations that the other agent is sufficiently 'like me' (cf. Meltzoff, 2007): we share the same prior beliefs to attune hidden dynamics. This provides an initial 'naive' confidence in beliefs about how one's action will influence another's prior beliefs (that, in turn, influence sensory outcomes via their actions). Borrowing from the language of social constructivist views of development (e.g., Rhodes and Wellman, 2017), our prior is a kind of naive certainty in one's intuitive theory about agential efficacy, with respect to the mental states of others (see also Kelso, 2016). This is to say that prior beliefs about the niche, e.g., others' mental states, bottom out just in their expected free energy. Belief-guided action (e.g., collaboration) may thus be constrained by salient policies entailed by a prior belief that, psychologically speaking, some hypothesis is in common ground. Put simply, to the extent that this hypothesis is sufficiently reliable, it will guide action and inference (see **Figure 3** and, e.g., Yoshida et al., 2008; Gallagher and Allen, 2018).

Pursuing this line of reasoning further provides a single, formally specified framework to subsume distinct proximate motivations for communication. That is, proximate motivations for communication (e.g., declarative, expressive, informative, interrogative motives; Begus and Southgate, 2012; Tomasello, 2019) surface as particular psychological manifestations of the same, species-typical tendency to align prior beliefs. Consider two proximate motivations for communication noted above; namely, a 'declarative' one motivated by the desired alignment of attentional states; and an 'interrogative' one motivated by a desire to learn about the niche. In the former case, individuals exploit their reliably shared beliefs to render the niche sufficiently similar to themselves (e.g., 'By ostensively pointing for that other agent, I expect to effectively align our mental states with respect to my intended referent'); and in the latter, individuals explore the precise, reliable parts of the niche (here, other agents) to improve their internal model of the niche (e.g., 'What is this thing called?'; reviewed in Harris and Corriveau, 2011; Harris et al., 2017). The underlying commonality in both cases is that individuals are effectively generating action-perception cycles that couple them to others, with the result being the alignment of mental states with respect to the niche.

Moving now to relevance optimization, we remind the reader that this process involves finessing the trade-off between the accuracy (e.g., meaningfulness, expressivity) and complexity (e.g., minimum description length, hierarchical depth of the policy) of their communicative constructions. Under active inference (see Pezzulo et al., 2013), if the prior beliefs of two individuals are inferred to be highly divergent on the basis of the evidence each provides to the other, and if both expect to minimize this divergence to a sufficient degree, then costlier (e.g., hierarchically deeper or more complex) policies should become relatively more salient as agents become increasingly dissimilar, as these policies will be necessary to resolve uncertainty or disambiguate the mental state of inscrutable others. This is in contrast to two individuals who 'speak the same language'. Here, less information needs to flow within the coupled action-perception cycle to attune mental states to a similar degree. In support of this view, one study (Kanwal et al., 2017) found that adults optimize the relevance of their communicative constructions during collaborative tasks as a function of their common ground, by using shorter words for common objects and longer words for uncommon objects (see also Winters et al., 2018). Related work suggests that children's adjective use (Bannard et al., 2017), turn-taking dynamics (Butko and Movellan, 2010), and question asking (Nelson et al., 2014) may be usefully cast as if they were optimizing the information content of produced communicative constructions with respect to processing and energy concerns (cf. Pea, 1979).

For a receiver, attention to the communicative stream enables updates to one's beliefs by providing 'contextual effects' (Sperber and Wilson, 1987); that is, orienting to a speaker influences the precision of hypotheses (about, e.g., the

interpretation of an utterance) through appropriate selection of ascending sensory information (indexed neurophysiologically by alpha suppression; Höhl et al., 2014; and increased theta; Begus et al., 2016; Köster et al., 2019). Specifically, individuals appear to explain away incoming sensory data by zeroing in on informative (useful) but parsimonious (i.e., efficient) explanations of hidden causes (Goodman and Frank, 2016; see also Gershman et al., 2015). For instance, Frank and Goodman (2014) report that adult and child listeners disambiguate ambiguous word meanings by optimizing their inferences of the relevance of a speaker's intended meaning[8]. In particular, these inferences can be captured as if individuals were maximizing model evidence for the prior belief that speakers are informative (see also Kao et al., 2014). This is captured by our extended formulation of cooperative communication, where inferences about mental states can be cast in terms of maximizing Bayesian model evidence (i.e., minimizing variational free energy) for the causes of one's sensation (e.g., another's mental states; Friston and Frith, 2015a).

Given an adaptive prior for alignment, one should tend to favor policies expected to reliably generate evidence of engagement in a coupled action-perception cycle. That is, such *ostensive policies* – policies expected to generate ostensive cues – are adaptive because they tend to generate sensory evidence for the hypothesis that one is engaged in a coupled action-perception cycle. Ostensive policies indicate to one's communicative partner that attending to one's action (i.e., to the individual generating ostensive cues) will likely be informative for them. Consequently, for a recipient, evidence provided by such cues increases the salience of certain policies; e.g., attentional orienting geared toward disambiguating the speaker's prior beliefs (Szufnarowska et al., 2014). As attention optimizes the precision of sensory cues, ostension in the coupled action-perception cycle plays a crucial (if indirect) role in reliably entraining and shaping prior beliefs (Axelsson et al., 2012; Butler and Tomasello, 2016; Kovács et al., 2017). Since prior beliefs generate action, ostensive cues are thus critical for guiding other individuals' actions and hence one's (attended) sensory states (e.g., Siposova et al., 2018).

By the same logic, in response to ostensive cues a recipient should (ostensively) signal their own inferred entrance into a communicative coupling (e.g., uptake signals; Austin, 1962); as well as, for example, their subjective degree of (and certainty in) the attunement of mental states (e.g., backchannel signals; Clark and Brennan, 1991). Indeed, other individuals – inferred to possess the same adaptive prior for alignment – preferentially leverage cooperative communication in turn; that is, respond to one's communicative bids (Kishimoto et al., 2007; Wu and Gros-Louis, 2015). This makes sense in light of the adaptive prior specified here: responding to another's communicative bids is something in the interest of both agents[9].

In summary, this subsection provided an active inference account of the microscale features of cooperative communication, from an individual's perspective, noted in the second section. We have thus outlined some important means by which individuals intentionally align their prior beliefs with respect to the dynamics of the niche (Constant et al., 2018), including others' mental states (Friston and Frith, 2015a). Indeed, a foundational facet of our account is that the alignment of the mental states of conspecifics manifests in the emergence of a novel scale of social and cultural dynamics constituted by synchronized component individuals (Ramstead et al., 2018). We turn to this now.

Dynamics at the Timescale of Mechanism: The Dyad

The precision of one's prior beliefs relative to another agent's, with whom one is coupled, has important implications for the degree and direction of attunement within and across couplings. In particular, the relative precision of the prior beliefs of each agent constrains the characteristic pattern of information flow between them – both at the level of turn taking in dialogical exchanges, and at the level of learning useful generative models of others[10] (and implicitly, of the self) (Friston and Frith, 2015a,b; Gencaga et al., 2015; e.g., Schippers et al., 2010). In terms of learning, this means that individuals endowed with relatively imprecise prior beliefs tend more, on average across time, to modify their own structure to fit that of their communicative partner(s), relative to individuals with relatively precise priors. This is a special case of generalized synchronization that is underwritten by the enslaving principle from cybernetics (Tschacher and Haken, 2007). To attune prior beliefs in such 'asymmetric' couplings, individuals with imprecise expectations in effect increase the precision of their sensory states (i.e., 'up the gain' afforded to sensory input; Moran et al., 2013; Auksztulewicz et al., 2017). This allows them to better change their own prior beliefs as a function of the evidence generated by their own (and others') action. This captures, for instance, the characteristic flow of information between agents following exposure to cues of prestige, with prestigious individuals being 'trend-setters' and others following suit (Henrich and Gil-White, 2001; Veissière et al., forthcoming).

[8]Interestingly, in machine learning, *automatic relevance determination* is a term used to denote model selection based upon variational free energy; namely, the removal of redundant model parameters to maximize efficiency or Bayesian model evidence. In turn, this is closely related to principles of minimum redundancy and maximum efficiency in perception (Barlow, 1974; Linsker, 1990; Wipf and Rao, 2007).

[9]It is useful to note that ostensive policies are salient insofar as they are (but one) intentional means for rapidly increasing the precision of (certain kinds) of hypotheses for another agent; e.g., that it is likely worthwhile to attend to the individual generating the ostensive cues. The account on offer therefore accommodates evidence suggesting that non-ostensive (unintentional) but nonetheless attention-grabbing actions, like shivering, may have similar effects on others' attentional orienting as ostensive cues (de Bordes et al., 2013; Szufnarowska et al., 2014).

[10]In numerical analyses of coupled communication, *turn taking* is usually implemented by a reciprocal augmentation and attenuation of sensory precision – so that one member of the dyad is listening while the other is speaking. Please see Friston and Frith (2015a) more details. A more enduring asymmetry relates to how one can learn from others, as illustrated using simulations of birdsong in Friston and Frith (2015b).

Additionally, such an asymmetry in information flow may capture the dynamics of the coupled action-perception cycles characteristic of interactions between human infants and children, and adults. Experimental and computational evidence suggests that older individuals possess relatively precise prior expectations, relative to those of younger, less experienced individuals (Wolpe et al., 2016; Karmali et al., 2018). Thus, younger individuals may ascribe greater precision to sensory information (Moran et al., 2014). The hypothesis here is, then, that repeated couplings between infants and children with adults (and more experienced peers) may cause the prior beliefs of inexperienced individuals to converge more toward the hidden causes generating sensory consequences (i.e., the mental states of more experienced others), rather than the other way around (Friston and Frith, 2015b; e.g., Fotopoulou and Tsakiris, 2017). That is, coupled action-perception cycles in such dyads tend to be characterized by an *asymmetric entrainment of prior beliefs* (for a closely related view, see Brownell, 2011).

What does this mean for the dynamics of (neural) belief updating during interaction? Technically, attunement to the niche instantiates the generalized synchronization of the statistics of prior beliefs and the niche (e.g., others' mental states); such that the structure and dynamics of individual brains come to recapitulate the structure and dynamics of the niche in which they are embedded[11] (Friston, 2012). This is depicted in **Figure 4**. Synchronization is a phenomenon that occurs in coupled chaotic dynamical systems (Pecora et al., 1997). Technically, it means that there is a (diffeomorphic) function relating the dynamics of the state of one system to those of the system with which it is coupled (Pecora et al., 1995). For instance, modeling results suggest that endowing two coupled hierarchical dynamical systems with an expectation to infer the hidden causes generating another's actions enables a bidirectional flow of information that synchronizes the statistics of their prior beliefs (Friston and Frith, 2015a; Constant et al., 2018). Alignment within and across coupled action-perception cycles means that the similarity (technically, the mutual information) of individuals' expectations increases (Friston and Frith, 2015b; Hasson and Frith, 2016). In this scheme, attention functions as a kind of coupling parameter, and its allocation is constrained by adaptive priors. Attention effectively increases the amount of information transferred from the system with precise priors to the system with imprecise priors (i.e., the system increasing the gain of its sensory states).

Indeed, studies in 'two-person' or hyperscanning neuroscience (Schilbach et al., 2013) have found evidence of the synchronizing effects of the usage of cooperative communication during, e.g., unidirectional person-to-person monologs (Stephens et al., 2010; Liu et al., 2017; Pérez et al., 2017), person-to-group monologs (Schmälzle et al., 2015), bidirectional person-to-person dialogues (Jiang et al., 2012), and even between classmates and their teacher during daily school activities (Dikker et al., 2017). Crucially, the degree of interbrain synchrony of neural dynamics appears to strongly predict psychological phenomena; for instance, the subjective meaningfulness of communication (Stolk et al., 2014), the accuracy of recall of the content of communication (Zadbood et al., 2017), and the perceived 'power' of political speech (Schmälzle et al., 2015; reviewed in Hasson et al., 2012; Feldman, 2015; Hasson and Frith, 2016; Schoot et al., 2016; Stolk et al., 2016). Indeed, the quality and amount of action-perception couplings over the course of early development better predicts later language ability (Hirsh-Pasek et al., 2015) and language-related brain function (Romeo et al., 2018) than more traditional measures, such as the number of words heard (Lindsay et al., 2019). Similarly, synchronous interbrain (limbic) dynamics in early infancy (i.e., prior to the onset of cooperative pointing) appears to be concomitant with several kinds of positive social experience, such as closeness and social bonding[12] (Atzil et al., 2014, 2017; reviewed in Feldman, 2015, 2017).

Dynamics at the Timescale of Ontogeny

The dynamics sketched above suggests a kind of Vygotskian scaffolding (Vygotsky, 1978; Moll and Tomasello, 2007) or 'co-construction' (Tomasello, 2019) of the dynamics of internal states; whereby – via recurrent engagement in loops of coupled action-perception with relatively 'entrenched' aspects of the niche – individuals learn (internalize) the salience of culturally anticipated policies used to infer hidden states. That is, by acting in a shared environment that contains older, relatively inflexible individuals that perform stereotyped behavior (characteristic of 'how we do things here'), younger individuals are able to learn the deontic value of policies (Ramstead et al., 2016; Veissière et al., forthcoming). For our purposes, this means that individuals' prior beliefs become more similar across couplings through (bidirectional) processes of (asymmetric) enculturation[13] (Renzi et al., 2017). That is, recurrent episodes of acutely increased alignment – of the kind typical of coupled action-perception cycles – are necessary for the creation and maintenance of species-typical states. In short, to gather evidence for an adaptive prior that mental states are aligned, one must act to bring about sensory states that are indicative of this belief (Chiel and Beer, 1997; Byrge et al., 2014).

[11]Technically, hidden states are characterized by their sufficient statistics. This denotes the minimum quantities needed to fully describe a probability distribution (Cover and Thomas, 1991). For the Gaussian distributions used in active inference, these are the mean and variance (see Buckley et al., 2017). Generalized synchronization implies that the mutual information of (the dynamics of) the states occupied by two (e.g.) chaotic dynamical systems is high (Pecora et al., 1995).

[12]To be clear, our claim is *not* that *only* the usage of communicative constructions can give rise to interbrain synchrony (see, e.g., evidence of non-verbal interbrain synchrony and associations with feelings of interpersonal closeness, Kinreich et al., 2017). Communicative constructions are merely a (highly useful) means to gather reliable evidence for the adaptive prior specified in the main text.

[13]For expository purposes we leave undiscussed the ontogenesis of critically important and later-appearing interactions with peers (e.g., Brownell and Carriger, 1990; Ashley and Tomasello, 1998; Brownell et al., 2006; see Brownell, 2016). Future explorations leveraging this approach should look to integrate data relating to peer-peer interactions in ontogenesis (reviewed in Brownell, 2011). Indeed, such phenomena are of great interest to the present account given the (possibly) more complex dynamics exhibited by the attunement of two systems embodying relatively imprecise expectations about how best to minimize uncertainty (e.g., Eckerman et al., 1989).

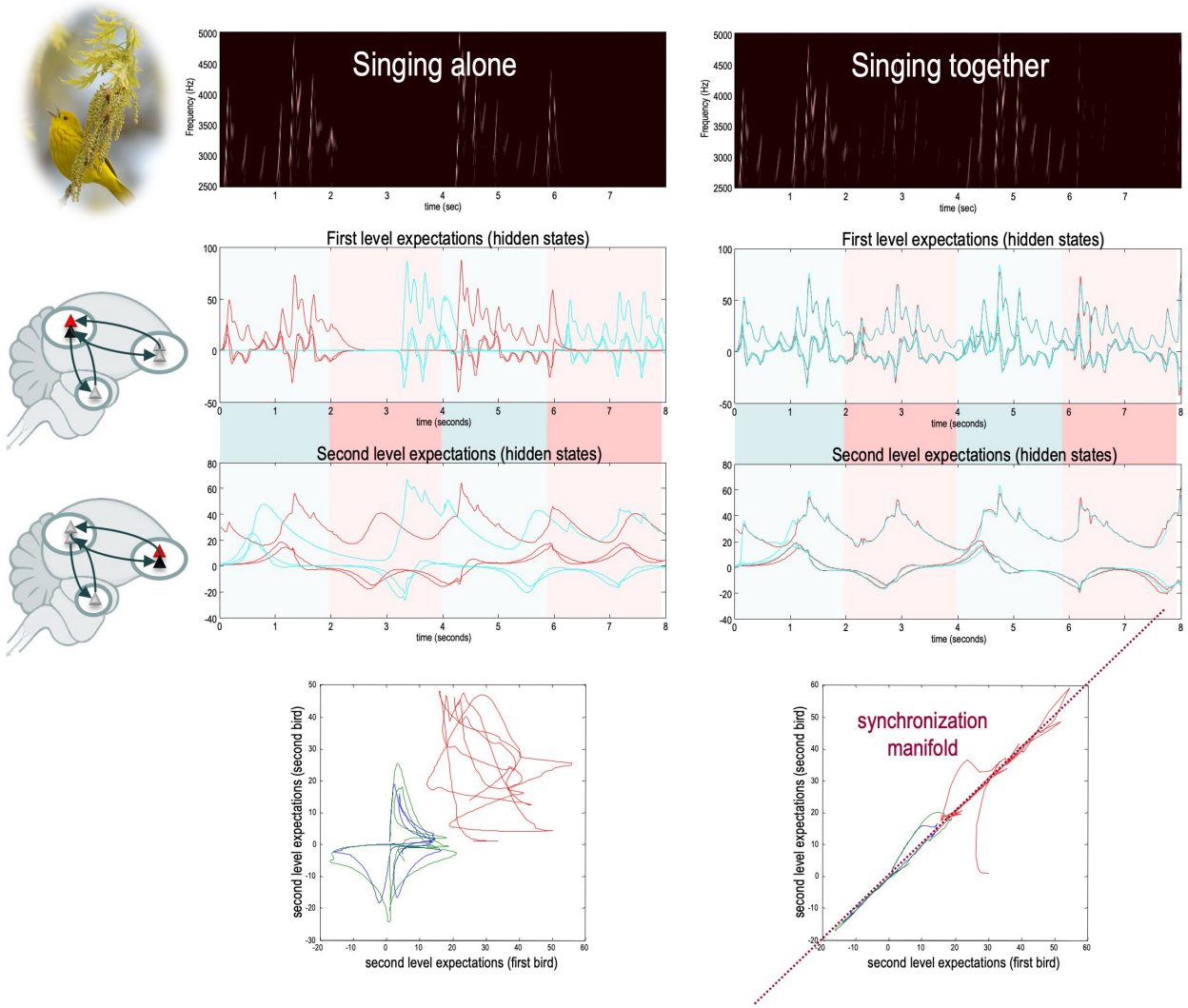

FIGURE 4 | A simulation of free-energy minimization of the sort implied by the coupled action-perception cycle. Two birds – endowed with prior expectations about the hidden states generating a shared (birdsong) narrative – sing for 2 s and then listen for a response. The posterior expectations for the first bird are shown in red; and the equivalent expectations for the second bird are shown in blue (both as a function of time). The left panel shows chaotic and uncoupled dynamics when the birds cannot hear each other ('singing alone'), while the right panel shows the synchrony in hidden states that emerges when the birds exchange sensory signals ('singing together'). The different colors correspond to the three hidden states for each bird. When singing alone, the birds cannot hear each other (because they are too far apart). Consequently, the dynamics diverge due to the sensitivity to initial conditions implicit in their (chaotic) generative models. The sonogram heard by the first bird is given in the **upper panel**. Because this bird can only hear itself, the sonogram reflects the predictions about action based upon its (first- and second-level) posterior expectations. Compare this to the case when the two birds can hear each other ('singing together'). Here, the posterior expectations encoded by internal states show (identical) synchrony at both the sensory and extrasensory levels, as shown in the **middle panels** (e.g., Pérez et al., 2017). Note that the sonogram is now continuous over the successive 2 s epochs, because the first bird can hear itself and the second bird. The ensuing synchronization manifold (i.e., the part of the joint state space that contains the generalized synchronization) is shown in the **lower panels**. These plot the second-level expectations in the second bird against the equivalent expectations in the first. The synchronization manifold for identical synchronization corresponds to the (broken) diagonal line. For details, see Friston and Frith (2015a). Obtained and adapted with permission from Constant et al. (2018).

Within and across interactions, such a dynamics increases the adaptive value of, e.g., collaborative foraging strategies by increasing inferred reliability in the hidden states generating observations (others' intentions; Han et al., 2015; Nakamura and Ohtsuki, 2016). This is because gathering evidence for the prior beliefs of other agents entails predicting how their beliefs relate to the niche; i.e., how others' beliefs relate to one's own mental states as well as non-social affordances. Consequently, gathering reliable evidence for others' mental states entails redirecting attention triadically (jointly). In this way, individuals become more reliable models of their interlocutor(s), and hence may leverage their own expectations about others' actions to guide expectations over sensory outcomes, like couplings with environmental

affordances[14] (e.g., Pezzulo, 2011; Gallotti and Frith, 2013; Bach et al., 2014).

A useful way to increase the degree of alignment of prior beliefs among individuals is to send more information to one's communicative partner. Holding the inferred common ground constant, one of the main ways to convey more information is to allow for hierarchically deeper policies (e.g., sequences of sequences) to generate action; that is, roughly, to provide more form (i.e., use longer communicative constructions). In effect, more information about mental states is thereby made observable. This perspective sheds interesting light on the species-typical trajectory from triadic attention (Striano and Stahl, 2005) to more reliably enacted forms of joint attention underwritten by reciprocal information flow – and the usage of pointing and gesture (Tomasello et al., 2007; Carpenter and Liebal, 2011) – to more complex constructions leveraged to transact with the hidden mental states of others (Aureli and Presaghi, 2010; see Colonnesi et al., 2010). The human agent appears to build up, nuance, and consolidate its (mutually expected) repertoire of action policies that, based on experience, have proven useful for adequately attuning with the mental states of conspecifics. That is, through this kind of continuous growth and hierarchical differentiation in communicative action policies (Goldin-Meadow, 2007; Tomasello, 2008), human individuals appear as though they were learning to tune themselves to the niche, and the niche to themselves.

Speaking generally, by repeatedly engaging in coupled action-perception cycles, individuals distil and abstract deeper observation-policy mappings (i.e., constructions) from the bottom up; that is, on an item-by-item basis (reviewed in Tomasello, 2000). In certain cases, individuals may then leverage learned hypotheses (about how best to disambiguate mental states) to reliably constrain the hypothesis space for learning and inference about constructions[15] (McClelland et al., 2010; see also Tenenbaum et al., 2011). That is, induction at higher layers of the model can serve to bootstrap learning at lower layers. Such 'domain-general' learning processes are illustrated by the model of Perfors et al. (2011). These authors provide a proof of principle account showing that several hours of child-directed input is sufficient for the posterior expectations of a hierarchical approximate Bayesian (i.e., active inference) learner – leveraging domain-general learning mechanisms – to converge toward a single, high level hypothesis about the causes of sensory input (here, a set of context-free grammars). That is, this set of context-free grammars had the greatest probability at the end of training. Consequently, this empirical prior functioned as abstract knowledge – it constrained expectations about likely hypotheses (in particular, auxiliary fronting) at lower layers[16] (see also Kemp et al., 2007).

Indeed, modeling schemes employing active inference provide evidence of their utility for modeling attunement to a communicative system (e.g., Kiebel et al., 2008, 2009; Friston et al., 2017b). For instance, Yildiz et al. (2013) used the active inference formalism to model word learning under optimal and noisy conditions and under variations in speaker accent. By attending to incoming input (i.e., increasing the precision of sensory signals), their model tuned its top–down beliefs to the structure of training data, which comprised sequences (of sequences) of spoken phonemes. The authors report that this model outperformed other computational learning schemes across a range of conditions and could be used to explain the judgments of adult second language learners. Future modeling work should investigate how an adaptive prior for alignment covers more ecologically valid instances of attunement to communicative constructions, such as the effects that 'starting small' and a prolonged period of developmental immaturity have on attuning to a communicative system (Elman, 1993; Bjorklund, 1997).

As noted, alignment with communicative partners means learning a set of 'automatic,' experientially robust (deontic) observation-policy mappings; e.g., the expectation (for English speakers) that a determiner typically precedes a noun (Meylan et al., 2017). Indeed, this view fits nicely with usage-based approaches to language acquisition (Tomasello, 2003; Lieven, 2016). Proponents of this view suggest that "constructions of all types are automatized motor routines and subroutines" that "come out of language use in context and... cognitive skills and strategies used in non-linguistic tasks" (Bybee, 2003, both p. 158).

Indeed, much of the structure and dynamics of the neural regions that underwrite the learning and usage of cooperative communication have been exapted (in particular, 'cooperativized') from their earlier evolutionary functions (Anderson, 2010; Bjorklund and Ellis, 2014). This has been emphasized, for instance, by embodied neurosemantics models of the neural underpinnings of the acquisition and comprehension of meaning in form-meaning pairings (reviewed in Pulvermüller, 2013, 2015). In such approaches, the meanings of both concrete and abstract constructions (e.g., 'kick' and 'love,' respectively) are grounded in low level sensorimotor dynamics and action-perception circuits contextualized by top–down input (Moseley et al., 2012; also, Harnad, 1990).

[14]From this perspective, it may be interesting for modeling work to investigate the notion of joint attention as an emergent property of coupling two hierarchical generative models attempting to infer the hidden states of each other (e.g., Friston and Frith, 2015a) while embedded in a broader ecological niche (e.g., Williams and Yaeger, 2017). In such a context, does joint attention effectively function to minimize a sensory Lagrangian over (jointly anticipated) sensory states (Sengupta et al., 2016)? What role does cooperative communication play in maintaining such a gauge invariance over the action of shared sensory states?

[15]Relatedly, because higher, contextualizing layers of a hierarchical model sample a larger space of inputs in estimating a smaller number of (more abstract) hypotheses (Tenenbaum et al., 2011), agents may, in certain instances, learn contextualizing 'overhypotheses' faster than learning at lower layers of the model (Gershman, 2017).

[16]We urge the reader to take the model of Perfors et al. (2011) with some caution. This is because part of the specification of their model was a set of (sets of) grammars; that is, their model came 'pre-equipped' with knowledge of various (formal, arbitrary) grammatical 'principles.' Thus, their model had simply to converge on the most probable set of grammars (hypothesis) given in its 'innate' repertoire. However, there is (i) no clear evidence for such innately specified (i.e., formal and arbitrary) linguistic principles in humans (Dąbrowska, 2015); (ii) no clear formulation of what may be included in such an innate repertoire (Tomasello, 2004); and (iii) numerous logical problems with the evolution of such innate structure (Christiansen and Chater, 2008; cf. e.g., Berwick et al., 2013).

To elaborate, human brains effectively combine two kinds – or two hierarchical levels – of general-purpose learning architecture to capitalize on the epistemic opportunities afforded by the action-perception cycle: (i) self-supervised (approximate Bayesian) learning, via the dynamic, hierarchical interplay between descending, neuronally encoded predictions and ascending prediction errors over time (Badcock et al., 2019b); and (ii) supervised (social) learning in a cultural niche via repeated, immersive practice in a set of culturally patterned routines (Roepstorff et al., 2010; Ramstead et al., 2016). In effect, attuning to a system of communicative constructions requires learning how to process and use form-meaning pairings in real-time communication. Thus, we are in agreement with Christiansen and Chater (2016a), who note that "learning [a cooperative communication system] involves creating a predictive model of the language, using online error-driven learning" (p. 121).

Deploying these learning processes in species typical communicative couplings means that, on average over time, individuals' communicative action policies become sufficiently similar; that is, not identical, but usable (Tomasello, 2003; Kidd et al., 2018). This is depicted in **Figure 5**. For instance, Bannard et al. (2009) found that the perplexity (an information theoretic measure that quantifies the fit of a distribution to a set of observations) of the (probabilistic) context-free grammar used to capture one child's (Brian's) utterances at age 2;0 (year;month) was able to account for approximately 15% of the utterances of another child (Annie, also 2;0). Similarly, the grammar imputed to Annie at 2;0 was able to explain approximately 36% of the utterances for Brian (2;0). Interestingly, at 3;0 model fit in either direction was increased. Thus, the grammar imputed to Brian at 3;0 accounted for roughly 59% of Annie's utterances (3;0), while the grammar imputed to Annie at 3;0 accounted for about 63% of Brian's utterances (3;0). Though the authors did not compute the significance of this change, this trend is precisely what one would expect under our model; namely, a trend toward statistically similar prior beliefs over hidden causes as individuals converge toward their cultural attractor (**Figure 5**).

In sum, by repeatedly 'filtering' one's action through others' mental states, one obtains a useful set of policies for flexibly and economically disambiguating prior beliefs. These correspond to policies with a high deontic value (Constant et al., 2019). In this way, one's set of constructions appears to converge on the set of constructions that constitute the communicative system(s) that predominantly generate one's sensory samples (i.e., those used by one's speaker community). This is to say that the prior beliefs of individuals converge toward an exploitable degree of similarity. One thus instantiates a sufficiently reliable model of the processes that generate sensory observations. This ontogenetic tendency repeats itself, cyclically, across generations. This has critical implications for the historical development of least effort communicative systems, to which we now turn.

Dynamics at the Timescale of Cultural Evolution

According to the model of cooperative communication proposed here, communicative systems (i.e., the dynamics of sets of form-meaning pairings) should appear, on average across time, to minimize their variational free energy (Ramstead et al., 2018). But what, exactly, does this mean; and how might this claim be investigated empirically? As noted above, a pointing gesture does not, in general, allow an agent to infer hidden causes as efficiently or reliably as the usage of a more complex construction. Therefore, in attempting to minimize their free energy, communicative systems should evolve toward a balance between usability and learnability (simplicity) on the one hand, and on the other, increasingly arbitrary, hierarchically deeper (complex) action sequences. This means that communicative systems should appear to optimize an accuracy-complexity, or expressivity-compressibility, trade-off (Tamariz and Kirby, 2016). Consider, for instance, the "drift to the arbitrary" (proposed by Tomasello, 2008, p. 219). Here, the suggestion is that 'grainy' bodily gestures like pointing give way to increasingly expressive, 'finer grained' gestures like pantomime. In turn, relatively expressive gestures give way to even more expressive, 'finely grained' vocal gestures like abstract and arbitrary communicative constructions (for similar views, see Wilcox, 2004; Fay et al., 2013; Perniss and Vigliocco, 2014). Interestingly, this roughly recapitulates the general ontogenetic trajectory of cooperative communication described above. It is as though – at multiple, nested scales of analysis – the human agent were becoming increasingly adept at flexibly deploying an increasingly sophisticated set of actions to resolve its sensory ambiguity.

These ideas are supported by the finding that the dynamics of relevance optimization across recurrent interactions, and generations of speakers, manifests in constructions that increase in expressivity with respect to production and processing costs (Tamariz and Kirby, 2016; e.g., Fay et al., 2010; Kirby et al., 2015). This means that human communicative systems, after a sufficiently long period of evolution, tend to cluster in a kind of 'least effort' subregion of a design space (i.e., parameter space) of communicative systems[17] (i Cancho and Solé, 2003; Seoane and Solé, 2018; see Evans and Levinson, 2009; Dediu et al., 2013). Expressed otherwise, relative to earlier generations of users of a particular communicative system, individuals in subsequent generations may be advantaged with respect to the range of communicative constructions that can be used to disambiguate mental states (Angus and Newton, 2015; perhaps, e.g., by coming to distinguish among previously undistinguished actions; Senghas, 2003). That is, communicative constructions themselves evolve to 'fit,' or gather evidence for, the adaptive priors favored by evolution and the specific demands of the local ecological niche (Kirby et al., 2007; Christiansen and Chater, 2008; Perfors and Navarro, 2014). This means that, over historical time, processes of cumulative cultural evolution (e.g., Richerson and Boyd, 2005; Henrich, 2015) tend to increase the deontic value of constructions by increasing the expressivity while minimizing the complexity of using and learning such constructions (for similar viewpoints, see Cornish et al., 2009; Dingemanse et al., 2015; Kirby et al.,

[17]To be clear, we are by no means claiming that this set of parameters exists in the brains of speakers. We are talking here about a scheme for modeling the dynamics of the cultural evolution of human communicative systems.

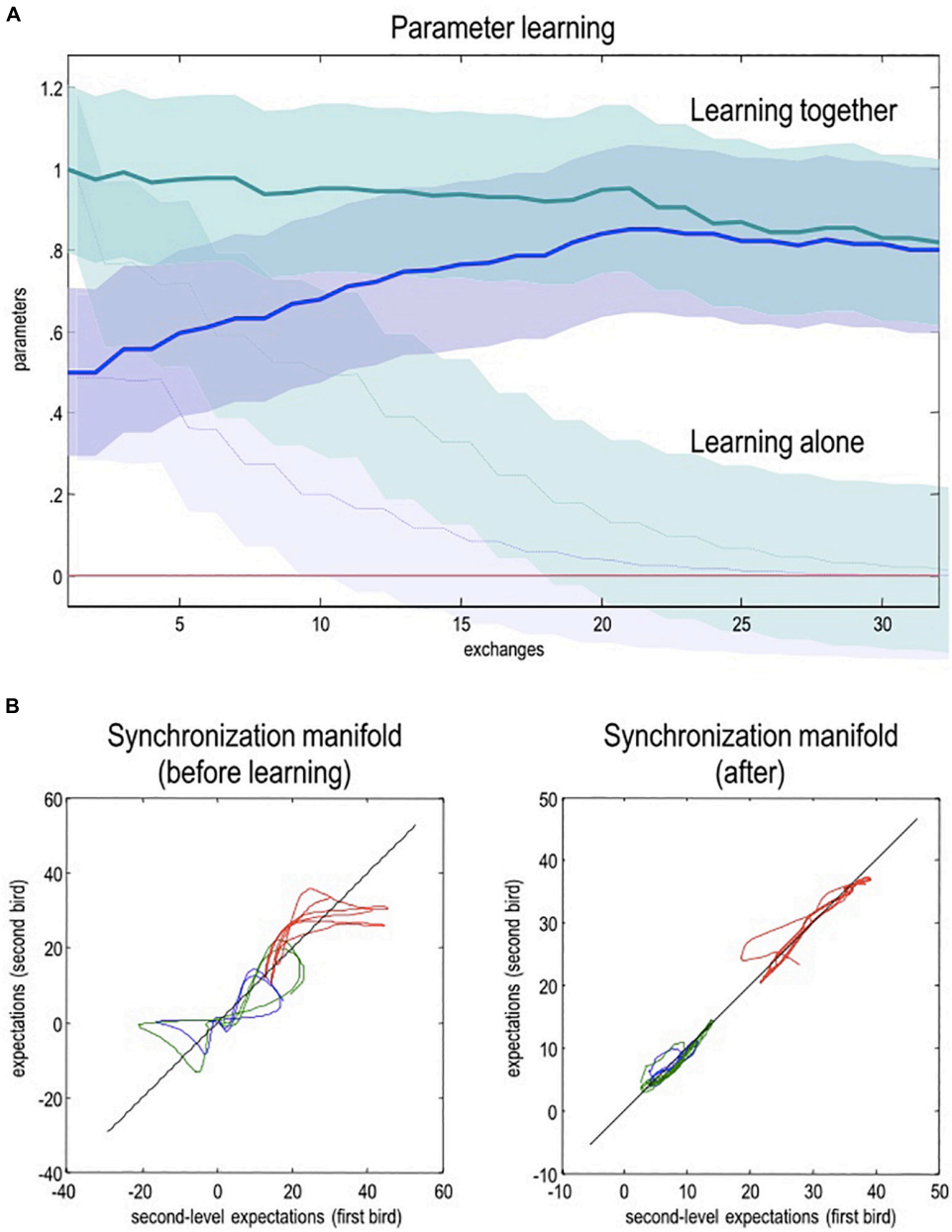

FIGURE 5 | A duet for one. This Figure depicts learning and communication via repeated engagement in coupled action-perception cycles in the context of an adaptive prior to align with conspecifics' hidden states. **(A)** Shows changes in the posterior expectations of an order parameter of the first bird (blue) and second bird (green) determining the chaotic structure of the songs depicted in **Figure 4** (by number of reciprocal sensory exchanges). The shaded areas correspond to 90% (prior Bayesian) confidence intervals. The broken lines (and intervals) report the results of the same simulation, but when the birds could not hear each other. **(B)** Shows the synchronization of posterior expectations encoded by extrasensory areas for the first (i) and subsequent (ii) exchanges, respectively. This synchronization is shown by plotting a mixture of expectations and their temporal derivatives from the second bird against the equivalent expectations of the first bird. This mixture is optimized by assuming a linear mapping between the birds' hidden states. In this example, the second (green) bird had more precise beliefs about its order parameter and, therefore, effectively, 'taught' the first bird. Parameter estimation (learning) converges toward the same value resulting in (generalized) synchrony between the two birds. For details, see Friston and Frith (2015b). Adapted with permission from Friston and Frith (2015b).

2015; Christiansen and Chater, 2016a; Tamariz and Kirby, 2016; Fay et al., 2018).

Cooperative communication emerges as a multiscale, self-organizing process that unfolds simultaneously across interaction, ontogeny, and cultural evolution (also, de Boer, 2011). Consequently, the adaptive prior under consideration enables, drives, and sustains each scale of dynamics. Circularly, each scale of dynamics generates actions that appear to gather evidence for the adaptive prior. Across developmental time, the contextualizing dynamics of cultural evolution appear

as a higher-order attractor – itself evolving in time, but sufficiently stable from the perspective of the developing individual – toward which individuals converge via recurrent engagement in coupled action-perception cycles that unfold in real-time. Taken together, interlocked dynamics at these three scales entrench the existence (i.e., probability) of the adaptive prior. In this way, cooperative communication becomes a self-fulfilling prophecy. That is, by gathering evidence for their adaptive priors, the low-level dynamics of interactants appear to create and maintain, at least for some period, the observable coherence of a contextualizing scale of (cultural) organization; namely, a communicative system (also see Szathmáry, 2015).

In active inference, the partitioning in the timescales that characterize a communicative system is formalized as between-scale differences in the precision of prior beliefs as one ascends scales (Constant et al., 2018; Ramstead et al., 2018). This is the result of, e.g., increasing the number of components (Smith et al., 2017) and the connectivity between components constituting a communicative system (Reali et al., 2018). This means that linear modifications to inputs to the system are associated with non-linear changes in its dynamics (Beckner et al., 2009; Shuai and Gong, 2014). Non-linearity is an inherent property of self-organizing systems (Prigogine and Stengers, 1984) and manifests in phenomena like critical slowing (i.e., phase transition; Gandhi et al., 1998; i Cancho and Solé, 2003), parameter reduction (Riley et al., 2011), and chaotic dynamics (Sanders et al., 2018). A change in the characteristic timescale of the dynamics of a cooperative communicative system is exemplified by Smith et al. (2017). These authors report experimental and simulation results suggesting that multi-person communicative systems exhibit slower regularization (decrease in conditional entropy) of a plurality marker across generations relative to communicative systems constituted by a single individual (for discussion of disparities of the pace of change across communicative systems, see Gray et al., 2013).

As noted above, the evolution of a communicative system may be cast as motion through a design space of communicative systems. Such spaces are effectively equivalent to the linguistic morphospace (Gray et al., 2013), or the space of states taken on by human communicative systems (e.g., linguistic networks; Seoane and Solé, 2018). Motion in design space may be relatively simple. For instance, Bybee (2010) has suggested that processes of grammaticalization – where flexible lexical forms gradually transition to fixed grammatical forms – may be modeled in terms of unidirectional (i.e., irreversible) motion through a continuous parameter space (also see Haspelmath, 1999). This might be modeled as a strange (Lorentz) attractor (Bybee, 2010), similar to that observed in models of communicative alignment (Friston and Frith, 2015a). In some cases, this motion may be more complex. For instance, the selection pressures acting on a system's constructions and, hence, the evolutionary trajectory of that set of constructions, varies as a function of the size of the population of speakers (Lupyan and Dale, 2010; Fay and Ellison, 2013; Reali et al., 2018; see Dingemanse et al., 2015).

In sum, the cultural niche construction implicit in free energy minimization in an ensemble of communicating conspecifics can be seen as a form of active inference on a (cultural) evolutionary level. In other words, selection pressures are just free energy gradients that allow us to cast *selection* (for useful communicative constructions) as a process of Bayesian *model selection* to maximize fitness; i.e., model evidence or the probability of communicative exchange, under a shared generative (phenotypic) model. This perspective nicely combines structure learning, evolution, and niche construction within the same formalism. For further discussion, please see Sella and Hirsh (2005), Frank (2012), Campbell (2016), and Constant et al. (2018).

FUTURE DIRECTIONS AND CONCLUSION

In this article, we have outlined an extension to existing theories of cooperative communication. Our extension is based on active inference and provides a novel, integrative take on the biobehavioral underpinnings of cooperative communication that complements existing psychological accounts (Tomasello, 2003, 2008, 2014, 2019). A more complete account of the dynamics entailed by the adaptive prior for alignment requires an integrative approach to research. To be sufficient, such research must aim to encapsulate the various timescales from which this prior emerges, particularly in a way that renders each scale of analysis complementary and mutually constraining with respect to the others (Tinbergen, 1963; Ramstead et al., 2018; Badcock et al., 2019b). The initial, though surely not exclusive, timescales of interest for cooperative communication were outlined in this paper. These range from the evolutionary history of early humans, to the intergenerational transmission of cultural patterns, down to individual development, and to two people conversing in real-time. This multiscale framework, arising from and underwriting the dynamics of the adaptive prior for alignment, should help to facilitate an understanding of inter- and intracultural similarities and differences in the structure and function of culture, mind, and brain. We conclude with a few comments about the limitations of the current proposal for the adaptive prior for alignment.

One limitation is the relative dearth of 'direct' evidence generated by empirical and computational studies of cooperative communication guided by the notion of an adaptive prior for alignment. We admit this is an important weakness, although one which can only be remedied through future research. Nevertheless, we have reviewed a substantial amount of indirect evidence generated by a range of empirical and simulation studies that speak to the integrative potential of the adaptive prior for alignment in making sense of cooperative communication.

For instance, our approach can be used to model the neuronal message-passing underwriting cooperative communication, as implied by active inference (e.g., Bastos et al., 2012; Parr and Friston, 2018). To illustrate this, regions in higher layers of cortex, such as anterior cingulate cortex (ACC; van den Heuvel and Sporns, 2013), integrate limbic afferents

encoding salience with control policies issued by motor cortex (Friston et al., 2014; Pezzulo et al., 2018). In turn, descending connections from paralimbic cortex convey signals that are unpacked as hierarchically nested sequences of cooperatively motivated action and inference, such as declarative pointing (Brunetti et al., 2014; see Holroyd and Yeung, 2012; Lavin et al., 2013; Haroush and Williams, 2015; Apps et al., 2016; Chambon et al., 2017). Interestingly, these neural considerations align with the psychological suggestions of Hare and Tomasello (2005), who suggest that early human selection pressures favored novel limbic dynamics that encode an increased tolerance and trust for conspecifics in the context of food (see the self-domestication hypothesis; Hare, 2017).

One specific, promising modeling approach pertains the usage of hierarchies of stable heteroclinic channels (SHCs; e.g., Rabinovich et al., 2014). SHCs are neuronally plausible models of hierarchically deep sequences (i.e., state trajectories of state trajectories) that may be scaled up to account for the acquisition and processing of more realistic cooperative communication data than have thus far been examined (Kiebel et al., 2009; see Rabinovich et al., 2015). In particular, it may be possible to use such a scheme to model the processing and use of communicative constructions, as these are hierarchically deep sequences of sequences (i.e., constructions are a statistically reliable ordering or 'chunk' of, e.g., word classes that entail chunks of morphemes that entail chunks of phonemes; relatedly, see 'chunk and pass' processing; Christiansen and Chater, 2016b). Indeed, the (re)use of hierarchical processing for language use may represent one instance of cooperativized, domain-general cognition exapted for usage in a cooperative social milieu. This is evidenced, for instance, by the presence of hierarchical processing of an artificial communication system in infants before 9 months of age (Kovács and Endress, 2014). Such a (developing) processing capacity may then be biased by the adaptive prior for alignment, after 9–12 months of age, toward disambiguating hierarchically organized communicative constructions (see Elman, 1993).

Another limitation of our proposal is that our consideration of the ontogenetic trajectory of cooperative communication focused exclusively on its typical trajectory. This was due to concerns about space. We readily acknowledge that there are all kinds of species atypical (i.e., unexpected) trajectories for the phenotypic expression of the adaptive prior for alignment. Arguably, studying how the dynamics of the adaptive prior for alignment may be perturbed in ontogenesis is crucial (e.g., discerning neurocomputational atypicalities or atypicalities in local niche dynamics; Thomas et al., 2019). Gaining a fuller grasp on the adaptive prior for alignment requires the integration of data and theory not just '*vertically*' (i.e., across scales), but also '*horizontally*' between the niche and its denizens. That is, the adaptive prior for alignment manifests distinctively not only across an array of timescales, but also at any given time across an array of cultural settings and, within cultures, neurotypical and neurodiverse populations.

For instance, in the fourth section, we discussed how, in neurotypical individuals, adequately explaining away sensory causes depends on a delicate, finely tuned balance of the top–down precision of hypotheses and the bottom–up precision of sensory fluctuations. But consider the case of autism, where neurocomputational atypicalities are thought to render the individual oversensitive to incoming error signals (Lawson et al., 2014; Mirza et al., 2019; see also Thomas et al., 2016). Such individuals would still expect to align mental states (Jaswal and Akhtar, 2019), and so would attend to others' communicative behaviors, but would be unable to attenuate the precision of sensory signals (Hadjikhani et al., 2017; Mirza et al., 2019).

Consequently, during initial interactive couplings, such individuals might initially look like they are typically developing (i.e., attending to others' eye gaze; Young et al., 2009). However, repeated attention to sources of sensory uncertainty (e.g., others' saccades), combined with an inability to adequately leverage predictions to explain away this uncertainty (owing to too much sensory precision), means that such individuals may develop idiosyncratic or atypical phenotypic expressions of the adaptive prior for alignment (e.g., avoiding eye gaze; Tanaka and Sung, 2016). In other words, early atypicalities in the internal dynamics generating evidence gathering cycles of action-perception may have downstream effects on joint attentional skills (Charman et al., 1997; Nyström et al., 2019), attunement to and use of communicative action policies (Loveland and Landry, 1986; Warlaumont et al., 2014), mental state inference (Tager-Flusberg, 2007), and other means for alignment (Heasman and Gillespie, 2019). In short, aberrant inference in a prosocial, developmental setting may easily lead to a pernicious kind of dyslexia – not for the written word – but for any communicative exchange (i.e., joint inference).

In summary, the adaptive prior for alignment 'sets the tone,' as it were, for species-typical patterns of evidence gathering, about oneself and the (social) world, that unfold over different timescales. The adaptive prior for alignment is, in effect, a kind of 'best guess' about the state occupied by the system at any point in time. Thus, for a human, processes of action, inference, learning, and (cultural) niche construction appear as if they were, on average across time, in the service of gathering evidence for the hypothesis that 'I' am like 'you,' that 'you' are like 'me,' and that 'we' exist.

AUTHOR CONTRIBUTIONS

JV and MR conceptualized the work for the manuscript. JV identified the manuscript's main target, i.e., cooperative communication. JV, PB, and MR worked out the argument and planned the manuscript. JV wrote the first draft of the manuscript. PB and AC helped with conceptualization work. AC drafted parts of the third section. PB helped edit the manuscript. KF and MR ensured that the technical aspects of the manuscript were presented accurately.

ACKNOWLEDGMENTS

We thank Michael J. Farrar, Robert D'Amico, Andreas Keil, Michael Tomasello, Leon Li, Josh Perlin, and Stephan Meylan

for helpful commentary, discussion, and feedback on early drafts of the manuscript. This research was produced thanks in part to funding from the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative (S. P. L. Veissière and MR), as well as a Postgraduate Research Scholarship in the Philosophy of Biomedicine as part of the ARC Australian Laureate Fellowship project A Philosophy of Medicine for the 21st Century (Ref: FL170100160) (AC), a Social Sciences and Humanities Research Council doctoral fellowship (Ref: 752-2019-0065) (AC), a Joseph-Armand Bombardier Canada Doctoral Scholarship and a Michael Smith Foreign Study Supplements award from the Social Sciences and Humanities Research Council of Canada (MR), and by a Wellcome Principal Research Fellowship (KF – Ref: 088130/Z/09/Z).

REFERENCES

Adams, R. A., Shipp, S., and Friston, K. J. (2013). Predictions not commands: active inference in the motor system. *Brain Struct. Funct.* 218, 611–643. doi: 10.1007/s00429-012-0475-5

Anderson, M. L. (2010). Neural reuse: a fundamental organizational principle of the brain. *Behav. Brain Sci.* 33, 245–266; discussion 266–313. doi: 10.1017/S0140525X10000853

Angus, S. D., and Newton, J. (2015). Emergence of shared intentionality is coupled to the advance of cumulative culture. *PLoS Comput. Biol.* 11:e1004587. doi: 10.1371/journal.pcbi.1004587

Apps, M. A. J., Rushworth, M. F. S., and Chang, S. W. C. (2016). The anterior cingulate gyrus and social cognition: tracking the motivation of others. *Neuron* 90, 692–707. doi: 10.1016/j.neuron.2016.04.018

Ashley, J., and Tomasello, M. (1998). Cooperative problem-solving and teaching in preschoolers. *Soc. Dev.* 7, 143–163. doi: 10.1111/1467-9507.00059

Attias, H. (2003). "Planning by probabilistic inference," in *Proceedings of the 9th International Workshop on Artificial Intelligence and Statistics*, Key West, FL.

Atzil, S., Hendler, T., and Feldman, R. (2014). The brain basis of social synchrony. *Soc. Cogn. Affect. Neurosci.* 9, 1193–1202. doi: 10.1093/scan/nst105

Atzil, S., Touroutoglou, A., Rudy, T., Salcedo, S., Feldman, R., Hooker, J. M., et al. (2017). Dopamine in the medial amygdala network mediates human bonding. *Proc. Natl. Acad. Sci. U.S.A.* 114, 2361–2366. doi: 10.1073/pnas.1612233114

Auksztulewicz, R., Barascud, N., Cooray, G., Nobre, A. C., Chait, M., and Friston, K. (2017). The cumulative effects of predictability on synaptic gain in the auditory processing stream. *J. Neurosci.* 37, 6751–6760. doi: 10.1523/JNEUROSCI.0291-17.2017

Aureli, T., and Presaghi, F. (2010). Developmental trajectories for mother–infant coregulation in the second year of life. *Infancy* 15, 557–585. doi: 10.1111/j.1532-7078.2010.00034.x

Austin, J. L. (1962). *How to Do Things with Words*. New York, NY: Oxford University Press.

Axelsson, E. L., Churchley, K., and Horst, J. S. (2012). The right thing at the right time: why ostensive naming facilitates word learning. *Front. Psychol.* 3:88. doi: 10.3389/fpsyg.2012.00088

Aylett, M., and Turk, A. (2004). The smooth signal redundancy hypothesis: a functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech* 47(Pt 1), 31–56. doi: 10.1177/00238309040470010201

Bach, P., Nicholson, T., and Hudson, M. (2014). The affordance-matching hypothesis: how objects guide action understanding and prediction. *Front. Hum. Neurosci.* 8:254. doi: 10.3389/fnhum.2014.00254

Badcock, P. B., Friston, K. J., and Ramstead, M. J. D. (2019a). The hierarchically mechanistic mind: a free-energy formulation of the human psyche. *Phys. Life Rev.* 31, 104–121. doi: 10.1016/j.plrev.2018.10.002

Badcock, P. B., Friston, K. J., Ramstead, M. J. D., Ploeger, A., and Hohwy, J. (2019b). The hierarchically mechanistic mind: an evolutionary systems theory of the human brain, cognition, and behavior. *Cogn. Affect. Behav. Neurosci.* 19, 1319–1351. doi: 10.3758/s13415-019-00721-3

Baker, C. L., and Tenenbaum, J. B. (2014). "Modeling human plan recognition using Bayesian theory of mind," in *Plan, Activity, and Intent Recognition*, eds G. Sukthankar, C. Geib, H. H. Bui, D. V. Pynadath, and R. P. Goldman (Boston, MA: Morgan Kaufmann Publishers), 177–204. doi: 10.1016/b978-0-12-398532-3.00007-5

Bannard, C., Lieven, E., and Tomasello, M. (2009). Modeling children's early grammatical knowledge. *Proc. Natl. Acad. Sci. U.S.A.* 106, 17284–17289. doi: 10.1073/pnas.0905638106

Bannard, C., Rosner, M., and Matthews, D. (2017). What's worth talking about? Information theory reveals how children balance informativeness and ease of production. *Psychol. Sci.* 28, 954–966. doi: 10.1177/0956797617699848

Barlow, H. B. (1974). Inductive inference, coding, perception, and language. *Perception* 3, 123–134. doi: 10.1068/p030123

Bastos, A. M., Usrey, W. M., Adams, R. A., Mangun, G. R., Fries, P., and Friston, K. J. (2012). Canonical microcircuits for predictive coding. *Neuron* 76, 695–711. doi: 10.1016/j.neuron.2012.10.038

Baumard, N., André, J.-B., and Sperber, D. (2013). A mutualistic approach to morality: the evolution of fairness by partner choice. *Behav. Brain Sci.* 36, 59–78. doi: 10.1017/S0140525X11002202

Beckner, C., Blythe, R., Bybee, J., Christiansen, M. H., Croft, W., Ellis, N. C., et al. (2009). Language is a complex adaptive system: position paper. *Lang. Learn.* 59(Suppl. 1), 1–26. doi: 10.1111/j.1467-9922.2009.00533.x

Begus, K., Gliga, T., and Southgate, V. (2014). Infants learn what they want to learn: responding to infant pointing leads to superior learning. *PLoS One* 9:e108817. doi: 10.1371/journal.pone.0108817

Begus, K., Gliga, T., and Southgate, V. (2016). Infants' preferences for native speakers are associated with an expectation of information. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12397–12402. doi: 10.1073/pnas.1603261113

Begus, K., and Southgate, V. (2012). Infant pointing serves an interrogative function. *Dev. Sci.* 15, 611–617. doi: 10.1111/j.1467-7687.2012.01160.x

Behne, T., Carpenter, M., and Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Dev. Sci.* 8, 492–499. doi: 10.1111/j.1467-7687.2005.00440.x

Berwick, R. C., Friederici, A. D., Chomsky, N., and Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends Cogn. Sci.* 17, 89–98. doi: 10.1016/j.tics.2012.12.002

Bjorklund, D. F. (1997). The role of immaturity in human development. *Psychol. Bull.* 122, 153–169. doi: 10.1037/0033-2909.122.2.153

Bjorklund, D. F., and Ellis, B. J. (2014). Children, childhood, and development in evolutionary perspective. *Dev. Rev.* 34, 225–264. doi: 10.1016/j.dr.2014.05.005

Blei, D. M., Kucukelbir, A., and McAuliffe, J. D. (2017). Variational inference: a review for statisticians. *J. Am. Stat. Assoc.* 112, 859–877. doi: 10.1080/01621459.2017.1285773

Bogacz, R. (2017). A tutorial on the free-energy framework for modelling perception and learning. *J. Math. Psychol.* 76(Pt B), 198–211. doi: 10.1016/j.jmp.2015.11.003

Botvinick, M., and Toussaint, M. (2012). Planning as inference. *Trends Cogn. Sci.* 16, 485–488. doi: 10.1016/j.tics.2012.08.006

Brownell, C. A. (2011). Early developments in joint action. *Rev. Philos. Psychol.* 2, 193–211. doi: 10.1007/s13164-011-0056-1

Brownell, C. A., and Carriger, M. S. (1990). Changes in cooperation and self-other differentiation during the second year. *Child Dev.* 61, 1164–1174.

Brownell, C. A., Ramani, G. B., and Zerwas, S. (2006). Becoming a social partner with peers: cooperation and social understanding in one- and two-year-olds. *Child Dev.* 77, 803–821. doi: 10.1111/j.1467-8624.2006.00904.x

Brownell, C. A. (2016). The Early Social Development Research Lab. Prosocial behavior in infancy: the role of socialization. *Child Dev. Perspect.* 10, 222–227. doi: 10.1111/cdep.12189

Bruineberg, J., Kiverstein, J., and Rietveld, E. (2018a). The anticipating brain is not a scientist: the free-energy principle from an ecological-enactive perspective. *Synthese* 195, 2417–2444. doi: 10.1007/s11229-016-1239-1

Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L., and Friston, K. J. (2018b). Free-energy minimization in joint agent-environment systems: a niche construction perspective. *J. Theor. Biol.* 455, 161–178. doi: 10.1016/j.jtbi.2018.07.002

Brunetti, M., Zappasodi, F., Marzetti, L., Perrucci, M. G., Cirillo, S., Romani, G. L., et al. (2014). Do you know what i mean? Brain oscillations and the understanding of communicative intentions. *Front. Hum. Neurosci.* 8:36. doi: 10.3389/fnhum.2014.00036

Buckley, C. L., Kim, C. S., McGregor, S., and Seth, A. K. (2017). The free energy principle for action and perception: a mathematical review. *J. Math. Psychol.* 81, 55–79. doi: 10.1016/j.jmp.2017.09.004

Bullinger, A. F., Zimmermann, F., Kaminski, J., and Tomasello, M. (2011). Different social motives in the gestural communication of chimpanzees and human children. *Dev. Sci.* 14, 58–68. doi: 10.1111/j.1467-7687.2010.00952.x

Butko, N. J., and Movellan, J. R. (2010). Detecting contingencies: an infomax approach. *Neural Networks* 23, 973–984. doi: 10.1016/j.neunet.2010.09.001

Butler, L. P., and Tomasello, M. (2016). Two- and 3-year-olds integrate linguistic and pedagogical cues in guiding inductive generalization and exploration. *J. Exp. Child Psychol.* 145, 64–78. doi: 10.1016/j.jecp.2015.12.001

Bybee, J. (2003). "Cognitive processes in grammaticalization," in *The New Psychology of Language*, ed. M. Tomasello (Mahwah, NJ: Lawrence Erlbaum Associates), 159–182.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge: Cambridge University Press.

Byrge, L., Sporns, O., and Smith, L. B. (2014). Developmental process emerges from extended brain-body-behavior networks. *Trends Cogn. Sci.* 18, 395–403. doi: 10.1016/j.tics.2014.04.010

Call, J. (2009). Contrasting the social cognition of humans and nonhuman apes: the shared intentionality hypothesis. *Top. Cogn. Sci.* 1, 368–379. doi: 10.1111/j.1756-8765.2009.01025.x

Call, J., and Tomasello, M. (Eds.) (2007). *The Gestural Communication of Apes and Monkeys*. New York, NY: Lawrence Erlbaum Associates.

Callaghan, T., Moll, H., Rakoczy, H., Warneken, F., Liszkowski, U., Behne, T., et al. (2011). Early social cognition in three cultural contexts. *Monogr. Soc. Res. Child Dev.* 76, vii–viii, 1–142. doi: 10.1111/j.1540-5834.2011.00603.x

Campbell, J. O. (2016). Universal Darwinism as a process of Bayesian inference. *Front. Syst. Neurosci.* 10:49. doi: 10.3389/fnsys.2016.00049

Carpenter, M., and Call, J. (2013). "How joint is the joint attention of apes and human infants?," in *Agency and Joint Attention*, eds J. Metcalfe and H. S. Terrence (New York, NY: Oxford University Press), 49–61. doi: 10.1093/acprof:oso/9780199988341.003.0003

Carpenter, M., and Liebal, K. (2011). "Joint attention, communication, and knowing together in infancy," in *Joint Attention: New Developments in Psychology, Philosophy of Mind, and Social Neuroscience*, ed. S. Seemann (Cambridge, MA: MIT Press), 159–181.

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63, i–vi, 1–143.

Chambon, V., Domenech, P., Jacquet, P. O., Barbalat, G., Bouton, S., Pacherie, E., et al. (2017). Neural coding of prior expectations in hierarchical intention inference. *Sci. Rep.* 7:1278. doi: 10.1038/s41598-017-01414-y

Charman, T., Swettenham, J., Baron-Cohen, S., Cox, A., Baird, G., and Drew, A. (1997). Infants with autism: an investigation of empathy, pretend play, joint attention, and imitation. *Dev. Psychol.* 33, 781–789. doi: 10.1037//0012-1649.33.5.781

Chiel, H. J., and Beer, R. D. (1997). The brain has a body: adaptive behavior emerges from interactions of nervous system, body and environment. *Trends Neurosci.* 20, 553–557. doi: 10.1016/s0166-2236(97)01149-1

Christiansen, M. H., and Chater, N. (2008). Language as shaped by the brain. *Behav. Brain Sci.* 31, 489–508. doi: 10.1017/S0140525X08004998 ; discussion 509–558,

Christiansen, M. H., and Chater, N. (2016a). *Creating Language: Integrating Evolution, Acquisition, and Processing*. Cambridge, MA: MIT Press.

Christiansen, M. H., and Chater, N. (2016b). The now-or-never bottleneck: a fundamental constraint on language. *Behav. Brain Sci.* 39:e62. doi: 10.1017/S0140525X1500031X

Christiansen, M. H., and Kirby, S. (2003). Language evolution: consensus and controversies. *Trends Cogn. Sci.* 7, 300–307. doi: 10.1016/s1364-6613(03)00136-0

Chukoskie, L., Snider, J., Mozer, M. C., Krauzlis, R. J., and Sejnowski, T. J. (2013). Learning where to look for a hidden target. *Proc. Natl. Acad. Sci. U.S.A.* 110(Suppl 2), 10438–10445. doi: 10.1073/pnas.1301216110

Cisek, P. (2007). Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1585–1599. doi: 10.1098/rstb.2007.2054

Clark, A. (2006). Language, embodiment, and the cognitive niche. *Trends Cogn. Sci.* 10, 370–374. doi: 10.1016/j.tics.2006.06.012

Clark, A. (2008). *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. Available online at: https://market.android.com/details?id=book-n5wRDAAAQBAJ (accessed October 29, 2008).

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behav. Brain Sci.* 36, 181–204. doi: 10.1017/S0140525X12000477

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H., and Brennan, S. E. (1991). "Grounding in communication," in *Perspectives on Socially Shared Cognition*, eds L. B. Resnick, J. M. Levine, and S. D. Teasley (Washington, DC: American Psychological Association), 127–149.

Clark, H. H., and Wilkes-Gibbs, D. (1986). Referring as a collaborative process. *Cognition* 22, 1–39. doi: 10.1016/0010-0277(86)90010-7

Colonnesi, C., Stams, G. J. J. M., Koster, I., and Noom, M. J. (2010). The relation between pointing and language development: a meta-analysis. *Dev. Rev.* 30, 352–366. doi: 10.1016/j.dr.2010.10.001

Constant, A., Clark, A., Kirchhoff, M., and Friston, K. J. (forthcoming). Extended active inference: constructing predictive cognition beyond skulls. *Mind Lang.*

Constant, A., Ramstead, M., Veissière, S., and Friston, K. J. (2019). Regimes of expectations: an active inference model of social conformity and decision making. *Front. Psychol.* 10:679. doi: 10.3389/fpsyg.2019.00679

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., Campbell, J. O., and Friston, K. J. (2018). A variational approach to niche construction. *J. R. Soc. Interface* 15:20170685. doi: 10.1098/rsif.2017.0685

Cornish, H., Tamariz, M., and Kirby, S. (2009). Complex adaptive systems and the origins of adaptive structure: what experiments can tell us. *Lang. Learn.* 59, 187–205. doi: 10.1111/j.1467-9922.2009.00540.x

Cover, T. M., and Thomas, J. A. (1991). *Elements of Information Theory*. Hoboken, NJ: Wiley.

Csibra, G. (2010). Recognizing communicative intentions in infancy. *Mind Lang.* 25, 141–168. doi: 10.1111/j.1468-0017.2009.01384.x

Dąbrowska, E. (2015). What exactly is Universal grammar, and has anyone seen it? *Front. Psychol.* 6:852. doi: 10.3389/fpsyg.2015.00852

de Boer, B. (2011). "Self−organization and language evolution," in *Oxford Handbook of Language Evolution*, 4th Edn, eds K. Gibson and M. Tallerman (Oxford: Oxford University Press), 612–620. doi: 10.1093/oxfordhb/9780199541119.013.0063

de Bordes, P. F., Cox, R. F. A., Hasselman, F., and Cillessen, A. H. N. (2013). Toddlers' gaze following through attention modulation: intention is in the eye of the beholder. *J. Exp. Child Psychol.* 116, 443–452. doi: 10.1016/j.jecp.2012.09.008

Dediu, D., Cysouw, M., Levinson, S. C., Baronchelli, A., Christiansen, M. H., and Croft, W. (2013). "Cultural evolution of language," in *Cultural Evolution: Society, Technology, Language, and Religion. Strüngmann Forum Reports*, Vol. 12, eds M. H. Christiansen and P. Richerson (Cambridge, MA: MIT Press), 303–332.

Demirdjian, D., Taycher, L., Shakhnarovich, G., Grauman, K., and Darrell, T. (2005). "Avoiding the "streetlight effect": tracking by exploring likelihood modes," in *Proceedings of the 10th IEEE International Conference on Computer Vision*, Vol. 1–2, (Beijing: Institute of Electrical and Electronics Engineers), 357–364.

Devaine, M., Hollard, G., and Daunizeau, J. (2014). Theory of mind: did evolution fool us? *PLoS One* 9:e87619. doi: 10.1371/journal.pone.0087619

Dikker, S., Wan, L., Davidesco, I., Kaggen, L., Oostrik, M., McClintock, J., et al. (2017). Brain-to-Brain synchrony tracks real-world dynamic group interactions in the classroom. *Curr. Biol.* 27, 1375–1380. doi: 10.1016/j.cub.2017.04.002

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., and Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends Cogn. Sci.* 19, 603–615. doi: 10.1016/j.tics.2015.07.013

Duguid, S., Wyman, E., Bullinger, A. F., Herfurth-Majstorovic, K., and Tomasello, M. (2014). Coordination strategies of chimpanzees and human children in a

Stag Hunt game. *Proc. Biol. Sci. R. Soc. Lond. B Biol. Sci.* 281:20141973. doi: 10.1098/rspb.2014.1973

Eckerman, C. O., Davis, C. C., and Didow, S. M. (1989). Toddlers' emerging ways of achieving social coordinations with a peer. *Child Dev.* 60, 440–453.

Elman, J. L. (1993). Learning and development in neural networks: the importance of starting small. *Cognition* 48, 71–99. doi: 10.1016/0010-0277(93)90058-4

Evans, N., and Levinson, S. C. (2009). The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* 32, 429–448; discussion 448–494. doi: 10.1017/S0140525X0999094X

Fay, N., Arbib, M., and Garrod, S. (2013). How to bootstrap a human communication system. *Cogn. Sci.* 37, 1356–1367. doi: 10.1111/cogs.12048

Fay, N., and Ellison, T. M. (2013). The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PLoS One* 8:e71781. doi: 10.1371/journal.pone.0071781

Fay, N., Ellison, T. M., Tylén, K., Fusaroli, R., Walker, B., and Garrod, S. (2018). Applying the cultural ratchet to a social artefact: the cumulative cultural evolution of a language game. *Evol. Hum. Behav.* 39, 300–309. doi: 10.1016/j.evolhumbehav.2018.02.002

Fay, N., Garrod, S., Roberts, L., and Swoboda, N. (2010). The interactive evolution of human communication systems. *Cogn. Sci.* 34, 351–386. doi: 10.1111/j.1551-6709.2009.01090.x

Feldman, H., and Friston, K. J. (2010). Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4:215. doi: 10.3389/fnhum.2010.00215

Feldman, R. (2015). The adaptive human parental brain: implications for children's social development. *Trends Neurosci.* 38, 387–399. doi: 10.1016/j.tins.2015.04.004

Feldman, R. (2017). The neurobiology of human attachments. *Trends Cogn. Sci.* 21, 80–99. doi: 10.1016/j.tics.2016.11.007

Fillmore, C. J., Kay, P., and O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: the case of let alone. *Language* 64, 501–538. doi: 10.2307/414531

Fonagy, P., and Allison, E. (2014). The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy* 51, 372–380. doi: 10.1037/a0036505

Fotopoulou, A., and Tsakiris, M. (2017). Mentalizing homeostasis: the social origins of interoceptive inference. *Neuropsychoanalysis* 19, 3–28. doi: 10.1080/15294145.2017.1294031

Frank, M. C., and Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cogn. Psychol.* 75, 80–96. doi: 10.1016/j.cogpsych.2014.08.002

Frank, S. A. (2012). Natural selection. V. How to read the fundamental equations of evolutionary change in terms of information theory. *J. Evol. Biol.* 25, 2377–2396. doi: 10.1111/jeb.12010

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nat. Rev. Neurosci.* 11, 127–138. doi: 10.1038/nrn2787

Friston, K. (2011). "Embodied inference: or "I think therefore I am, if I am what I think."," in *The Implications of Embodiment (Cognition and Communication)*, eds W. Tschacher and C. Bergomi (Exeter: Imprint Academic), 89–125.

Friston, K. (2012). A free energy principle for biological systems. *Entropy* 14, 2100–2121. doi: 10.3390/e14112100

Friston, K. (2013). Life as we know it. *J. R. Soc. Interface* 10:20130475. doi: 10.1098/rsif.2013.0475

Friston, K., Adams, R. A., Perrinet, L., and Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Front. Psychol.* 3:151. doi: 10.3389/fpsyg.2012.00151

Friston, K., and Ao, P. (2012). Free energy, value, and attractors. *Comput. Math. Methods Med.* 2012:937860. doi: 10.1155/2012/937860

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., and Pezzulo, G. (2017a). Active inference: a process theory. *Neural Comput.* 29, 1–49. doi: 10.1162/NECO_a_00912

Friston, K., and Frith, C. (2015a). A Duet for one. *Conscious. Cogn.* 36, 390–405. doi: 10.1016/j.concog.2014.12.003

Friston, K., and Frith, C. D. (2015b). Active inference, communication and hermeneutics. *Cortex* 68, 129–143. doi: 10.1016/j.cortex.2015.03.025

Friston, K., Levin, M., Sengupta, B., and Pezzulo, G. (2015a). Knowing one's place: a free-energy approach to pattern regulation. *J. R. Soc. Interface* 12:20141383. doi: 10.1098/rsif.2014.1383

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., and Pezzulo, G. (2015b). Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214. doi: 10.1080/17588928.2015.1020053

Friston, K., Rosch, R., Parr, T., Price, C., and Bowman, H. (2017b). Deep temporal models and active inference. *Neurosci. Biobehav. Rev.* 77, 388–402. doi: 10.1016/j.neubiorev.2017.04.009

Friston, K., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2014). The anatomy of choice: dopamine and decision-making. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130481. doi: 10.1098/rstb.2013.0481

Gadamer, H.-G. (2004). *Truth and Method*. London: Continuum. Available at: mvlindsey.files.wordpress.com/2015/08/truth-and-method-gadamer-2004.pdf (accessed March 5, 2020).

Gallagher, S., and Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. *Synthese* 195, 2627–2648. doi: 10.1007/s11229-016-1269-8

Gallotti, M., and Frith, C. D. (2013). Social cognition in the we-mode. *Trends Cogn. Sci.* 17, 160–165. doi: 10.1016/j.tics.2013.02.002

Gandhi, A., Levin, S., and Orszag, S. (1998). "Critical slowing down" in time-to-extinction: an example of critical phenomena in ecology. *J. Theor. Biol.* 192, 363–376. doi: 10.1006/jtbi.1998.0660

Ganea, P. A., and Saylor, M. M. (2007). Infants' use of shared linguistic information to clarify ambiguous requests. *Child Dev.* 78, 493–502. doi: 10.1111/j.1467-8624.2007.01011.x

Gencaga, D., Knuth, K., and Rossow, W. (2015). A recipe for the estimation of information flow in a dynamical system. *Entropy* 17, 438–470. doi: 10.3390/e17010438

Gershman, S. J. (2017). On the blessing of abstraction. *Q. J. Exp. Psychol.* 70, 361–365. doi: 10.1080/17470218.2016.1159706

Gershman, S. J., Horvitz, E. J., and Tenenbaum, J. B. (2015). Computational rationality: a converging paradigm for intelligence in brains, minds, and machines. *Science* 349, 273–278. doi: 10.1126/science.aac6076

Goldberg, A. E. (2003). Constructions: a new theoretical approach to language. *Trends Cogn. Sci.* 7, 219–224. doi: 10.1016/S1364-6613(03)00080-9

Goldin-Meadow, S. (2007). Pointing sets the stage for learning language—and creating language. *Child Dev.* 78, 741–745. doi: 10.1111/j.1467-8624.2007.01029.x

Goodman, N. D., and Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends Cogn. Sci.* 20, 818–829. doi: 10.1016/j.tics.2016.08.005

Goodwin, C. (2000). Action and embodiment within situated human interaction. *J. Pragmat.* 32, 1489–1522. doi: 10.1016/s0378-2166(99)00096-x

Grau-Moya, J., Hez, E., Pezzulo, G., and Braun, D. A. (2013). The effect of model uncertainty on cooperation in sensorimotor interactions. *J. R. Soc. Interface* 10:20130554. doi: 10.1098/rsif.2013.0554

Gray, R. D., Greenhill, S. J., and Atkinson, Q. D. (2013). "Phylogenetic models of language change: three new questions," in *Cultural Evolution: Society, Technology, Language, and Religion*, eds P. J. Richerson and M. H. Christiansen (Cambridge, MA: MIT Press), 285–302. doi: 10.7551/mitpress/9780262019750.003.0015

Hadjikhani, N., Åsberg Johnels, J., Zürcher, N. R., Lassalle, A., Guillon, Q., Hippolyte, L., et al. (2017). Look me in the eyes: constraining gaze in the eye-region provokes abnormally high subcortical activation in autism. *Sci. Rep.* 7, 1–7. doi: 10.1038/s41598-017-03378-5

Han, S. (2015). Understanding cultural differences in human behavior: a cultural neuroscience approach. *Curr. Opin. Behav. Sci.* 3, 68–72. doi: 10.1016/j.cobeha.2015.01.013

Han, T. A., Santos, F. C., Lenaerts, T., and Pereira, L. M. (2015). Synergy between intention recognition and commitments in cooperation dilemmas. *Sci. Rep.* 5:9312. doi: 10.1038/srep09312

Hare, B. (2017). Survival of the friendliest: homo sapiens evolved via selection for prosociality. *Annu. Rev. Psychol.* 68, 155–186. doi: 10.1146/annurev-psych-010416-044201

Hare, B., and Tomasello, M. (2005). Human-like social skills in dogs? *Trends Cogn. Sci.* 9, 439–444. doi: 10.1016/j.tics.2005.07.003

Harnad, S. (1990). The symbol grounding problem. *Physica D* 42, 335–346.

Haroush, K., and Williams, Z. M. (2015). Neuronal prediction of opponent's behavior during cooperative social interchange in primates. *Cell* 160, 1233–1245. doi: 10.1016/j.cell.2015.01.045

Harris, P. L., Bartz, D. T., and Rowe, M. L. (2017). Young children communicate their ignorance and ask questions. *Proc. Natl. Acad. Sci. U.S.A.* 114, 7884–7891. doi: 10.1073/pnas.1620745114

Harris, P. L., and Corriveau, K. H. (2011). Young children's selective trust in informants. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 366, 1179–1187. doi: 10.1098/rstb.2010.0321

Haspelmath, M. (1999). Why is grammaticalization irreversible? *Linguist. Philos.* 37, 1043–1068. doi: 10.1515/ling.37.6.1043

Hasson, U., and Frith, C. D. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 371:20150366. doi: 10.1098/rstb.2015.0366

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., and Keysers, C. (2012). Brain-to-brain coupling: a mechanism for creating and sharing a social world. *Trends Cogn. Sci.* 16, 114–121. doi: 10.1016/j.tics.2011.12.007

Heasman, B., and Gillespie, A. (2019). Neurodivergent intersubjectivity: distinctive features of how autistic people create shared understanding. *Autism* 23, 910–921. doi: 10.1177/1362361318785172

Heine, B., and Kuteva, T. (2002). "On the evolution of grammatical forms," in *The Transition to Language*, ed. I. A. Wray (Oxford: Oxford University Press), 376–397.

Henrich, J. (2015). *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., and Gil-White, F. J. (2001). The evolution of prestige: freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evol. Hum. Behav.* 22, 165–196. doi: 10.1016/s1090-5138(00)00071-4

Hinton, G. E. (2007). Learning multiple layers of representation. *Trends Cogn. Sci.* 11, 428–434. doi: 10.1016/j.tics.2007.09.004

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., et al. (2015). The contribution of early communication quality to low-income children's language success. *Psychol. Sci.* 26, 1071–1083. doi: 10.1177/0956797615581493

Höhl, S., Michel, C., Reid, V. M., Parise, E., and Striano, T. (2014). Eye contact during live social interaction modulates infants' oscillatory brain activity. *Soc. Neurosci.* 9, 300–308. doi: 10.1080/17470919.2014.884982

Hohwy, J. (2016). The self-evidencing brain. *Noûs* 50, 259–285. doi: 10.1111/nous.12062

Holroyd, C. B., and Yeung, N. (2012). Motivation of extended behaviors by anterior cingulate cortex. *Trends Cogn. Sci.* 16, 122–128. doi: 10.1016/j.tics.2011.12.008

i Cancho, R. F., and Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proc. Natl. Acad. Sci. U.S.A.* 100, 788–791. doi: 10.1073/pnas.0335980100

Jaswal, V. K., and Akhtar, N. (2019). Being versus appearing socially uninterested: challenging assumptions about social motivation in autism. *Behav. Brain Sci.* 42:e82. doi: 10.1017/S0140525X18001826

Jiang, J., Dai, B., Peng, D., Zhu, C., Liu, L., and Lu, C. (2012). Neural synchronization during face-to-face communication. *J. Neurosci.* 32, 16064–16069. doi: 10.1523/JNEUROSCI.2926-12.2012

Kanai, R., Komura, Y., Shipp, S., and Friston, K. (2015). Cerebral hierarchies: predictive processing, precision and the pulvinar. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 370:20140169. doi: 10.1098/rstb.2014.0169

Kanwal, J., Smith, K., Culbertson, J., and Kirby, S. (2017). Zipf's Law of abbreviation and the principle of least effort: language users optimise a miniature lexicon for efficient communication. *Cognition* 165, 45–52. doi: 10.1016/j.cognition.2017.05.001

Kao, J. T., Wu, J. Y., Bergen, L., and Goodman, N. D. (2014). Nonliteral understanding of number words. *Proc. Natl. Acad. Sci. U.S.A.* 111, 12002–12007. doi: 10.1073/pnas.1407479111

Karmali, F., Whitman, G. T., and Lewis, R. F. (2018). Bayesian optimal adaptation explains age-related human sensorimotor changes. *J. Neurophysiol.* 119, 509–520. doi: 10.1152/jn.00710.2017

Kelso, J. A. S. (2016). On the self-organizing origins of agency. *Trends Cogn. Sci.* 20, 490–499. doi: 10.1016/j.tics.2016.04.004

Kemmer, S. (2003). "Human cognition and the elaboration of events: some universal conceptual categories," in *The New Psychology of Language: Cognitive and Functional Approaches to Language Structure*, Vol. II, ed. M. Tomasello (London: Psychology Press), 95–124.

Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical Bayesian models. *Dev. Sci.* 10, 307–321. doi: 10.1111/j.1467-7687.2007.00585.x

Kidd, E., Donnelly, S., and Christiansen, M. H. (2018). Individual differences in language acquisition and processing. *Trends Cogn. Sci.* 22, 154–169. doi: 10.1016/j.tics.2017.11.006

Kiebel, S. J., Daunizeau, J., and Friston, K. J. (2008). A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4:e1000209. doi: 10.1371/journal.pcbi.1000209

Kiebel, S. J., von Kriegstein, K., Daunizeau, J., and Friston, K. J. (2009). Recognizing sequences of sequences. *PLoS Comput. Biol.* 5:e1000464. doi: 10.1371/journal.pcbi.1000464

Kinreich, S., Djalovski, A., Kraus, L., Louzoun, Y., and Feldman, R. (2017). Brain-to-Brain synchrony during naturalistic social interactions. *Sci. Rep.* 7:17060. doi: 10.1038/s41598-017-17339-5

Kirby, S., Dowman, M., and Griffiths, T. L. (2007). Innateness and culture in the evolution of language. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5241–5245. doi: 10.1073/pnas.0608222104

Kirby, S., Tamariz, M., Cornish, H., and Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition* 141, 87–102. doi: 10.1016/j.cognition.2015.03.016

Kishimoto, T., Shizawa, Y., Yasuda, J., Hinobayashi, T., and Minami, T. (2007). Do pointing gestures by infants provoke comments from adults? *Infant Behav. Dev.* 30, 562–567. doi: 10.1016/j.infbeh.2007.04.001

Köster, M., Langeloh, M., and Höhl, S. (2019). Visually entrained theta oscillations increase for unexpected events in the infant brain. *Psych. Sci.* 30, 1656–1663. doi: 10.1177/0956797619876260

Kovács, Á. M., and Endress, A. D. (2014). Hierarchical processing in seven-month-old infants. *Infancy* 19, 409–425. doi: 10.1111/infa.12052

Kovács, Á. M., Tauzin, T., Téglás, E., Gergely, G., and Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy* 19, 543–557. doi: 10.1111/infa.12060

Kovács, Á. M., Téglás, E., Gergely, G., and Csibra, G. (2017). Seeing behind the surface: communicative demonstration boosts category disambiguation in 12-month-olds. *Developmental Science* 20:e12485. doi: 10.1111/desc.12485

Kuperberg, G. R., and Jaeger, T. F. (2016). What do we mean by prediction in language comprehension? *Lang. Cogn. Neurosci.* 31, 32–59. doi: 10.1080/23273798.2015.1102299

Lavin, C., Melis, C., Mikulan, E., Gelormini, C., Huepe, D., and Ibañez, A. (2013). The anterior cingulate cortex: an integrative hub for human socially-driven interactions. *Front. Neurosci.* 7:64. doi: 10.3389/fnins.2013.00064

Lawson, R. P., Rees, G., and Friston, K. J. (2014). An aberrant precision account of autism. *Front. Hum. Neurosci.* 8:302. doi: 10.3389/fnhum.2014.00302

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature* 521, 436–444. doi: 10.1038/nature14539

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition* 106, 1126–1177. doi: 10.1016/j.cognition.2007.05.006

Lewis, D. (1969). *Convention: A Philosophical Study*. Hoboken, NJ: Wiley-Blackwell.

Liebal, K., Behne, T., Carpenter, M., and Tomasello, M. (2009). Infants use shared experience to interpret pointing gestures. *Dev. Sci.* 12, 264–271. doi: 10.1111/j.1467-7687.2008.00758.x

Liebal, K., Carpenter, M., and Tomasello, M. (2010). Infants' use of shared experience in declarative pointing. *Infancy* 15, 545–556. doi: 10.1111/j.1532-7078.2009.00028.x

Liebenberg, L. (2006). Persistence hunting by modern hunter-gatherers. *Curr. Anthropol.* 47, 1017–1026. doi: 10.1086/508695

Lieven, E. (2016). Usage-based approaches to language development: where do we go from here? *Lang. Cogn.* 8, 346–368. doi: 10.1017/langcog.2016.16

Lieven, E., and Stoll, S. (2013). Early communicative development in two cultures: a comparison of the communicative environments of children from two cultures. *Hum. Dev.* 56, 178–206. doi: 10.1159/000351073

Lindsay, L., Gambi, C., and Rabagliati, H. (2019). Preschoolers optimize the timing of their conversational turns through flexible coordination of language

comprehension and production. *Psychol. Sci.* 30, 504–515. doi: 10.1177/0956797618822802

Linsker, R. (1990). Perceptual neural organization: some approaches based on network models and information theory. *Annu. Rev. Neurosci.* 13, 257–281. doi: 10.1146/annurev.ne.13.030190.001353

Liszkowski, U., Brown, P., Callaghan, T., Takada, A., and de Vos, C. (2012). A prelinguistic gestural universal of human communication. *Cogn. Sci.* 36, 698–713. doi: 10.1111/j.1551-6709.2011.01228.x

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., and Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Dev. Sci.* 7, 297–307. doi: 10.1111/j.1467-7687.2004.00349.x

Liszkowski, U., Carpenter, M., and Tomasello, M. (2007). Reference and attitude in infant pointing. *J. Child Lang.* 34, 1–20. doi: 10.1017/s0305000906007689

Liszkowski, U., Schäfer, M., Carpenter, M., and Tomasello, M. (2009). Prelinguistic infants, but not chimpanzees, communicate about absent entities. *Psychol. Sci.* 20, 654–660. doi: 10.1111/j.1467-9280.2009.02346.x

Liu, Y., Piazza, E. A., Simony, E., Shewokis, P. A., Onaral, B., Hasson, U., et al. (2017). Measuring speaker–listener neural coupling with functional near infrared spectroscopy. *Sci. Rep.* 7:43293. doi: 10.1038/srep43293

Loveland, K. A., and Landry, S. H. (1986). Joint attention and language in autism and developmental language delay. *J. Autism Dev. Disord.* 16, 335–349. doi: 10.1007/bf01531663

Lucca, K., and Wilbourn, M. P. (2018a). Communicating to learn: infants' pointing gestures result in optimal learning. *Child Dev.* 89, 941–960. doi: 10.1111/cdev.12707

Lucca, K., and Wilbourn, M. P. (2018b). The what and the how: information-seeking pointing gestures facilitate learning labels and functions. *J. Exp. Child Psychol.* 178, 417–436. doi: 10.1016/j.jecp.2018.08.003

Lupyan, G., and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One* 5:e8559. doi: 10.1371/journal.pone.0008559

MacKay, D. J. (1995). Free-energy minimisation algorithm for decoding and cryptoanalysis. *Electron. Lett.* 31, 445–447.

MacLean, E. L. (2016). Unraveling the evolution of uniquely human cognition. *Proc. Natl. Acad. Sci. U.S.A.* 113, 6348–6354. doi: 10.1073/pnas.1521270113

Marno, H., Guellai, B., Vidal, Y., Franzoi, J., Nespor, M., and Mehler, J. (2016). Infants' selectively pay attention to the information they receive from a native speaker of their language. *Front. Psychol.* 7:1150. doi: 10.3389/fpsyg.2016.01150

Matthews, D., Behne, T., Lieven, E., and Tomasello, M. (2012). Origins of the human pointing gesture: a training study. *Dev. Sci.* 15, 817–829. doi: 10.1111/j.1467-7687.2012.01181.x

McCauley, S. M., and Christiansen, M. H. (2014). Prospects for usage-based computational models of grammatical development: argument structure and semantic roles. *Wiley Interdiscip. Rev. Cogn. Sci.* 5, 489–499. doi: 10.1002/wcs.1295

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., et al. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* 14, 348–356. doi: 10.1016/j.tics.2010.06.002

McLoone, B., and Smead, R. (2014). The ontogeny and evolution of human collaboration. *Biol. Philos.* 29, 559–576. doi: 10.1007/s10539-014-9435-1

Meltzoff, A. N. (2007). "Like me": a foundation for social cognition. *Dev. Sci.* 10, 126–134. doi: 10.1111/j.1467-7687.2007.00574.x

Meylan, S. C., Frank, M. C., Roy, B. C., and Levy, R. (2017). The emergence of an abstract grammatical category in children's early speech. *Psychol. Sci.* 28, 181–192. doi: 10.1177/0956797616677753

Mirza, M. B., Adams, R. A., Friston, K., and Parr, T. (2019). Introducing a Bayesian model of selective attention based on active inference. *Sci. Rep.* 9:13915. doi: 10.1038/s41598-019-50138-8

Mirza, M. B., Adams, R. A., Mathys, C. D., and Friston, K. J. (2016). Scene construction, visual foraging, and active inference. *Front. Comput. Neurosci.* 10:56. doi: 10.3389/fncom.2016.00056

Mitani, J. (2009). Cooperation and competition in chimpanzees: current understanding and future challenges. *Evol. Anthropol.* 18, 215–227. doi: 10.1002/evan.20229

Moll, H., Carpenter, M., and Tomasello, M. (2007). Fourteen−month−olds know what others experience only in joint engagement. *Dev. Sci.* 10, 826–835. doi: 10.1111/j.1467-7687.2007.00615.x

Moll, H., and Tomasello, M. (2007). Cooperation and human cognition: the Vygotskian intelligence hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 639–648. doi: 10.1098/rstb.2006.2000

Moran, R. J., Campo, P., Symmonds, M., Stephan, K. E., Dolan, R. J., and Friston, K. J. (2013). Free energy, precision and learning: the role of cholinergic neuromodulation. *J. Neurosci.* 33, 8227–8236. doi: 10.1523/JNEUROSCI.4255-12.2013

Moran, R. J., Symmonds, M., Dolan, R. J., and Friston, K. J. (2014). The brain ages optimally to model its environment: evidence from sensory learning over the adult lifespan. *PLoS Comput. Biol.* 10:e1003422. doi: 10.1371/journal.pcbi.1003422

Moseley, R., Carota, F., Hauk, O., Mohr, B., and Pulvermüller, F. (2012). A role for the motor system in binding abstract emotional meaning. *Cereb. Cortex* 22, 1634–1647. doi: 10.1093/cercor/bhr238

Moutoussis, M., Fearon, P., El-Deredy, W., Dolan, R. J., and Friston, K. J. (2014). Bayesian inferences about the self (and others): a review. *Conscious. Cogn.* 25, 67–76. doi: 10.1016/j.concog.2014.01.009

Nakamura, M., and Ohtsuki, H. (2016). Optimal decision rules in repeated games where players infer an opponent's mind via simplified belief calculation. *Games Vol.* 7:19. doi: 10.3390/g7030019

Nelson, J. D., Divjak, B., Gudmundsdottir, G., Martignon, L. F., and Meder, B. (2014). Children's sequential information search is sensitive to environmental probabilities. *Cognition* 130, 74–80. doi: 10.1016/j.cognition.2013.09.007

Nyström, P., Thorup, E., Bölte, S., and Falck-Ytter, T. (2019). Joint attention in infancy and the emergence of autism. *Biol. Psychiatry* 86, 631–638. doi: 10.1016/j.biopsych.2019.05.006

Parr, T., and Friston, K. (2017). Working memory, attention, and salience in active inference. *Sci. Rep.* 7:14678. doi: 10.1038/s41598-017-15249-0

Parr, T., and Friston, K. J. (2018). The anatomy of inference: generative models and brain structure. *Front. Comput. Neurosci.* 12:90. doi: 10.3389/fncom.2018.00090

Pea, R. D. (1979). Can information theory explain early word choice. *J. Child Lang.* 6, 397–410. doi: 10.1017/s0305000900002440

Pecora, L. M., Carroll, T. L., and Heagy, J. F. (1995). Statistics for mathematical properties of maps between time series embeddings. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Topics* 52, 3420–3439. doi: 10.1103/physreve.52.3420

Pecora, L. M., Carroll, T. L., Johnson, G. A., Mar, D. J., and Heagy, J. F. (1997). Fundamentals of synchronization in chaotic systems, concepts, and applications. *Chaos* 7, 520–543. doi: 10.1063/1.166278

Pérez, A., Carreiras, M., and Duñabeitia, J. A. (2017). Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening. *Sci. Rep.* 7:4190. doi: 10.1038/s41598-017-04464-4

Perfors, A., and Navarro, D. J. (2014). Language evolution can be shaped by the structure of the world. *Cogn. Sci.* 38, 775–793. doi: 10.1111/cogs.12102

Perfors, A., Tenenbaum, J. B., and Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition* 118, 306–338. doi: 10.1016/j.cognition.2010.11.001

Perniss, P., and Vigliocco, G. (2014). The bridge of iconicity: from a world of experience to the experience of language. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369:20130300. doi: 10.1098/rstb.2013.0300

Pezzulo, G. (2011). Shared representations as coordination tools for interaction. *Rev. Philos. Psychol.* 2, 303–333. doi: 10.1007/s13164-011-0060-5

Pezzulo, G., Donnarumma, F., and Dindo, H. (2013). Human sensorimotor communication: a theory of signaling in online social interactions. *PLoS One* 8:e79876. doi: 10.1371/journal.pone.0079876

Pezzulo, G., Rigoli, F., and Friston, K. (2015). Active Inference, homeostatic regulation and adaptive behavioural control. *Prog. Neurobiol.* 134, 17–35. doi: 10.1016/j.pneurobio.2015.09.001

Pezzulo, G., Rigoli, F., and Friston, K. J. (2018). Hierarchical active inference: a theory of motivated control. *Trends Cogn. Sci.* 22, 294–306. doi: 10.1016/j.tics.2018.01.009

Powers, A. R. III, Kelley, M., and Corlett, P. R. (2016). Hallucinations as top-down effects on perception. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* 1, 393–400. doi: 10.1016/j.bpsc.2016.04.003

Prigogine, I., and Stengers, I. (1984). *Order Out of Chaos: Man's New Dialogue with Nature*. New York, NY: Bantam New Age Books.

Pulvermüller, F. (2013). How neurons make meaning: brain mechanisms for embodied and abstract-symbolic semantics. *Trends Cogn. Sci.* 17, 458–470. doi: 10.1016/j.tics.2013.06.004

Pulvermüller, F. (2015). "Language, action, interaction: neuropragmatic perspectives on symbols, meaning, and context-dependent function," in *The Pragmatic Turn: Toward Action-Oriented Views in Cognitive*, eds A. K. Engel, K. J. Friston, and D. Kragic (Boston, MA: MIT Press).

Rabinovich, M. I., Simmons, A. N., and Varona, P. (2015). Dynamical bridge between brain and mind. *Trends Cogn. Sci.* 19, 453–461. doi: 10.1016/j.tics.2015.06.005

Rabinovich, M. I., Varona, P., Tristan, I., and Afraimovich, V. S. (2014). Chunking dynamics: heteroclinics in mind. *Front. Comput. Neurosci.* 8:22. doi: 10.3389/fncom.2014.00022

Ramstead, M. J. D., Badcock, P. B., and Friston, K. J. (2018). Answering Schrödinger's question: a free-energy formulation. *Phys. Life Rev.* 24, 1–16. doi: 10.1016/j.plrev.2017.09.001

Ramstead, M. J. D., Veissière, S. P. L., and Kirmayer, L. J. (2016). Cultural affordances: scaffolding local worlds through shared intentionality and regimes of attention. *Front. Psychol.* 7:1090. doi: 10.3389/fpsyg.2016.01090

Reali, F., Chater, N., and Christiansen, M. H. (2018). Simpler grammar, larger vocabulary: how population size affects language. *Proc. Biol. Sci. R. Soc. Lond. B Biol. Sci.* 285:20172586. doi: 10.1098/rspb.2017.2586

Reddy, V. (2003). On being the object of attention: implications for self–other consciousness. *Trends Cogn. Sci.* 7, 397–402. doi: 10.1016/S1364-6613(03)00191-8

Rekers, Y., Haun, D. B. M., and Tomasello, M. (2011). Children, but not chimpanzees, prefer to collaborate. *Curr. Biol.* 21, 1756–1758. doi: 10.1016/j.cub.2011.08.066

Renzi, D. T., Romberg, A. R., Bolger, D. J., and Newman, R. S. (2017). Two minds are better than one: cooperative communication as a new framework for understanding infant language learning. *Transl. Issues Psychol. Sci.* 3, 19–33. doi: 10.1037/tps0000088

Rhodes, M., and Wellman, H. (2017). Moral learning as intuitive theory revision. *Cognition* 167, 191–200. doi: 10.1016/j.cognition.2016.08.013

Richerson, P. J., and Boyd, R. (2005). *Not By Genes Alone*. Chicago, IL: University of Chicago Press.

Riley, M. A., Richardson, M. J., Shockley, K., and Ramenzoni, V. C. (2011). Interpersonal synergies. *Front. Psychol.* 2:38. doi: 10.3389/fpsyg.2011.00038

Roepstorff, A., Niewöhner, J., and Beck, S. (2010). Enculturing brains through patterned practices. *Neural Netw.* 23, 1051–1059. doi: 10.1016/j.neunet.2010.08.002

Romberg, A. R., and Saffran, J. R. (2010). Statistical leaning and language acquisition. *Wiley Interdiscip. Rev. Cogn. Sci.* 1, 906–914. doi: 10.1002/wcs.78

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., et al. (2018). Beyond the 30-million-word gap: children's conversational exposure is associated with language-related brain function. *Psychol. Sci.* 29, 700–710. doi: 10.1177/0956797617742725

Sanders, J. B. T., Farmer, J. D., and Galla, T. (2018). The prevalence of chaotic dynamics in games with many players. *Sci. Rep.* 8:4902. doi: 10.1038/s41598-018-22013-5

Saylor, M. M., and Ganea, P. (2007). Infants interpret ambiguous requests for absent objects. *Dev. Psychol.* 43, 696–704. doi: 10.1037/0012-1649.43.3.696

Schegloff, E. A. (2006). "Interaction: the infrastructure for social institutions, the natural ecological niche for language, and the arena in which culture is enacted," in *Roots of Human Sociality: Culture, Cognition and Interaction*, eds N. J. Enfield and S. C. Levinson (London: Berg Publishers), 70–96.

Schegloff, E. A., Jefferson, G., and Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language* 53, 361–382. doi: 10.1353/lan.1977.0041

Schelling, T. C. (1960). *The Strategy of Conflict*. Cambridge, MA: Harvard University Press.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., et al. (2013). Toward a second-person neuroscience. *Behav. Brain Sci.* 36, 393–414. doi: 10.1017/S0140525X12000660

Schippers, M. B., Roebroeck, A., Renken, R., Nanetti, L., and Keysers, C. (2010). Mapping the information flow from one brain to another during gestural communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 9388–9393. doi: 10.1073/pnas.1001791107

Schmälzle, R., Häcker, F. E. K., Honey, C. J., and Hasson, U. (2015). Engaged listeners: shared neural processing of powerful political speeches. *Soc. Cogn. Affect. Neurosci.* 10, 1137–1143. doi: 10.1093/scan/nsu168

Schmidhuber, J. (2010). Formal theory of creativity, fun, and intrinsic motivation (1990-2010). *IEEE Trans. Auton. Ment. Dev.* 2, 230–247. doi: 10.1109/tamd.2010.2056368

Schoot, L., Hagoort, P., and Segaert, K. (2016). What can we learn from a two-brain approach to verbal interaction? *Neurosci. Biobehav. Rev.* 68, 454–459. doi: 10.1016/j.neubiorev.2016.06.009

Sella, G., and Hirsh, A. E. (2005). The application of statistical physics to evolutionary biology. *Proc. Natl. Acad. Sci. U.S.A.* 102, 9541–9546. doi: 10.1073/pnas.0501865102

Senghas, A. (2003). Intergenerational influence and ontogenetic development in the emergence of spatial grammar in Nicaraguan Sign Language. *Cogn. Dev.* 18, 511–531. doi: 10.1016/j.cogdev.2003.09.006

Sengupta, B., Tozzi, A., Cooray, G. K., Douglas, P. K., and Friston, K. J. (2016). Towards a Neuronal Gauge theory. *PLoS Biol.* 14:e1002400. doi: 10.1371/journal.pbio.1002400

Seoane, L. F., and Solé, R. (2018). The morphospace of language networks. *Sci. Rep.* 8:10465. doi: 10.1038/s41598-018-28820-0

Sewell, W. H. (1992). A theory of structure: duality, agency, and transformation. *Am. J. Sociol.* 98, 1–29. doi: 10.1086/229967

Seyfarth, R. M., and Cheney, D. L. (2003). Signalers and receivers in animal communication. *Annu. Rev. Psychol.* 54, 145–173. doi: 10.1146/annurev.psych.54.101601.145121

Shuai, L., and Gong, T. (2014). Language as an emergent group-level trait. *Behav. Brain Sci.* 37, 274–275. doi: 10.1017/S0140525X13003026

Shweder, R. A., and Sullivan, M. A. (1993). Cultural psychology: who needs it? *Annu. Rev. Psychol.* 44, 497–523. doi: 10.1146/annurev.ps.44.020193.002433

Siposova, B., and Carpenter, M. (2019). A new look at joint attention and common knowledge. *Cognition* 189, 260–274. doi: 10.1016/j.cognition.2019.03.019

Siposova, B., Tomasello, M., and Carpenter, M. (2018). Communicative eye contact signals a commitment to cooperate for young children. *Cognition* 179, 192–201. doi: 10.1016/j.cognition.2018.06.010

Skyrms, B. (2001). The Stag Hunt. *Proc. Address. Am. Philos. Assoc.* 75, 31–41. doi: 10.2307/3218711

Smith, K., Perfors, A., Fehér, O., Samara, A., Swoboda, K., and Wonnacott, E. (2017). Language learning, language use and the evolution of linguistic variation. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 372:20160051. doi: 10.1098/rstb.2016.0051

Sperber, D., and Wilson, D. (1986). *Relevance: Communication and Cognition*. Hoboken, NJ: Wiley-Blackwell.

Sperber, D., and Wilson, D. (1987). Précis of relevance: communication and cognition. *Behav. Brain Sci.* 10, 697–710. doi: 10.1017/S0140525X00055345

Stephens, G. J., Silbert, L. J., and Hasson, U. (2010). Speaker–listener neural coupling underlies successful communication. *Proc. Natl. Acad. Sci. U.S.A.* 107, 14425–14430. doi: 10.1073/pnas.1008662107

Stolk, A., Noordzij, M. L., Verhagen, L., Volman, I., Schoffelen, J.-M., Oostenveld, R., et al. (2014). Cerebral coherence between communicators marks the emergence of meaning. *Proc. Natl. Acad. Sci. U.S.A.* 111, 18183–18188. doi: 10.1073/pnas.1414886111

Stolk, A., Verhagen, L., and Toni, I. (2016). Conceptual alignment: how brains achieve mutual understanding. *Trends Cogn. Sci.* 20, 180–191. doi: 10.1016/j.tics.2015.11.007

Striano, T., and Stahl, D. (2005). Sensitivity to triadic attention in early infancy. *Dev. Sci.* 8, 333–343. doi: 10.1111/j.1467-7687.2005.00421.x

Suzuki, S. (1970/2014). *Zen Mind, Beginner's Mind: Informal Talks on Zen Meditation and Practice*. Boulder, CO: Shambhala Publications.

Szathmáry, E. (2015). Toward major evolutionary transitions theory 2.0. *Proc. Natl. Acad. Sci. U.S.A.* 112, 10104–10111. doi: 10.1073/pnas.1421398112

Szufnarowska, J., Rohlfing, K. J., Fawcett, C., and Gredebäck, G. (2014). Is ostension any more than attention? *Sci. Rep.* 4:5304. doi: 10.1038/srep05304

Tager-Flusberg, H. (2007). Evaluating the theory-of-mind hypothesis of autism. *Curr. Dir. Psychol.Sci.* 16, 311–315. doi: 10.1111/j.1467-8721.2007.00527.x

Tamariz, M., and Kirby, S. (2016). The cultural evolution of language. *Curr. Opin. Psychol.* 8, 37–43. doi: 10.1016/j.copsyc.2015.09.003

Tanaka, J. W., and Sung, A. (2016). The "eye avoidance" hypothesis of autism face processing. *J. Autism Dev. Disord.* 46, 1538–1552. doi: 10.1007/s10803-013-1976-7

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., and Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science* 331, 1279–1285. doi: 10.1126/science.1192788

Thomas, M. S. C., Davis, R., Karmiloff-Smith, A., Knowland, V. C. P., and Charman, T. (2016). The over-pruning hypothesis of autism. *Dev. Sci.* 19, 284–305. doi: 10.1111/desc.12303

Thomas, M. S. C., Fedor, A., Davis, R., Yang, J., Alireza, H., Charman, T., et al. (2019). Computational modeling of interventions for developmental disorders. *Psychol. Rev.* 126, 693–726. doi: 10.1037/rev0000151

Tinbergen, N. (1963). On aims and methods of Ethology. *Z. Tierpsychol.* 20, 410–433. doi: 10.1111/j.1439-0310.1963.tb01161.x

Tomasello, M. (1995). "Joint attention as social cognition," in *Joint Attention: Its Origins and Role in Development*, eds C. Moore and P. J. Dunham (Hillsdale, NJ: Lawrence Erlbaum Associates), 103–130.

Tomasello, M. (2000). The item-based nature of children's early syntactic development. *Trends Cogn. Sci.* 4, 156–163. doi: 10.1016/s1364-6613(00)01462-5

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Approach to Child Language Acquisition*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2004). What kind of evidence could refute the UG hypothesis? Commentary on Wunderlich. *Stud. Lang.* 28, 642–645. doi: 10.1075/sl.28.3.13tom

Tomasello, M. (2008). *Origins of Human Communication*. Cambridge, MA: MIT Press.

Tomasello, M. (2014). *A Natural History of Human Thinking*. Cambridge, MA: Harvard University Press.

Tomasello, M. (2019). *Becoming Human: A Theory of Ontogeny*. Cambridge, MA: Harvard University Press.

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691; discussion 691–735. doi: 10.1017/S0140525X05000129

Tomasello, M., Carpenter, M., and Liszkowski, U. (2007). A new look at infant pointing. *Child Dev.* 78, 705–722. doi: 10.1111/j.1467-8624.2007.01025.x

Tomasello, M., and Haberl, K. (2003). Understanding attention: 12- and 18-month-olds know what is new for other persons. *Dev. Psychol.* 39, 906–912. doi: 10.1037/0012-1649.39.5.906

Tomasello, M., Melis, A. P., Tennie, C., Wyman, E., and Herrmann, E. (2012). Two key steps in the evolution of human cooperation: the interdependence hypothesis. *Curr. Anthropol.* 53, 673–692. doi: 10.1086/668207

Tschacher, W., and Haken, H. (2007). Intentionality in non-equilibrium systems? The functional aspects of self-organized pattern formation. *New Ideas Psychol.* 25, 1–15. doi: 10.1016/j.newideapsych.2006.09.002

van den Heuvel, M. P., and Sporns, O. (2013). Network hubs in the human brain. *Trends Cogn. Sci.* 17, 683–696. doi: 10.1016/j.tics.2013.09.012

Veissière, S. P. L., Constant, A., Ramstead, M. J. D., Friston, K. J., and Kirmayer, L. J. (forthcoming). Thinking through other minds: a variational approach to cognition and culture. *Behav. Brain Sci.* 1–97. doi: 10.1017/S0140525X19001213

Verhagen, A. (2007). "Construal and perspectivisation," in *The Oxford Handbook of Cognitive Linguistics*, eds D. Geeraerts and H. Cuyckens (Oxford: Oxford University Press), 48–81.

Vygotsky, L. S. (1978). *Mind in Society: The development of higher mental process*. Cambridge, MA: Harvard University Press.

Wallace, C. S., and Dowe, D. L. (1999). Minimum message length and kolmogorov complexity. *Comput. J.* 42, 270–283. doi: 10.1093/comjnl/42.4.270

Warlaumont, A. S., Richards, J. A., Gilkerson, J., and Oller, D. K. (2014). A social feedback loop for speech development and its reduction in autism. *Psychol. Sci.* 25, 1314–1324. doi: 10.1177/0956797614531023

Whiten, A., and Erdal, D. (2012). The human socio-cognitive niche and its evolutionary origins. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 2119–2129. doi: 10.1098/rstb.2012.0114

Wilcox, S. (2004). Language from gesture. *Behav. Brain Sci.* 27, 525–526. doi: 10.1017/s0140525x04470113

Williams, S., and Yaeger, L. (2017). Evolution of neural dynamics in an ecological model. *Geosci. J.* 7:49. doi: 10.3390/geosciences7030049

Winters, J., Kirby, S., and Smith, K. (2018). Contextual predictability shapes signal autonomy. *Cognition* 176, 15–30. doi: 10.1016/j.cognition.2018.03.002

Wipf, D. P., and Rao, B. D. (2007). An empirical Bayesian strategy for solving the simultaneous sparse approximation problem. *IEEE Trans. Signal Process.* 55, 3704–3716. doi: 10.1109/tsp.2007.894265

Wolpe, N., Ingram, J. N., Tsvetanov, K. A., Geerligs, L., Kievit, R. A., Henson, R. N., et al. (2016). Ageing increases reliance on sensorimotor prediction through structural and functional differences in frontostriatal circuits. *Nat. Commun.* 7:13034. doi: 10.1038/ncomms13034

Wu, Z., and Gros-Louis, J. (2015). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *J. Child Lang.* 42, 538–561. doi: 10.1017/S0305000914000221

Wyman, E., Rakoczy, H., and Tomasello, M. (2013). Non-verbal communication enables children's coordination in a "Stag Hunt" game. *Eur. J. Dev. Psychol.* 10, 597–610. doi: 10.1080/17405629.2012.726469

Yildiz, I. B., von Kriegstein, K., and Kiebel, S. J. (2013). From birdsong to human speech recognition: Bayesian inference on a hierarchy of nonlinear dynamical systems. *PLoS Comput. Biol.* 9:e1003219. doi: 10.1371/journal.pcbi.1003219

Yoshida, W., Dolan, R. J., and Friston, K. J. (2008). Game theory of mind. *PLoS Comput. Biol.* 4:e1000254. doi: 10.1371/journal.pcbi.1000254

Young, G. S., Merin, N., Rogers, S. J., and Ozonoff, S. (2009). Gaze behavior and affect at 6 months: predicting clinical outcomes and language development in typically developing infants and infants at risk for autism. *Dev. Sci.* 12, 798–814. doi: 10.1111/j.1467-7687.2009.00833.x

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., and Hasson, U. (2017). How we transmit memories to other brains: constructing shared neural representations via communication. *Cereb. Cortex* 27, 4988–5000. doi: 10.1093/cercor/bhx202

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Vasil, Badcock, Constant, Friston and Ramstead. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.