Behavioral and Brain Sciences

cambridge.org/bbs

Thinking through other minds: A variational approach to cognition and culture

Samuel P. L. Veissière[a,b,c], Axel Constant[b,d,e], Maxwell J. D. Ramstead[a,b,e], Karl J. Friston[e] and Laurence J. Kirmayer[a,b,c]

[a]Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, Montreal, Quebec, Canada H3A 1A1; [b]Culture, Mind, and Brain Program, McGill University, Montreal, Quebec, Canada H3A 1A1; [c]Department of Anthropology, McGill University, Montreal, Quebec, Canada H3A 2T7; [d]Charles Perkins Centre, The University of Sydney, Sydney, New South Wales, Australia 2006 and [e]Wellcome Centre for Human Neuroimaging, University College London, London WC1N 3AR, UK.
samuel.veissiere@mcgill.ca axel.constant.pruvost@gmail.com Maxwell.ramstead@mcgill.ca
k.friston@ucl.ac.uk Laurence.kirmayer@mcgill.ca

Target Article

Cite this article: Veissière SPL, Constant A, Ramstead MJD, Friston KJ, Kirmayer LJ. (2020) Thinking through other minds: A variational approach to cognition and culture. *Behavioral and Brain Sciences* **43**, e90: 1–75. doi:10.1017/S0140525X19001213

Target Article Accepted: 5 May 2019
Target Article Manuscript Online: 30 May 2019
Commentaries Accepted: 11 January 2020

Keywords:
Cognition and culture; cultural affordances; cultural co-evolution; cultural learning; embodiment; enactment; enculturation; epistemic affordances; niche construction; social learning; variational free-energy principle

What is Open Peer Commentary? What follows on these pages is known as a Treatment, in which a significant and controversial Target Article is published along with Commentaries (p. 23) and an Authors' Response (p. 55). See bbsonline.org for more information.

Abstract

The processes underwriting the acquisition of culture remain unclear. How are shared habits, norms, and expectations learned and maintained with precision and reliability across large-scale sociocultural ensembles? Is there a unifying account of the mechanisms involved in the acquisition of culture? Notions such as "shared expectations," the "selective patterning of attention and behaviour," "cultural evolution," "cultural inheritance," and "implicit learning" are the main candidates to underpin a unifying account of cognition and the acquisition of culture; however, their interactions require greater specification and clarification. In this article, we integrate these candidates using the variational (free-energy) approach to human cognition and culture in theoretical neuroscience. We describe the construction by humans of social niches that afford epistemic resources called cultural affordances. We argue that human agents learn the shared habits, norms, and expectations of their culture through immersive participation in patterned cultural practices that selectively pattern attention and behaviour. We call this process "thinking through other minds" (TTOM) – in effect, the process of inferring other agents' expectations about the world and how to behave in social context. We argue that for humans, information from and about other people's expectations constitutes the primary domain of statistical regularities that humans leverage to predict and organize behaviour. The integrative model we offer has implications that can advance theories of cognition, enculturation, adaptation, and psychopathology. Crucially, this formal (variational) treatment seeks to resolve key debates in current cognitive science, such as the distinction between internalist and externalist accounts of theory of mind abilities and the more fundamental distinction between dynamical and representational accounts of enactivism.

> [Humans] form with others joint goals to which both parties are normatively committed, they establish with others domains of joint attention and common conceptual ground, and they create with others symbolic, institutional realities that assign deontic powers to otherwise inert entities.
>
> —Michael Tomasello (2009, p. 105)

> Choosing a swimsuit—
> when did his eyes replace mine?
> *(mizugi erabu itsu shika kare no me to natte)*
>
> —Mayuzumi Madoka (2003, p. 232)[1]

1. Introduction: Learning in cultural context

1.1. The puzzle of implicit cultural learning

Since the advent of the social sciences in the late nineteenth century, a recurring trope casts "society" or, in its Durkheimian formulation, "regulatory social forces" (Durkheim 1985/2014) as superordinate to individual human agency. As the story goes, humans acquire norms, tastes, preferences, and ways of doing things that are consistent with those of others in their local world and communities – that is, the relevant social and cultural groups (in-groups and out-groups) to which they belong and with whom they interact (Kurzban and Neuberg 2005).

Group variations in learned and structured dispositions extend to such domains as culturally shaped body practices like walking, sitting, eating, and sleeping (Mauss 1973); differentiated patterns of prejudice or bias against certain kinds of persons (e.g., racism, sexism, and classism; Machery 2016); proneness to optical illusions (McCauley & Henrich 2006); colour perception (Goldstein et al. 2009); food preferences (Wright et al. 2001); desirable body types (Swami et al. 2010); and thresholds for pain (Zatzick & Dimsdale 1990) and other

© The Author(s), 2019. Published by Cambridge University Press

CAMBRIDGE UNIVERSITY PRESS

Samuel Veissière, Ph.D., is Assistant Professor of Psychiatry and Anthropology and Co-Director of the Culture, Mind, and Brain Program at McGill University. He has published broadly on the co-evolution of cognition and culture, sociality, and social dimensions of attention, psychopathology, and healing. His work on the Cultural Affordances model combines experimental, theoretical, and ethnographic approaches to the study of such phenomena as hypnosis and placebo effects, hypersociality in Internet and smartphone addiction, ideology and polarization, and cultural shifts in gender relations.

Axel Constant is a doctoral candidate at the University of Sydney. He has authored a series of articles applying active inference to areas including evolutionary biology, cognitive psychology, and the philosophy of psychiatry, including 'A Variational Approach to Niche Construction' in the *Journal of the Royal Society Interface*. He has held honorary positions as a visiting fellow at the Wellcome Center for Human Neuroimaging of the University College London and the Culture Mind and Brain Program at McGill University. He received a talent doctoral award from the Social Sciences and Humanities Research Council of Canada.

Maxwell Ramstead, Ph.D. is currently the Utting Fellow in the Department of Psychiatry at the Jewish General Hospital in Montréal. He is affiliated with the Division of Social and Transcultural Psychiatry and the Culture, Mind, and Brain program at McGill University, and the Wellcome Trust Centre for Neuroimaging of University College London. His research explores active inference and multiscale explanation in psychiatry, the cognitive sciences, and the computational neurosciences. He is the author of over a dozen peer reviewed publications in journals such as *Physics of Life Reviews* and *Synthese*.

Karl Friston is a theoretical neuroscientist and authority on brain imaging. He invented statistical parametric mapping (SPM), voxel-based morphometry (VBM) and dynamic causal modelling (DCM). His main contribution to theoretical neurobiology is a free-energy principle for action and perception (active inference). Friston was elected a Fellow of the Academy of Medical Sciences (1999). In 2003 he was awarded the Minerva Golden Brain Award and was elected a Fellow of the Royal Society in 2006. In 2008 he received a Medal, College de France and an Honorary Doctorate from the University of York in 2011. He became of Fellow of the Royal Society of Biology in 2012, received the Weldon Memorial prize and Medal in 2013 and was elected as a member of EMBO in 2014 and the Academia Europaea in (2015). He was the 2016 recipient of the Charles Branch Award and the Glass Brain Award. He holds Honorary Doctorates from the University of Zurich and Radboud University.

Laurence J. Kirmayer, M.D., F.R.C.P.C., F.C.A.H.S., F.R.S.C. is James McGill Professor and Director, Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, and Co-Director of the McGill-FPR Culture, Mind and Brain Program. He also directs the Culture & Mental Health Research Unit at the Institute of Community and Family Psychiatry, Jewish General Hospital, in Montreal, where he conducts research on culturally responsive mental health services, mental health promotion, and the anthropology of psychiatry. His publications include the co-edited volumes, *Cultural Consultation: Encountering the Other in Mental Health Care* (Springer, 2013); *Re-Visioning Psychiatry: Cultural Phenomenology, Critical Neuroscience, and Global Mental Health* (Cambridge, 2015) and *Culture, Mind and Brsain: Emerging Concepts, Methods, and Applications* (Cambridge, 2020). He is a Fellow of the Canadian Academy of Health Sciences and the Royal Society of Canada.

forms of suffering and affliction that are shaped by culture (Kirmayer 1989; Kirmayer & Young 1998; Kirmayer et al. 2017) and historical context (Gold & Gold 2015; Hacking 1998). As developmental psychologists have argued, it is precisely because of the existence of intergroup behavioural and cognitive variations that arise through social learning within members of the same species that we can speak of culture (Tomasello 2009). We know there is such a "thing" as culture, in other words, because there are cultural differences (Brown 2004). Although it is clear that specific developmental experiences – governed by explicit social norms and contexts – shape these perceptual, cognitive, and attitudinal processes, most of cultural learning appears to be *implicit*, in the sense that it occurs without explicit instruction.

Implicit cultural learning poses a classical "poverty of stimulus" problem, in that acquired knowledge, attitudes, and dispositions appear to go far beyond what can be learned by direct experience (Berwick et al. 2013; Chomsky 1996) – they evince a special, ampliative form of abductive inference. For example, alongside the many rules and facts about the world that are explicitly taught, human children learn a large and stable set of implicit beliefs that govern action without needing to be stated explicitly, described, or explained (Sperber 1996; 1997). By age 7, children are already proficient in complex, though mostly tacit intergroup relational rules and dynamics of power, and already form implicit judgments about the "value" of members of other groups, and that of their group in relation to others (e.g., children of minority groups often internalize preferences for prestige-laden groups different from their own ethnic group; for a review, see Clark 1988; Clark & Clark 1939; Huneman & Machery 2015; Kelly et al. 2010; Kinzler & Spelke 2011; Machery & Faucher 2017; Navarrete & Fessler 2005; Pauker et al. 2016).

Clearly, we are continuously immersed in culturally shaped environments and interactions from before birth. Despite advances in developmental psychology (Csibra & Gergely 2009; Tomasello 2014) and cognitive anthropology (Boyd and Richerson 2005), we still lack a formal account of the mechanisms of enculturation. The processes that enable implicit cultural habits and norms to arise from inference and imitation, and to be learned and maintained with a high degree of precision and reliability across large-scale sociocultural phenomena, involving multiple interlocking minds and institutional structures, are only partly understood. This is our puzzle.

1.2. The theory of mind debates

In this article, we will propose a solution to the puzzle of implicit cultural learning. We present a model of the ability to perform inferences about the shared beliefs that underwrite social norms and patterned cultural practices derived from first principles. In helping to solve the puzzle of the *implicit* acquisition of culture, our model provides an integrative view of what has variously been called *mind reading*, perspective taking, joint intentionality, *folk psychology*, *mentalizing*, or *theory of mind* (TOM) – in short, the human ability to ascribe mental states, intentions, and feelings to other human agents and to oneself. To simplify, we will use the term TOM to refer to this ability. Of pertinence to our argument here, TOM (in its various theoretical formulations) is generally described as a key mechanism underwriting the human capacity to form joint goals leading to cultural forms of life (Tomasello 2009).

As a generative framework, TOM has been the subject of sometimes fierce and still ongoing debate in cognitive science

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

(Michael et al. 2014; for a comprehensive review, see Heyes & Frith 2014). Historically, much of the debate has occurred between three camps that have advanced alternative explanations for the human ability to infer the mental states of others – namely, the theory theory (TT), simulation theory (ST), and embodied cognition (EC) accounts.

Whether one considers the debate settled depends on one's disciplinary and theoretical position. Outside of the field of developmental psychology, which seems to have adopted some arguments from embodied cognition in favour of an enriched TT account, philosophers in the enactivist camp – and, to different extents, anthropologists – still disagree with the mainstream "cognitivist" psychological account of TOM.

Revisiting the TOM debate from the perspective of cognitive and evolutionary anthropology is helpful to contextualise current critiques (e.g., Christensen & Michael 2016; Michael et al. 2014). These critiques stress the importance of considering culture-specific, embodied, and shared interactions with the environment, over the manipulation of internal representations about other minds (reviewed in sections 1.2.1–1.2.3). Beyond extending debates in the philosophy of mind, the arguments here will be helpful to anthropologists – who are today, attributable in part to the popularity of the so-called ontological turn (e.g., De Castro 2009), largely committed to anti-cognitivist accounts – and psychologists, who largely fail to consider the extent to which cognition is "collective."

The basic idea behind TT is that human agents acquire knowledge about the ways in which mental states should be ascribed, which takes the form of a (literal) theory of how minds operate (Carruthers & Smith 1996; Gopnik & Wellman 2012). Proponents of TT hold that social coordination and social cognition require the capacity to make inferences about other people's mental states and propositional attitudes *as such* (i.e., an ability to explicitly formulate to oneself that others also think "silently," that they may hold beliefs that are true or false, and that there may be a difference between their stated and true intentions, beliefs, or needs – the ability, in other words, to hold a folk theory about other people's minds).

According to a large body of related critiques in the social sciences and phenomenological philosophy, the TT account fails to describe a species-wide mechanism on several counts:

1. TT is a construct derived from Western contexts and fails to describe universal human mechanisms – we call this the *cross-cultural critique*.
2. TT is a dualistic cognitivist construct and thus fails to account for the embodied nature of cognition – we call this the *embodiment critique*.
3. TT is committed to a Machiavellian view of the evolution of cognition that fails to account for the cooperative nature of cognition and behaviour – we call this the *cooperativity critique*.

1.2.1. The cross-cultural critique

For many anthropologists, the TT account reflects a culture-bound, historically specific notion of "mind" and the person that is biased towards individualistic Western folk models popularized by enlightenment philosophers (e.g., Locke's notion of personhood as psychological interiority, Cartesian mind-body dualism, and Kant's notion of phenomenal reality and selfhood). Critics in this camp point out that in many non-Western cultures, folk reasoning about human action does not emphasize individuals' intentions or mental states (Astuti & Bloch 2015; Duranti 2015; Geertz 1973; Keane 2015; Luhrmann 2011; Rosaldo 1982).

Instead, actions may be explained in terms of their perlocutionary effects – that is, in terms of their purported consequences according to locally relevant norms, such as "what would upset the ancestors" (Astuti & Bloch 2015). Extreme versions of this claim have pointed to ethnographic examples from a group of primarily Melanesian cultures described as having a folk psychology characterized by an "opacity of mind" in which the notion of mental states and psychological interiority is reportedly absent (Ramsey 2007; Robbins & Rumsey 2008).

Recent reviews of this controversy, however, noted that there is no experimental evidence to verify whether and how Melanesians make inferences about others' mental states based on others' behaviour (Robbins et al. 2011), while a close reading of the ethnographic record suggests that folk notions of opacity are *normative* rather than descriptive. This is suggested by ethnographic reports of children being reprimanded for overt curiosity about others' actions or intentions. On this view, Melanesians are simply taught that they *ought not to* wonder about what people are thinking (Robbins 2008; Robbins & Rumsey 2008; Rumsey 2013). Moreover, reports from other Melanesian contexts indicate that it is widely recognized that people "think silently" (e.g., in the context of courtship among the Korowai of New Guinea; Luhrmann 2011; Stasch 2009).

Although the current balance of evidence does not support critiques that TT describes a process that is exclusively found in Western cultural contexts, ethnographic studies document wide variation in the ways that people inquire into and talk about others' states of mind that must be accommodated by any account of TOM.

1.2.2. The embodiment critique

Philosophers and psychologists in the embodied cognition camp have also objected to the TT account on the grounds that understanding others or responding to social cues is characterized by "quick," "intuitive," "embodied" responses that need not entail interpretations about other minds or any notion of mental states (Michael et al. 2014). "Some of these critics of TT have proposed an alternative approach based on the idea that, rather than mobilizing an explicit theory to ascribe mental states to others, human agents use their own experiences and intuitions to understand other human agents through a process of 'simulation' - other people's propositional attitudes are on this view 'simulated' from one's own mental experience, but are not 'theorized' as such" (Goldman 2006). On the view of such simulation theories (ST), TOM abilities involve processes of modelling others' actions, which may be embodied and automatic (Gallese & Goldman 1998). Embodied cognition need not involve anything that looks like a theory because it uses bodily sensorimotor systems to provide analogical models of human motivation, intention, and action (Shapiro 2010).

Radical enactivist cognitive science takes this emphasis on embodied cognition further to argue that basic cognition does not entail any kind of mental content – particularly not about others' mental states and propositional attitudes (Hutto & Myin 2013). In more recent accounts (Hutto & Myin 2017; Hutto & Satne 2015), enactivists grant the existence of explicit inferences about others, but only in situations that are developmentally contingent on language. Learning to make explicit ascriptions is then a separate, later, developmentally achieved result of narrative practices (Hutto 2012).

As Heyes and Frith (2014) point out, some current accounts have adopted a compromise position, which gives credence to both sides of the debate, through recognizing multiple processes and progressive elaboration over development. In Apperly and Butterfill's (2009) two-systems model, for example, most social cognition may be largely automatic, while a process akin to TT

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

may underpin specific types of language-dependent inferences. Apperly and Butterfill's account stemmed from a growing consensus in cognitive science – famously exemplified in Daniel Kahneman's *Thinking Fast and Slow* (2011) – that cognition can be divided into two "systems": one evolutionarily old, innate, implicit, "cheap" automatic system of informational foraging supported by a series of largely social biases, and a developmentally older, evolutionarily young, effortful, relatively inefficient modality of volitional, voluntary reflection. Apperly and Butterfill proposed that the distinction between TT and ST could be cast along this spectrum, with explicit mentalizing about others entailing a situationally specific, relatively rare sort of reflexivity acquired later in developmental.

Others still have proposed a "multi-system," progressive scaffolding of socio-cognitive inferences ranging from the fully automatic to the effortfully explicit (Michael et al. 2014). These later "interactionist" models offer a more nuanced and dynamic account of the gradients of inferences, which, rather than being "located" in discrete cognitive systems, likely occur on a continuum of attunement to different statistical regularities. This is a point elaborated on in detail in Hugo Mercier and Dan Sperber's *Enigma of Reason* (2017), in which they also recast so-called system 2 reflexivity as varieties of automatic inference *about other's inferences* triggered by communicative cues – actual or imaginary (e.g., in engaging in, or mentally rehearsing, conversation and interaction with others). Crucially, these recent models (two systems, multi-systems, and interactionist) all study the manner in which agents optimize the metabolic cost of cognition by tuning attentional preference to different domains of statistical regularities, emphasizing the function of social and cultural modulations of automaticity. These models, as we argue in section 1.3, lend themselves to a culturally informed free-energy principle (FEP) model.

1.2.3. The cooperativity critique

TOM has played a key role in evolutionary psychology. Early accounts of evolutionary psychology described the evolution of human intelligence and TOM abilities by appealing to the so-called Machiavellian intelligence hypothesis (Dunbar 2003; Gavrilets & Vose 2006; Pinker 1999; Trivers 2000). On this view, the ability to rightly infer others' mental states – human mind reading – and propositional attitudes about others' mental states evolved through a cognitive arms race between cheaters (who need to understand others so as to deceive them) and cheater detectors (who need to understand others to detect deception).

In contrast, scholars in the mutualist camp (Henrich 2015; Tomasello 2014) contend that individual human fitness is best maximized by cooperation with others, leading to an evolved preference for promoting group fitness through the cooperative division of labour. Such cooperation requires knowledge of others' states of mind or intentions. In support of these views, natural pedagogy (Csibra & Gergely 2009; 2011), interactionist (Mercier & Sperber 2017), and other cultural intelligence paradigms have emphasized the evolved propensity for a non-Machiavellian, cooperative division of cognitive labour, in which mind reading evolved for the purpose of *outsourcing contextually relevant information* to specific others from our in-groups and to leverage knowledge, skills, and attitudes from a cumulative cultural repertoire. In more radical versions of mutualist models, such as Hrdy's cooperative breeding hypothesis (Burkart et al. 2009; Hrdy 2011), mind reading is thought to have evolved in the pre–*Homo sapiens* lineage as a result of a "cuteness and care" arms race, because selection favoured individuals who were, at once, good caregivers and good at eliciting care from others.

Heyes and Frith (2014) have proposed an account of the cultural co-evolutionary elaboration of TOM abilities, suggesting that the internalist, brain-centred accounts provided by proponents of TT and ST need to be augmented by an account of how cultural evolution and cultural inheritance sculpt an innate mind reading "start-up kit," in ways that are analogous to how cultural practices of reading harnessed an evolutionarily older linguistic "start-up kit" (Dehaene & Cohen 2007).

The extent to which the evolution of perspective-taking abilities requires mental content about other minds is still hotly debated. In the *mind-shaping* hypothesis (Mameli 2001; Zawidzki 2008; 2013), for example, mind reading likely emerges from an evolutionarily older and developmentally earlier capacity to imitate, learn, teach, and directly influence others. Nevertheless, current work suggests that the ability to engage with others as agents with interior states and intentions is central to the cooperative forms of social life we call "culture."

1.3. Piecing together the puzzle of implicit learning: A new portrait of TOM

1.3.1. Conceptualization

The cultural, embodiment, and cooperative critiques of TOM emphasize either *internal* cognitive processes of theory building or simulation or *external*, social-cultural processes of interaction and cooperation. Clearly, these are differences in emphasis, and a more complete picture must show how they fit together.

In this article, we complete this picture by proposing a model of implicit cultural learning that we call "thinking through other minds" (TTOM). In recognizing the virtues (and limitations) of both internalist and externalist accounts, the TTOM model proposes a resolution of the dialectic – and false dichotomy – between so-called internalist (TT and ST) and externalist (mutualist, interactionist, and cultural evolutionist) positions.

TTOM integrates a number of recent approaches to the study of cognition – in particular, the cultural intelligence hypothesis in evolutionary anthropology (Boyer 2018; Henrich 2015; Tomasello 2014), the niche construction perspective in evolutionary biology (Laland et al. 2015; Odling-Smee et al. 2003), the interactionist approach to the evolution of reasoning in cognitive science (Mercier & Sperber 2017), and the sociocultural enactivist approach to mind reading (Fabry 2018; Gallagher 2017; Gallagher & Allen 2018; Hutto 2012; Hutto et al. 2014).

1.3.2. What the variational model affords

At a formal level, we integrate these approaches within the framework of the variational free-energy principle (FEP; Friston 2005; 2010) in theoretical neuroscience and biology. Framing this integration in terms of the FEP allows us to derive, from first principles, an interactional model that can explain the acquisition, production, and stabilization of cultural expectations (Friston 2013; Friston & Stephan 2007; Ramstead et al. 2018). See Box 1.

We will argue from the formal perspective of embodied (i.e., active) inference, which rests upon our species' remarkable capacity to infer or assign conspecifics to some pragmatic (i.e., prosocial) categories. A successful inference about the "sort of person you are" enables a host of conditional inferences, many of which have a direct bearing on "how I should behave." This is particularly true if I infer that "you are like me." We will unpack this view

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Box 1. The formal structure of the FEP model adds significantly to the general approach we outline in this article in two ways.

1. *Conceptually*, the FEP provides us with an explanation *from first principles* of the processes involved in, and the adaptive value of, implicit cultural learning and mind-reading abilities. It gives us a formal grip on the underlying dynamics of these two phenomena (for a schematic overview, see Figs. 1–4 and the mathematical appendix). The main challenge confronting TTOM is that of making sense of the dynamics involved when agents learn domains of socially relevant expectations – that are involved in the acquisition of culture – and how these domains are scaffolded from joint intentionality, basic perspective-taking abilities, and evolved attentional dispositions for learning from and through others. These domains are internal (e.g., neural scale) and external (environmental scale) to individual agents. Without a formal apparatus, it is difficult to make sense of these multiscale learning dynamics or to examine how they interact. We employ the FEP to formulate TTOM for the simple reason that it is, to our knowledge, the *only* theory that has produced formal models (supported by computer simulations) of many of the cognitive mechanisms involved in the learning dynamics of TTOM, including, for example, action, perception, learning and attention (Friston et al. 2016), visual foraging (Mirza et al. 2016), communication (Friston & Frith 2015b), decision making (Friston et al. 2014a), planning and navigation (Kaplan & Friston 2018), emotions (Joffily & Coricelli 2013), curiosity and insights (Friston et al. 2017b), and niche construction (Bruineberg et al. 2018b; Constant et al. 2018b).
2. *Empirically*, the FEP offers a set of equations that can be used to develop computational models of data acquired in studies of social interaction, in which implicit cultural learning and mind reading are at play. These models can then be used to identify new dynamics and make predictions that can, in turn, be tested in real-world situations. The scope of the current argument is limited to discussing the theoretical relevance of the FEP. That said, we can indicate candidate tasks to produce data amenable to FEP modelling. Notably, the different variants of two-person psychophysiology in social interaction studies (e.g., Bolis & Schilbach 2018a; Bolis et al. 2017; Schilbach 2016; Timmerman et al. 2012; von der Lühe et al. 2016) are target modelling candidates, as they already rely on core principles of active inference and involve the manipulation of what we call "epistemic resources."

with a special focus on epistemic action, via the selective patterning of salience and attention – and how this is mediated via cultural affordances. We hope to show that these epistemic resources arise naturally from cultural niche construction when, and only when, I share an environment with other "creatures like me."

The formalism of the FEP allows us to take further steps towards operationalizing the process of implicit cultural learning and mind reading that we describe as thinking through other minds (see Box 2). In brief, the set of equations that model the process of TTOM could be implemented in computational models, to study simulations of, for example, psychophysical, neuronal, and behavioural measurements of the processes involved in a mind-reading or cultural learning task.

On the one hand, such simulations would allow researchers to generate hypotheses about mind reading and cultural learning that may be tested with other empirical methods. On the other hand, FEP simulations can be employed to replicate in vivo experiments (e.g., Kiebel & Friston 2011; Schwartenbeck & Friston 2016). One can then use the model to explore the dynamic consequences of changes in parameters associated with the causal factors that led to the generation of the experimental outcomes that were studied empirically. With this method, one also might identify potential contributors to pathological and healthy responses to the task by manipulating the parameters and generating new simulated psychophysical, neural, and behavioural measurements based on the model that has been fitted with in vivo data (e.g., Cullen et al. 2018).

1.3.3. Outline of the argument

Section 2 of this article introduces the notions of expectations and cultural affordances. We describe *shared attention* and *evolved attentional biases* as crucial mechanisms for engaging with and stabilizing sociocultural niches. We describe the *selective patterning of salience and attention* as the main process behind enculturation, which in turn enables the engagement of human agents with the sets of possible actions (or cultural affordances) that make up their local world (Ramstead et al. 2016).

Section 3 presents our solution to the puzzle of implicit cultural learning. Human beings acquire the shared habits, norms, and expectations that constitute their culture through their immersive engagement within specific cultural practices, which we call *regimes of attention* (Veissière 2016). Regimes of attention mark off certain contextually adequate actions as especially salient and help agents learn to respond to the norms and resources of their local cultural niche. The most important of these resources are the *epistemic resources* that indicate salient information deemed relevant and reliable (Bertolotti & Magnani 2017; Clark 2006; Pinker 2003; Whiten & Erdal 2012).

As we elaborate through the notion of *epistemic authority*, we show that humans are typically biased towards the *source* rather than the *content* of information (Mercier & Sperber 2017). As amply documented in the literature on so-called cognitive errors (Kahneman 2011), this tendency can also direct humans towards low-quality, but otherwise high-fidelity, information, particularly when it can be intuitively associated with social proof and other mechanisms of social influence (Cialdini & Goldstein 2004). We identify the prestige bias in particular (Henrich & Gil-White 2001) as a central attentional mechanism in the mediation of salience for humans.

The notion of *salience* understood as expected information gain is a central theme of the FEP (Friston et al. 2016; Kaplan & Friston 2018; Parr & Friston 2017a; 2017b). Recent FEP-based models of cognition in context cast niche construction behaviour as the process whereby organisms "outsource" the computation of salience to statistical structures of the physical environment. The environmental niche then registers information about salience (what an organism trusts or preferentially attends to for it will lead to information gain).

This information corresponds to epistemic resources of the niche (Bruineberg et al. 2018b; Constant et al. 2018a; 2018b). Niche construction allows the scaffolding of complex networks of shared expectations encoded across brains, bodies, constructed environments, and other agents, which modulate attention, guide action, and entail the learning of patterned behaviours. Human niches are fundamentally social and cultural – built and

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

constituted by interactions with other people. In the general human niche or any local sub-niche, behaviour is to a large extent *culturally* patterned. Hence, in addition to (and, as we will argue, often prior to) observable statistical regularities in external states of the world, human behaviour is patterned through expectations about *what other people also expect of the world.* It is this domain of expectations about salience and the process of leveraging these expectations that we call "thinking through other minds" (TTOM).

The processes that make up TTOM extend from the conventionalized, normative behaviour of enculturated individual agents (e.g., stopping at a red traffic light), which only in some cases requires making inferences about agents, to cases that require bona fide inferences about others' mental states for proper (i.e., situationally appropriate) modes of engagement.

Section 4 of this article shows how TTOM integrates standard TOM approaches to tackle the cultural, embodiment, and cooperative critiques. TTOM argues for a compromise position between *internalist*, brain-based approaches (e.g., simulation and theory-theory theories), which emphasize the neural machinery in individual human brains that is necessary to read other minds, and *externalist* approaches (e.g., radical enactive and cultural evolutionary theory). Indeed, one of the main motivations for the FEP is to capture the two-way traffic between the organism and the world, to emphasize both the enactment of shared cultural expectations and norms, and the brain-based cognitive abilities that make such an enactment possible, adaptive, and situationally appropriate. Under the FEP, there is no justification for any strict distinction between *dynamics* (as emphasized by externalists) and *inference* (the focus of internalist models).

The conclusion discusses the implications of this model for future research on enculturation and the cultural shaping of cognition in health and illness.

2. Expectations and cultural affordances

In this section, we show that human agents learn most of their expectations through the selective patterning of attention, based on immersive participation in cultural practices. At the outset, we should define what we mean by "expectations."[2] We use the term to describe a rich *repertoire* or *spectrum* of priors or beliefs that reflect action readiness, which ranges from the fully automatic to the effortfully deliberate. Our concept of expectation describes the patterns of action readiness that modulate and direct the adaptive action of agents; it is therefore very broad in its applicability and ranges from the implicit, embodied expectations that we enact continuously, often without noticing, to the more consciously held, effortful, psychologically contentful expectations that characterize enculturated human consciousness.

2.1. The concept of expectation

On the more automatic end of the spectrum, we can speak of expectations when one's stomach prepares a digestive response upon expecting that food is coming from mastication, or when one's hand and arm prepare an adequate muscle response to lift a half-full glass of wine. Each of these processes reflects different kinds or levels of prior engagement of the world, across different timescales, which include evolutionarily old dispositions common to all vertebrates that have been exapted for new uses, as well as distinctive developmental experiences, and learning histories. Together, these elicit physiological, bodily, and emotional orientations towards the possibilities for action available in a specific context. Immersion in cultural contexts, moreover, will *structure* such low-level expectations *through participation in patterned cultural practices* (e.g., contextually patterned modes of affect associated with specific kinds of food and drink and ritual contexts of consumption).

Human expectations, thus, are always scaffolded through "levels" (or scales) of evolutionary and developmentally inscribed prior dispositions that come to be modulated by higher-level symbolic conventions (Kirmayer & Ramstead 2017). The intuitive distrust of other people symbolically marked as belonging to an out-group, for example, has been shown to recruit evolutionarily old disgust responses (Phillips et al. 1997; Rozin et al. 2009; Tybur et al. 2013). This involves another level of implicit "expectations" in which evolutionarily old threat and poison-detection dispositions are activated by (differently implicit) symbolic conventions or affordances.

At the other end of the spectrum, many of the expectations that guide behaviour are explicitly taught, effortfully learned, and can be reflected upon (e.g., "sit up straight," "do not fidget in class"). Such expectations, however, are also more difficult to learn, and least likely to become fully patterned. Indeed, one may sit badly most of the time, fidget in class despite my embarrassment, and face disappointment when one's daughter chooses to become an engineer. Later developing forms of explicit inference require abstract thought, formal instruction, and perhaps deliberation to learn; but once the agent is properly enculturated, new practices usually can be figured out without the direct presence or instruction of other agents. The learner learns the metacognitive strategy of how to access, offload, and work with conventional forms of presented cultural knowledge (Heyes 2018b). This process, however, will generally entail different modes of indirect social learning, for example, from instructional codes devised by others (such as learning a cooking skill from a written guide or YouTube video).

Examining these processes of acquiring conventional or normative behaviours, social scientists have pointed to the important difference between *dogma* (official doctrine) and *doxa* (common belief; Bourdieu 1977). The explicit rules and conventions established in dogma (what people know they must do) and reported in everyday speech are poor indicators of the regularities of a culture – and how humans learn cultural behaviour in general. *Doxa*, in Pierre Bourdieu's famous formulation, refers to *all that is taken for granted* in any given context or society. For example, in his "dramaturgical" account of social life, sociologist Erving Goffman (2009) describes the gradients of effort and explicit performance required in the obedience to and enactment of social conventions in everyday life. Goffman notes that in some spaces (such as the home), which are symbolically marked as the "backstage," people tend to relax their effortful behaviour and ignore or disobey many social rules; they trade off the dogma for the doxa. Nevertheless, their behaviours necessarily draw from the culturally shaped repertoire of normative and conventional forms.

What interests us here is how the doxa of backstage behaviour (indeed most of solitary cognition) is itself already culturally patterned, despite the immediate absence of others' enforcing gaze (and the foregrounding of inferences we make about what others know and expect in context). A first hint is the fact that human agents are constantly (deliberately or automatically) adjusting what they are doing to what relevant others (e.g., role models or anti-role models, specific or generalized) *expect*, and *expect*

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

them to expect, and so on. Much of this is accomplished implicitly (Tomasello et al. 2005), usually through nonverbal communication with gesture, facial expression, posture, and pantomime, but also through language when necessary. Evidence that this kind of expectation does not depend on language comes from the observation that infants as young as 15 months are able to make *implicit inferences* about others' mental states (Onishi & Baillargeon 2005) and actions well before they can formulate *explicit statements* to this effect (Michael et al. 2014).

2.2. The concept of affordance

In Gibson's ecological approach to perception (Gibson 1979), things and features of the world are said to *afford* possibilities for engagement (Chemero 2009; van Dijk & Rietveld 2017). An affordance is a relation between an agent's abilities and the physical states of its environment. For example, water affords drinking, cups afford drinking out of, bridges afford crossing, axes afford cutting, handles afford holding, and so on. Affordances are defined in terms of physical properties of the thing in the world (e.g., being graspable, being able to support the weight of a person) and in terms of the *abilities* or expectations of the agent (e.g., knowing how to sit straight). Abilities can be described in terms of the spectrum of *expectations* with which the agent is endowed (Gibson 1979; Pezzulo & Cisek 2016; Rietveld & Kiverstein 2014; Tschacher & Haken 2007). It takes an agent with a mouth, throat, stomach, and so on (to drink); and hands and opposable thumbs (to grasp a cup); and a certain set of skills (hand-eye coordination, for example) to be able to "discover" the relationship of water and cups to the action of drinking.

The relation of affordances to the notion of expectations is a recent extension of the ecological approach that explains perception as conditioned on the beliefs of the agent (Bruineberg & Rietveld 2014; Chemero 2009). Hence, affordances are not simply static features of the environment, independent of the presence and engagement of an agent, nor are they states of the cognitive agent alone. Affordances are "invariant variables" or structures of relatedness (Gibson 1979, p. 134). In the case of sensorimotor affordances, for example, they are invariant, in that they are grounded in the physics and geometry of the agent's interaction with the environment, which results in relationships that are highly reliable and stable across time and are ready to be perceived or (re)discovered by the agent; and they are variable, in that they are specified dynamically by the sensorimotor and other cognitive abilities of the agent. In the case of affective affordances and expectations, the stability may reside in the neurobiology of organisms' learning and memory systems coupled with the persistence of the environmental cues to which particular patterns of recollection and enactment have become linked. The relational space of possibilities between agents and their environments constitutes an *ecological niche*. Agents and their environments are modified, and become attuned to each other, as the result of their history of co-adaptive interactions (Bruineberg & Rietveld 2014; Gibson 1979).

These examples are congruent with work on the evolution and cultural learning of tool use (Stout & Chaminade 2007; Stout et al. 2008), which illustrates the need for humans to learn to hierarchically structure actions with long-term consequences. "Hierarchical" here means that actions are nested within one another, and that complex behaviours require planning a whole chain of nested actions, not just the immediate optimization of current actions or a simple sequence. This kind of executive control of behaviours is characteristic of *enculturation*, in which complex sequences of action are built out of iterative structures of simpler components strung together in ways that reflect the results of collective experiences of trial and error. An individual is therefore able to borrow from and integrate the experimentation and learning of others in the cultural group.

Direct or "natural" affordances in the humanly constructed ("anthropogenic") environment can be supplemented, modified, or supplanted by "conventional" affordances (Ramstead et al. 2016), which depend on shared cultural conventions, based on skills learned through immersive social practices. Hence, bodies of water ("naturally") afford drowning for all humans, and swimming for those with the acquired skills that allow them access to that specific *cultural* affordance. Mastering swimming, like all cultural affordances and most of what humans do and think, requires immersive participation (Hutto 2012; Roepstorff et al. 2010), which includes imitation, practice, repetition, and a grasp of norms and conventions. Hence, affordances are contextually sensitive. For example, for the right kind of agent, a formal suit and tie might function as a cue that indicates authority and affords deference; but when additional cues are added (e.g., a napkin draped over the forearm and a silver tray with glasses), the affordances will change whose enculturation enables them to respond appropriately to the cues.

2.3. Learning cultural affordances

How are the affordances of the niche learned? What does it mean to learn to recognize and engage a specific field of affordances? This is a puzzle, because affordance theory tends to collapse basic categories of learning like "knowing how" and "knowing that." For example, there is no necessary precedence of the knowing "that" a cup is for drinking over the knowing of "how" to drink from a cup, and vice versa. Even in domains where knowing "that" seems to precede knowing "how," such a distinction does not hold, because knowing "that" is leveraged as a skill interiorized and integrated to normal implicit motor practice – for example, architectural design (Rietveld & Brouwers 2017) and mathematical thinking (Menary 2010). Put simply, knowing "that" is only knowing "that" when it becomes know "how," and acquiring know "how" requires interiorizing and embodying know "that." This circularity can be understood through a process of scaffolding that occurs on multiple temporal scales associated with the cultural co-evolution of particular niches, communities, or traditions; the developmental trajectory of individuals; and the process of learning to engage with new social contexts.

What, then, are the underpinnings of scaffolding? Some anthropologists, like Tim Ingold, have argued that human niches comprise affordances that can be figured out, rediscovered, or rebuilt by human individuals in each generation without the "transmission" of a purportedly separate realm of "cultural representations" (Ingold 2001). Critics of Ingold (e.g., Howes 2011) have pointed out that most of what humans learn over their life spans in order to become proficient at functioning in their local worlds, is learned *socially* – that is to say, learned primarily from other humans, and not just from what things or situations themselves afford. However, Ingold maintains that many aspects of human life are simply emulated (Hamilton 2008), "shown," or "pointed to," and left to be explored, "figured out," and experimented with by individual learners (for example, in play).

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

The main role of others in this kind of social learning is to direct attention rather than to convey specific semantic content (Tomasello 2014). In effect, social learning involves immersion in local contexts through what we call *regimes of attention* and imitation that direct human agents to engage differentially in forms of shared intentionality. We have argued that such regimes of attention play a central role in the enculturation of human agents (Ramstead et al. 2016). Indeed, human beings seem particularly specialized for such forms of social learning (Sterelny 2012).

Humans mostly learn deictically (in context) and pragmatically by participating in cultural practices and by being immersed in the ways of doing things that characterize a given local culture. Some of this involves following the "tracks" laid down in local environments by others, or following the norms and rules presented through institutions, without engaging with others' interiority. But many convention-dependent forms of learning require inferences based on prior knowledge about how we expect others to think and behave in specific settings (e.g., adjusting to culturally specific turn-taking rituals in public space; Ramstead et al. 2016).

The process of learning how to engage cultural affordances to think through other minds likely begins in infancy when we seek or accept guidance from our caregivers, and it further develops through exposure to social hierarchies of prestige, themselves embodied in kinds of high-status agents that can be leveraged as models (Feinman 1982), which are knowledgeable or skilful in-group members, educators, community and religious leaders, celebrities, and imaginative reconstructions of folk or historical personages with high epistemic prestige (e.g., "What would Wittgenstein think of this theory?"). Individual action, in turn, is guided by what agents expect relevant agents to expect of them ("What would mother expect me to do?").

Others in our social world present us with cultural affordances, as well as solicitations, for action. Engagement with these realizes a specific social niche, context, group, or community. The reliance on social and cultural affordances co-constructed with and maintained by other people makes it important for us to distinguish between those who think like us and those whose thinking is either systematically different from our own or else unfamiliar and, hence, unpredictable – and inherently surprising. This distinction marks off domains of in-group and out-group, with corresponding epistemic authority. Regimes of attention then make the right kinds of social solicitations stand out in context, thereby allowing the learning of socially relevant affordances in a given cultural niche, community, or local world.

2.4. The phylogeny and ontogeny of cultural affordances

In human ontogeny, it is likely that affordances are first learned implicitly, automatically, and with little conscious effort, through imitation, repetition, and rewards. Phylogenetically, the human mind evolved to support a series of adaptive "content biases" (Henrich 2015) for features of the world that possess high intrinsic learnability, and feed-forward potential through teachability and memorability. Fire, edible foods, and simple tools, for example, all have been amply documented as possessing these heuristic properties (Henrich 2015). In the realm of more conventional affordances, compared with other primates, humans are also unusually adept at tracking other agents' social status and shifts in symbolically assigned prestige through gossip (Dunbar 2004; Henrich & Gil-White 2001).

Status among social animals generally provides a guide for whom to follow and obey, and from whom or what to learn. As cultural evolutionists have pointed out (Henrich & Gil-White 2001; Mercier & Sperber 2017), social status among humans serves a primarily *epistemic* function. One seeks guides for thought, behaviour, and affect in agents who embody sources of relevant cultural information that are deemed to be of high quality in relevant social contexts (e.g., we learn from professors in the classroom and seek help from good students, or we seek to publish in high-impact journals). Among humans, symbolically conferred *prestige* has largely replaced sheer physical dominance as a way to find, acquire, and signal status (Henrich 2015). In social context, marks of distinctions (Bourdieu 1984) such as styles of dress, forms of speech, and other techniques of the body provide a shortcut that signal an agent's status on the various prestige scales deemed relevant. *Gossip*, in turn, serves the more fine-grained communicative function of keeping track of an agent's conferred prestige and epistemic status.

The aforementioned mechanisms rely on evolved cognitive biases for cultural transmission that have been hypothesised to serve an information-tracking function (Henrich 2015) – that is, as enabling humans to outsource their decision making to other *agents*, through patterned *interactions* with them and the *shared places* in which they dwell. The physical structure of the environment – including artefacts, practices, and other socially constructed aspects of the ecological niche – embody or encode adaptive, context-relevant cultural information endowed with *salience* (i.e., as high-quality or "useful" sources of information in context). A dramatic illustration of this is provided by the infamous Milgram experiments (Milgram 1963), which demonstrated the extent to which human agents are ready to outsource their actions to those that symbolically display the right credentials and wield epistemic authority.

Social status serves the epistemic function of locating the person in a locally relevant hierarchy – a process that can also be described in terms of affordances as prestigious agents solicit imitation through such perceived qualities as *trustworthiness* (Mercier & Sperber 2017) and *credibility* (Henrich 2015). How well or badly agents respond to such *affordances* – as indexed through gossip (e.g., circulating stories about cheating spouses, embezzling chiefs, or free-riding subordinates) – thus will largely determine the levels of trust that they inspire in others. Furthermore, the hierarchy that locates the person is not only material, but also symbolic, as expressed through historically acquired and socially displayed marks of distinction. This poses a challenge to an account of affordances in terms of immediately present features.

Humans are accustomed to attending to certain people, in certain places for tones of voice, facial expressions, shifts in body posture, and so on, which signal approbation, disapproval, or moral concern and hence convey (in context) normative information (Ignatow 2009; Williams 2011). As we have seen, beyond what they naturally afford, human material environments have additional, *symbolically inscribed* normative and deontic powers that deeply permeate the way that individuals affectively approach and engage with their niches (Kaufmann & Clément 2014). For example, in the European Middle Ages, children may have been socialized to fear forests as dark and dangerous spaces full of beasts, witches, and evil spirits through folktales and bedtime stories. In contrast, in many hunter-gatherer cultures, like the Aka of Central Africa, children are equipped with cultural knowledge to expect the forest to offer a safe, nurturing space (Hewlett 1994; 2017).

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

The physical environments occupied by various human groups and sub-groups also characterize *group-specific* affordances (e.g., a neighbourhood or a city; Einarsson & Ziemke 2017). Consider how a space (e.g., a university or museum) that is symbolically marked with group-general standards of prestige – a space, thus, that has been historically inaccessible to low-status individuals – will afford radically different experiences to high- and low-status individuals depending on how their respective sub-group is valorized in their macro-cultural niche. Pierre Bourdieu's concept of *habitus* (as the internalization of social norms in techniques of the body) is one way of approaching the varying effects of a socio-cultural niche on individuals with different status or position. To expand on Bourdieu's (1977) reflections on the effects of cultural capital on *habitus*, we note that a similar space can be marked as "welcoming" for some, but as "intimidating" or outright "hostile" to others (e.g., for minority groups). This reflects a related, orthogonal distinction between the familiar (predictable) versus the unfamiliar (unpredictable). From a cultural affordances perspective, being socially marked and positioned at a particular place in a cultural niche enables automatic responses in one's patterns of movement, posture, breathing, and gaze, as well as in neurobiological responses, such as fluctuations in cortisol (Bijleveld et al. 2012), oxytocin (Hrdy 2011; Luo et al. 2015), or testosterone (Cheng et al. 2013).

The co-existence of habitus or internal physiological dispositions with external features of an adaptive niche points to a crucial feature of affordance theory – namely, that the affordances of the environment and the capacities of an individual are inextricably interwoven, and co-determining. However, developmentally, and in shared social contexts, culture precedes individual action and experience. In a sense, culture confers on the environment latent affordances such that, if one learns the right repertoire of skills (including attentional strategies) from one's forebears (by acquiring specific cultural knowledge and practices), one can "read" the environment in new ways, thereby discovering "new" affordances (that were, in a sense, there all along, insofar as they engaged other or prior skilled actors). Moreover, because one of the functions of cultural affordances is to allow improvisation (and hence the creation of new cultural forms), the affordances of a niche that are being actively engaged are always in the process of discovery, elaboration, and extension. Clarifying the temporal move from group or cooperative affordances to individual ones (and back) is part of explaining developmental enculturation, skill acquisition, and culture production.

So far, we have described regimes of attention and symbolic layering as cultural affordances of the *conventional* and *normative* variety. Over the course of human ontogeny, this "conventional" domain of culture eventually becomes superordinate to the natural domain. Past a certain developmental stage, language can be used to install superordinate frames through which subsequent affordances are perceived and engaged (cf. Bengio 2014). This linguistic capacity to leverage affordances can include cooperative behaviours that reflect social norms and cultural forms of life. The statistical regularities exploited in learning cultural affordances, thus, are primarily situated in the realm of *expectations that humans learn to form about other people in the niche* – that is, in the realm of *folk psychology*. We call this intersubjective process of engaging others' expectations and inferences "*thinking through other minds*." In the next section, drawing on the FEP, we turn to the question of how cultural affordances can be acquired and maintained to coordinate large cultural groups, through selective patterns of attention and learning.

3. TTOM: Learning cultural affordances under the free-energy principle

3.1. The free-energy principle as applied to individual cognition

To explain cultural affordances and implicit cultural learning, we draw on the variational free-energy principle. The FEP is a mathematical statement of the fact that living systems act to limit the repertoire of physiological (interoceptive) and perceptual (exteroceptive) states in which they can find themselves (Friston 2013; Friston et al. 2006) (See Box 1). Although even simple organisms have autoregulatory mechanisms to restrict themselves to a limited number of sensory states (compatible with their survival), humans additionally accomplish this feat by leveraging cognitive functions and socioculturally installed behaviour. For example, if core body temperature drops from its usual 37 degrees Celsius, internal processes of shivering are automatically evoked, and externally oriented actions are initiated to move the agent towards a heat source or to put on a jacket or parka.

This requires the agent to learn about the structure of its environment, which, from the point of view of the brain, is not a small business, because the (skull-bound) brain is secluded from the causal regularities in the environment it seeks to learn (Hohwy 2013).

The brain only has direct access to the way its sensory states fluctuate (i.e., sensory input), and not the causes of those inputs, which it must learn to guide adaptive action (Clark 2013a) – where "adaptive" action solicits familiar, unsurprising (interoceptive and exteroceptive) sensations from the world. The brain overcomes this problematic seclusion by matching the statistical organization of its states to the statistical structure of causal regularities in the world. To do so, the brain needs to re-shape itself, self-organizing so as to expect, and be ready to respond with effective action to, patterned changes in its sensory states that correspond to adaptively relevant changes "out there" in the world (Bruineberg & Rietveld 2014). Because action selection and response conforms to such expectations, behaviour can effectively maintain the agent within expected states.

The FEP describes this complex adaptive learning process in terms of variational inference (also called approximate Bayesian inference). Briefly, the idea is that the agent learns a statistical model of sensory causes in the world, called a *generative model*. This model represents the agent's relation to the environment and enables it to predict how sensory inputs are generated, by modelling their causes (including, crucially, the actions of the agent itself).

The generative model underwrites the agent's perception and action as they unfold over time. The parameters of the generative model encode the beliefs of the agent about its relation to the environment (e.g., when I move my finger to flip the switch, the light goes off). This is realized by neural network dynamics that change over short timescales (reflecting external states of the world) and slower changes in network connectivity that encode parameters that change over longer timescales to reflect the contingencies that underlie the agent's representations of the transitions among the states of the world (e.g., the probability of my finger moving the switch to change its state from "down/off" to "up/on"; Kiebel et al. 2008).

The generative model functions as a point of reference in a cyclical (action-perception) process that allows the organism to engage in active inference. Internal states of the agent (e.g., the states of its brain) encode a *recognition* density – that is, a

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

probability distribution or Bayesian belief about the current state of affairs and contingencies causing sensory input. This (posterior) belief is encoded by neuronal activity, synaptic efficacy, and connection strength (Friston 2010). The mathematical formulation behind the FEP claims that all of these internal brain states change in a way to minimize variational free energy. By construction, the variational free energy is always greater than a quantity known as *surprisal*, *self-information*, or, more simply, *surprise* in information theory. This means that minimizing free energy minimizes *surprise*, which can be quantified as the negative logarithm of the probability that "a creature like me" would sample "these sensations."

Crucially, in minimizing free energy, the posterior beliefs encoded by neuronal quantities approximate the true posterior density over the causes of sensations (see Fig. 1 for details). Intuitively, the variational principle of least free energy is just a description of systems (like you and me) that seek out expected sensations. An equivalent and complementary interpretation follows from the fact that surprise is the converse of Bayesian model evidence in statistics. This means that we can understand active inference as gathering sensory evidence for an agent's model of its world – sometimes referred to as self-evidencing.

Put another way, this can take the form of seeking expected sensations associated with novelty or danger (e.g., thrill seeking) or, in more maladaptive cases (e.g., depression), of "confirming" the negative valence of one's world through rumination (Badcock et al. 2017). As we discuss in section 3.3, accounting for novelty seeking in free-energy minimization is an important contribution of the model. On the face of it, humans seem to find a certain kind of surprise desirable. To understand this mathematically, it is useful to appreciate that *expected surprise* (i.e., expected free energy) is *uncertainty* (i.e., entropy). This means that certain acts such as "attending to this" or "looking over there" become attractive if they afford the opportunity to reduce uncertainty. Think of the game of "peek-a-boo" played with infants as a case in point, in which the infant (as learned through repeated practice) attends earnestly in pleasurable anticipation of resolving uncertainty about where her mother will reveal herself. Generally speaking, epistemic affordance of this sort has a positive valence because it entails a reduction of uncertainty, both about states of affairs in the world and "what will happen if I do that."

In summary, the FEP – as applied to individual cognition – describes the process by which an agent updates its (Bayesian)

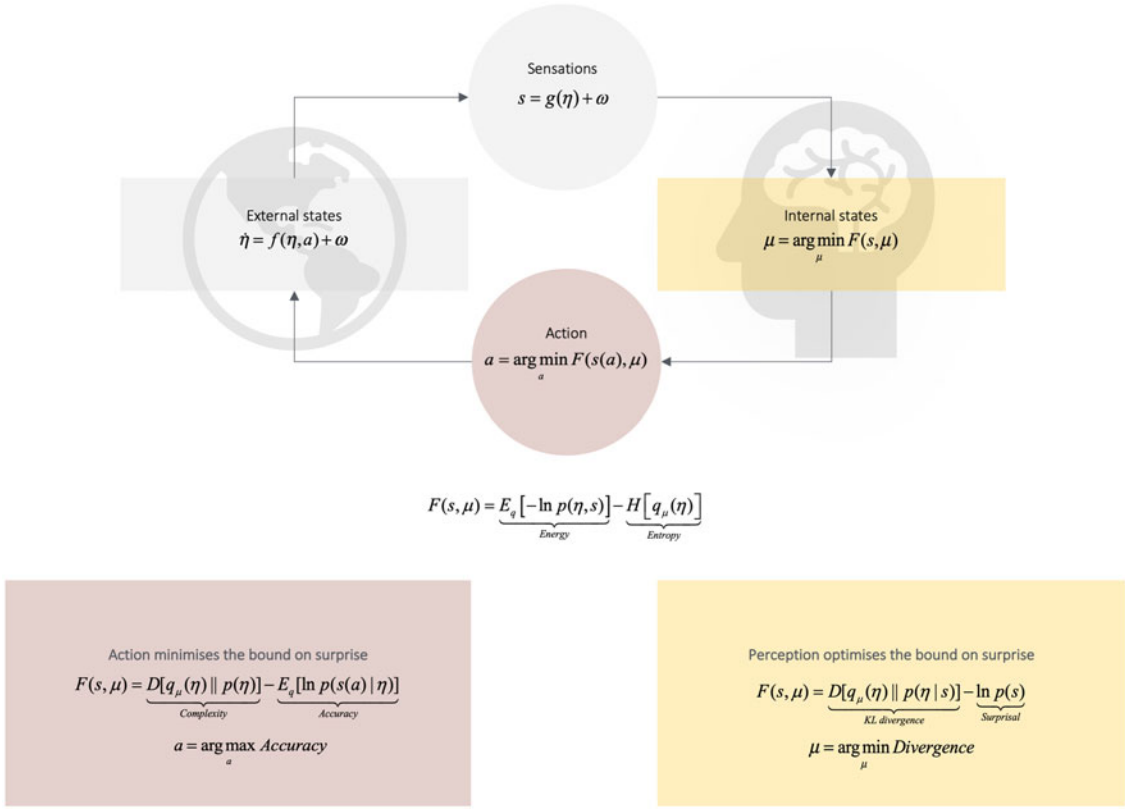


Figure 1. Self-evidencing and the Bayesian brain. Upper panel: Schematic of the quantities that define an agent and its coupling to the world. These quantities include the internal states of the agent (e.g., a brain) and quantities describing exchange with the world – namely, sensory input and action that changes the way the environment is sampled. The environment is described by equations of motion that specify the dynamics of (hidden) states of the world. Internal states and action both change to minimize free energy or self-information, which is a function of sensory input and a probabilistic belief encoded by the internal states. Lower panel: Alternative expressions for free energy illustrating what its minimization entails. For action, free energy (i.e., self-information) can only be suppressed by increasing the accuracy of sensory data (i.e., selectively sampling data that are predicted). Conversely, optimizing internal states makes the representation an approximate conditional density on the causes of sensory input (by minimizing a Kullback-Leibler divergence between the approximate and true posterior density). This optimization makes the free-energy bound on self-information tighter and enables action to avoid surprising sensations (because the divergence can never be less than zero). When selecting actions that minimize the expected free energy, the expected divergence becomes (negative) epistemic value or salience, whereas the expected surprise becomes (negative) extrinsic value – namely, the expected likelihood that prior preferences will be realized following an action. See the Appendix for a technical explanation – and description of the variables in this figure.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

beliefs, encoded by brain states, to optimize a generative (in the sense that it makes predictions) model of the world. When these beliefs are realized by action upon the world, this process is known as *active inference* (Friston 2011; Friston et al. 2017a). Active inference involves the coordination of sensorimotor patterns (1) by selectively sampling sensations that minimize expected surprise (i.e., by actions that include orientation, attention, and exploration) and (2) by updating expectations about the most probable causes of sensory inputs (i.e., perception). Perception entails optimizing beliefs about states of the world and learning the parameters of generative models, via Hebbian processes of associative learning (Friston 2010).

3.2. Attention and learning

Not all kinds of sensory inputs are equal in their significance or reliability, and therefore, they need to be differentially weighted when updating beliefs via free-energy minimization. For example, interoceptive signals might merely be tracking physiological noise (Feldman 2013; Seth & Friston 2016), or, again, exteroceptive sensory streams can stem from anomalous events that are unlikely to recur. Nevertheless, a priori, any signal can indicate relevant information that is worth accumulating, insofar as it enables an agent to track statistical regularities of the niche. An important aspect of self-evidencing involves updating beliefs about the reliability or precision of sources of information, particularly sensory input. *Sensory precision* corresponds to the precision of sensory information (e.g., how much confidence or reliability can be afforded auditory input when a rabbit listens out for a fox sneaking in the grass).

Because the agent has to navigate a capricious and context-sensitive environment, it also needs to assess the *precision of its own expectations* – namely, how far expectations depart from typical beliefs. This corresponds to *prior precision* (e.g., how much confidence or precision a rabbit should afford its prior beliefs, *given its expectations* about the presence of foxes in the area at that time of the day). Note the subtle but fundamental difference between expectations or beliefs about the (first-order) causes of sensations and expectations about precision, which constitute (second-order) estimates of statistical context (Hohwy 2013). In short, precision reflects the reliability of expectations about states of affairs – that is, whether or not sensory evidence or prior beliefs can be *trusted* (and not what they concern per se).

Using the FEP, we can distinguish two complementary, but computationally distinct, aspects of the folk-psychological concept of "attention" (Parr & Friston 2017a; 2017b; 2019): (1) as the process of directing the organism to selective sampling of the world (through shifting attention, sensory modulation, movement, or exploratory behaviour) such as to resolve uncertainty (i.e., expected surprise)[3]; and (2) as the calibration or weighting of this information as it is gathered to minimize surprise. Both play a crucial role in what follows. Under the FEP, *salience* is considered the main candidate for the implementation of attentional processes in the first sense – namely, the information gain or resolution of uncertainty afforded by the active sampling of the sensorium. The second sort of attentional selection corresponds to *precision weighting* (the modulation of belief updating as a function of estimated precision). This attentional process selects certain (neuronal) messages for belief updating through differential selection or modulation (Stephan et al. 2008). In short, *salience* is an attribute of action, in the sense that a particular way of sampling the world epistemic affordances, whereas attentional *selection* via precision weighting is an attribute of perception, in the sense of accumulating the right sort of information after it has been sampled.

Figure 2 illustrates the attentional selection of messages using a predictive coding formulation of free-energy minimization. In this formulation, prediction errors are passed upwards through hierarchical connectivity architectures in the brain to update higher-order expectations. In turn, the expectations provide descending predictions to create prediction errors. In this scheme, sensory precision is assigned to prediction errors at the sensory level of the hierarchy, whereas prior precision is assigned to prediction errors at higher levels. This precision weighting is thought to underwrite attentional selection of sensory input and is a crucial aspect of perceptual inference (Feldman & Friston 2010; Hohwy 2013). In what follows, we will subsume both sorts of attentional mechanisms under salience, given that overt sampling and covert attentional selection both conform to the same variational principles, under the FEP.

Attentional salience plays a central role in learning to engage with culturally constructed niches, both to select sensory evidence relative to the individual's goals and to identify sources with high reliability. The cultural affordances model proposes that human agents acquire culture by being immersed in specific, culturally patterned practices that modulate salience, which we call "regimes of attention" (Ramstead et al. 2016; Veissière 2016). Most regimes of attention do not involve isolated independent features of the environment, but correlated cues and opportunities for epistemic action that are organized in terms of local, cultural forms of cooperative activity, norms, and practices.

As we will describe in section 3.4, and as shown in Figure 3, these epistemic actions are supported by epistemic resources offered by the local cultural niche. In turn, regimes of attention correspond to the *salience* or epistemic affordance of sources of cultural information embodied in the epistemic cues of the niche. As shown in Figure 2, through active inference over the local cultural niche, humans can learn the norms and other contingencies that govern their local cultures.

Crucially, the configuration of regimes of attention by cultural practices and the ensuing attribution of salience to cultural information is only one of two aspects of cultural learning under active inference. The other aspect is the modulation of salience *via the modification of the environmental aspects of the patterned cultural practices* (e.g., people and material artefacts). As we will see in section 3.4, this "external" modulation of salience is enabled by mechanisms that we associate with developmental niche construction broadly construed (by analogy to internal mechanisms, such as perception and learning in the brain; Bruineberg et al. 2018b; Constant et al. 2018a; 2018b). Indeed, most predictions made by human agents result from – and pertain to – interactions with other human agents that co-construct a shared local culture and its niches. Through these niches, this culture furnishes feedback for the neurocognitive processes that serve the cultural patterning of attention (Seligman et al. 2016). As such, it follows that what we call "culture" is an extensive process that recruits elements both within the brain and in the shared cultural world (e.g., constructed places and designed artefacts).

3.3. Novelty, salience, and surprise

One might argue that there is an important design specification issue here; that is, to what patterns is salience or epistemic affordance attached (e.g., specific sensory information, families of

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

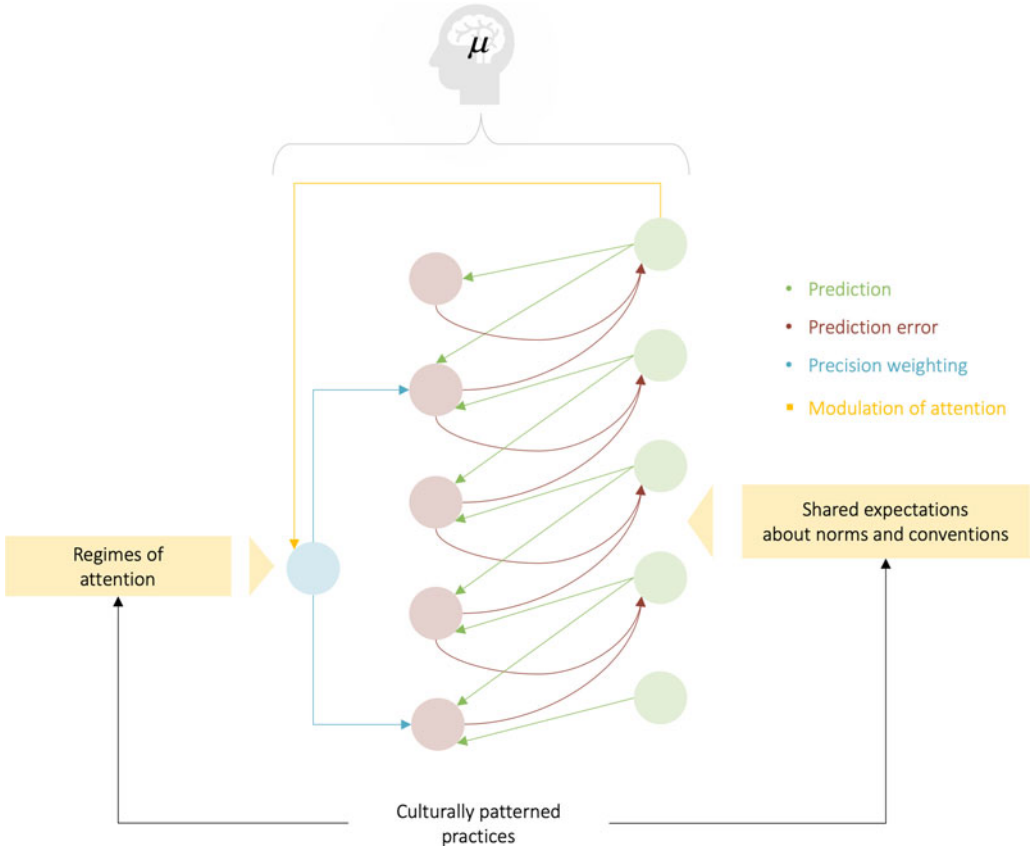


Figure 2. Cultural affordances. A schematic illustration of the looping effects that modulate social learning by human agents through expectations that, in turn, enable their interaction with cultural affordances. The attentional processes of individual agents are modulated by regimes of attention and by the shared expectations, norms, and conventions that characterize their local culture. In this example, the key point is that the yellow arrows effectively bias self-evidencing towards or away from (certain kinds of) sensory evidence – and that the optimal selection (i.e., salience) has to be both learned and learnable in the right sort of cultural context. Adapted from Ramstead et al. (2016).

similar events, and sources of information)? Any such assignment implies a pre-existing conceptual structure that allows for parsing the flow of information and that imparts some kind of hierarchical organization to available information. Precision and salience estimates are judged against some notion of what is salient (and this cannot just be what is stable over time, because that could result in a small, self-satisficing circle).

Under the FEP, these design specification issues are addressed by assuming that the agent embodies expectations that are established through histories of learning and, ultimately, through natural selection (Badcock 2012; Badcock et al. 2019; Friston 2010). Prior expectations are heritable through genetic, epigenetic, and exogenetic mechanisms (Constant et al. 2018b). These specify the epistemic value of sensations and, by the same token, the extent to which they should be considered. Priors that are inherited by the agent thus mandate the occupation of a limited repertoire of sensory states with high epistemic value that are revisited again and again (Friston 2010; Friston et al. 2015; Pezzulo & Cisek 2016), thus giving the impression that the agent maintains its organization (i.e., limits or minimizes the free energy of its phenotypic states with regard to the states in its niche). Our account thus focuses on the *conservative nature of human culture* – its ability to ensure that certain well-bounded and highly valuable states are frequented.[4]

Conservation is essential to cultural continuity and enculturation, but cultural niches also constantly change through creative innovation and adaptation. This raises the question of how free-energy minimization and dynamical coupling can account for creativity and innovation in social coordination, behaviour patterning, and the organization of sociocultural ensembles. Proponents of the FEP face a similar issue at the level of individual cognition, known as the "dark room problem" (Friston et al. 2012a; Kiverstein et al. 2019). The problem is simple: If agents aim to avoid unexpected encounters with their environment, we should expect minimally changing sensory environments like dark rooms and correspondingly monotonous sensations to be the most frequently (re)visited states of an organism. Yet, there are countless examples in every aspect of life (from art and politics to eroticism, contemplation, and drug taking, to name but a few) in which humans seem motivated (or driven) to *maximize* novelty and evanescent states of being (Veissière 2018). What, then, prompts novelty seeking behaviour at the level of individuals and social ensembles?

The FEP deals with the issue of novelty-seeking behaviour by formalizing action as being in the game of maximizing the *epistemic value of action* (or *epistemic affordance*). In essence, free-energy minimizing agents seek to sample the world in the most efficient way possible. Because the information gain (i.e., salience) is the amount of uncertainty resolved, it makes good sense for the agent to selectively sample regions of environment with high uncertainty, which will yield the most informative observations. This relates to the development of *artificial curiosity* in

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

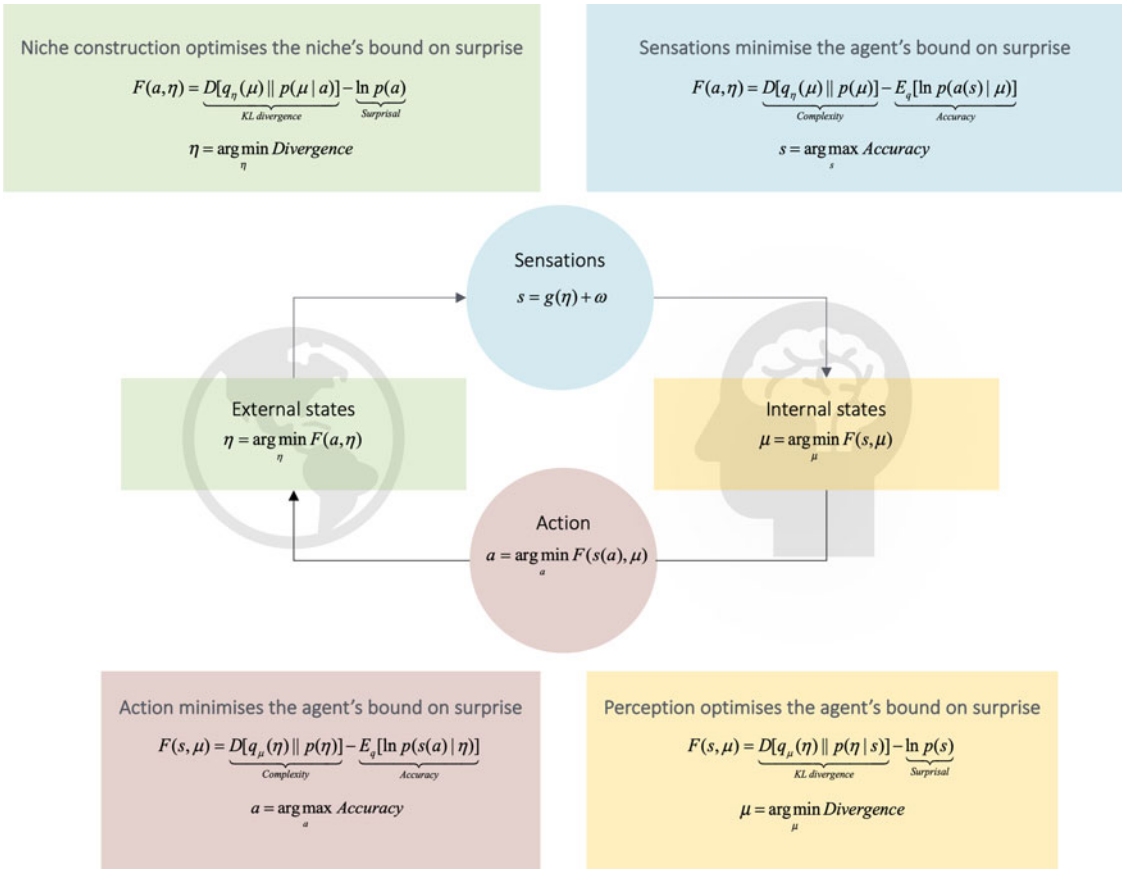


Figure 3. Summary of the variational approach to niche construction. As in Figure 1, internal states and action change to minimize free energy based on sensations and beliefs. Heuristically, one can think of niche construction as the process whereby the agent's action creates a symmetry between internal and external states. The agent changes the statistical structure of the world as it acts on the world. The statistical structure of the world here simply refers to the actual probability of finding some causes of outcomes at a given location in the environment (e.g., the bread being the cause of pleasant smell in the bakery). From the point of view of niche construction, such probability changes as a function of the agent's action and in a way that is consistent with the agent's beliefs. Indeed, a simple consequence of agents acting to optimize action based on beliefs is that the traces produced by agents' action will tend to be consistent with their beliefs. Another intriguing consequence of this is that, over time, traces in the world will effectively "learn" agents' beliefs, in the sense that those traces will encode statistical regularities that relate to those beliefs. For example, consider a well-worn path cut through the grass in the park. Such a "desire path" encodes a robust probability that the location of the path in the environment will map onto the probability outcome "being walked on." The value of that probability mapping increases over time as people wear down the path. This means that changes in the niche mirror changes in agents' beliefs enacted via action. With the mathematical apparatus of the free-energy principle, one can model "environmental learning" about the agents' action in the same way that one models "agents' learning" of the environment's sensory causes. The only twist is that the quantities are inversed (compare blue and green vs. yellow and red boxes). From the point of view of the environment's generative process, actions play the same role as sensations in the agent's generative model (for a detailed mathematical description, see Bruineberg et al. 2018; Constant et al. 2018b).

neurorobotics as a form of *intrinsic motivation* – so called because the resolution of uncertainty is itself intrinsically valuable and drives exploration (Friston 2017a; 2017b; Oudeyer & Kaplan 2007; Schmidhuber 2006).

In effect, agents will act to optimize the *epistemic value or affordance* of an action before acting on its *pragmatic value*, which is essentially its expected utility (Friston et al. 2015; Pezzulo et al. 2016). For example, if one enters a dimly lit kitchen to grab a midnight snack from the pantry, one is more likely to turn the light switch on before heading to the pantry. Turning the light on allows one to get an optimal grip and disambiguate the situation, before one acts on the pragmatic value (i.e., the utility) offered by snack foods. In short, the dark room objection fails because it simply does not take into account the formal description of action under the free-energy principle. In selecting action, an active inference agent (also known as a free-energy minimizing agent) attributes an intrinsic value to the reduction of uncertainty, which entails exploration. Hence, under active inference, policy selection fundamentally is guided by intrinsic, epistemic (belief-based) imperatives. This formally differentiates approaches based on the FEP from non-epistemic (belief-free) formulations, such as reinforcement learning (Cullen et al. 2018).

Intrinsic motivation[5] and artificial curiosity enable the agent to explore novel, transient, and unexpected regions of the space of policies open to them. This can be an "adaptive" exploration or epistemic foraging, because it allows for the exploration of this space; over longer timescales, the local increase in free energy serves the more general process of reducing free energy (for either the individual or the group, because it prepares the organism for potential changes in adaptive contexts and enlarges the repertoire of responses for the individual or the group). Similarly, cultural diversity allows individuals and groups to explore alternative niches that may provide adaptive advantage in the larger fitness landscape (Bengio 2014).

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

This can be seen on the temporal scale of human cultural co-evolution. The 7R variant of the *DRD4* gene (which encodes the D4 subtype of the dopamine receptor) appears to have become more widespread 50,000 years ago at a time of great migrations and a revolution in hunting technology among early *Homo sapiens* (Andrews et al. 2002; Shelley-Tremblay & Rosén 1996; Swanson et al. 2002). Traits like novelty seeking, creativity, high energy, and willingness to take risks associated with that gene likely conferred adaptive advantages in the environment of our ancestors. These may have become less valuable or even maladaptive later as human niches became safer, more standardized, and more predictable. Indeed, this shift in adaptive value with cultural context is invoked in evolutionary explanations of some forms of behavioural dysfunction, such as attention deficit hyperactivity disorder (Shelley-Tremblay & Rosén 1996; Tovo-Rodrigues et al. 2013). Of course, even maladaptive (non-optimal) traits may come to be culturally valued or exploited by individuals and communities, perhaps to their own detriment. Only the first of these pathways relates to the normal, adaptive acquisition of culture, which is the main focus of this article. However, both forms of epistemic foraging might contribute to cultural evolution.

3.4. Niche construction and learning

Culturally competent agents must learn regimes of attention across similar kinds of situations. For example, drivers must learn how pedestrians waiting at a red traffic light or crosswalk behave. The norms of pedestrian-vehicle behaviour vary in different cultural contexts. In some local contexts, pedestrians have the right of way and cars must stop, or pedestrians may observe red lights more laxly and attempt to cross against a red light, if the traffic is sparse. Within a given context, individuals' behaviour may vary. Drivers must learn how to respond quickly in such varying situations. To do this, drivers may internalize different estimates of precision (i.e., rates of variability) for different classes of agents (e.g., children might be more likely to cross the street without warning), and in turn, when travelling, drivers will re-adjust their expectations in light of local cultural variations in official rule obeying (e.g., in a country where people are more likely to jaywalk). In addition to the internal updating of precision estimates, one can think of epistemic affordances as encoded in the social-ecological niche (Constant et al. 2018b), in the patterned cultural practices that direct the epistemic foraging of agents (Ramstead et al. 2016), and in the specifically constructed aspects of the material environment (Constant et al. 2018a). For example, drivers and pedestrians learn not only how to assess the information afforded by traffic lights, but also how to leverage the traffic light's probable influence on others to improve the quality of their assessment (Constant et al. 2018a) – for example, checking that the bus driver can see his red light, before stepping out onto a pedestrian crossing.

Responding to a culturally constructed niche depends on a developmental history of learning to negotiate similar niches (a developmental history that is shared with all conspecifics within the same econiche). In the process of development, however, humans not only respond to niches, but also take part actively in their (re)construction. For example, based on the frequency of traffic accidents at an intersection, the location or timing of traffic lights may be modified by collective action. This (re) construction of the niche occurs in more rudimentary ways constantly throughout the development of individuals and groups in local niches.

From the point of view of the FEP, *developmental niche construction* can be viewed as the process whereby agents make their niche conform to their expectations (Constant et al. 2018a). Developmental niches are the set of exogenetic, physically and behaviourally grounded resources necessary to guide the reproduction of the adaptive life cycle (Stotz 2017; Stotz & Griffiths 2017). Because actions are guided by salience, and change the physical architecture (and epistemic affordance) of the environment, they tend to make the niche a good statistical "mirror" of the agent's epistemic foraging, functional anatomy, and, ultimately, brain-based expectations (Constant et al. 2018b; Fig. 3). In short, if we all act successfully to minimize uncertainty, our econiche will become inherently more predictable – if, and only if, epistemic affordances become encultured.

The exploitation of regimes of attention – encoded in the niche – is especially useful to track regularities unfolding over longer timescales of the history of a community, whose variability would be harder to assess over the timescale of an individual's perceptual and procedural learning. In humans, the epistemic affordance offered by niches constitutes *epistemic resources* that shape learning, and shared cultural practices (Hutto 2012; Roepstorff et al. 2010), as well as social relationships necessary for cooperative activities like breeding of animals (Burkart et al. 2009). Many of these epistemic resources involve specific kinds of patterned cultural practice that we associate with regimes of attention (Burkart et al. 2009; Hutto 2012; Roepstorff et al. 2010; Veissière 2016). These epistemic resources are states of the environment that, when repeatedly engaged by agents, shape their neurally encoded precision and salience expectations and, thereby, direct their future patterns of attention, epistemic foraging and learning, and subsequent patterns of engagement through perception and action. Epistemic resources help agents learn (from others) how to attend to or forage the niche for relevant affordances and how to weigh the cues associated with different affordances. Epistemic resources allow the agent to track and evaluate the relevance of more abstract, temporally extended, stable, and general statistical regularities structuring agent-niche relationships, like conventionalized patterns of interaction shared among multiple agents.

3.5. Learning cultural affordances under the free-energy principle

Epistemic affordances are encoded by – or installed in – the environment, as repeated physical actions leave traces that change the structure of the developmental niche in ways that influence agents' expectations (e.g., "I can trust that by taking this trail, which other people have also taken, I will end up at the other side of the park"). Over time, these traces of the actions of other people (e.g., traffic signals, dirt paths across a park, and shelters for hikers along a mountain trail) make certain affordances stand out as especially relevant. These are the affordances that yield highly reliable actions (i.e., uncertainty minimizing action, or actions that are expected to guide the agent towards goals or expected states) (see Fig. 4).

In many situations, affordances based on the history of human action will be more salient than those that reflect simple optimization (e.g., cutting across a lawn might afford getting to the other side faster, but many people will walk along a winding path, even in the absence of other humans). The well-worn path reflects an implicit consensus among many previous walkers. Individualized expectations guiding behaviour in context may thus be inferred

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

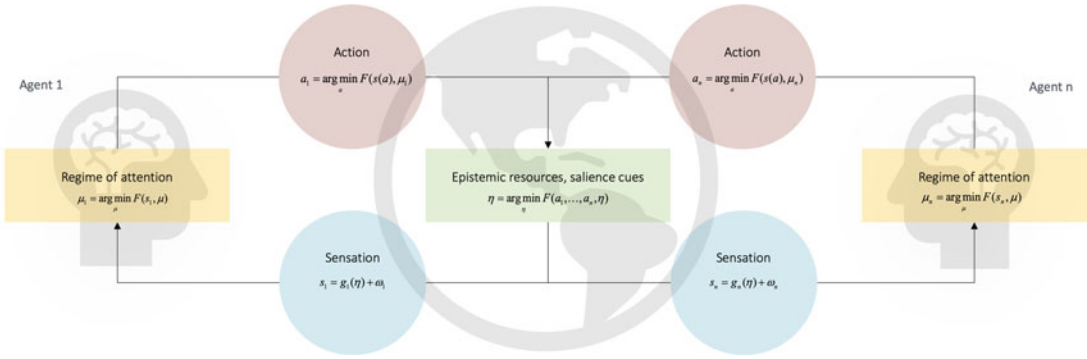


Figure 4. Thinking through other minds (see Figs. 1 and 3 for the equations). This figure depicts the loop between action, sensations, and niche construction that lead to the acquisition and production of cultural habits, and to the inference and learning about other minds. The shared epistemic resources in the constructed niche (i.e., external states modified by actions from agents 1 to n) and the regimes of attention (i.e., internal state) constitute the domains of statistical regularities that tune to one another via the physical engagement of the niche. Those domains are finessed (i.e., mutual learning of internal and external states) by a community of practices (agents from 1 to n) over ontogenetic (e.g., over development) and phylogenetic timescales (e.g., via the inheritance of material resources). The learning and deployment of internal and external domains of statistical regularities is what we call "thinking through other minds" (TTOM). TTOM entails, and depends on, the production of culturally patterned practices. Cultural practices and associated artefacts are epistemic resources that guide the attention (and learning) of members in the community by shaping sensory perception.

from a *continuum of expectations about other agents*, ranging from reflective to fully intuitive, and, in turn, from actually present to probable and generalized others. Under the FEP, the dynamics and acquisition of all these expectations by a group of agents are mediated by the very same inference mechanisms.

Developmental niche construction can be cast as an interactional process between agents and a shared environment, producing affordances that support the reproduction of a normative life trajectory, through the norm-guided development of each new member of the community (cf. Constant et al. 2018b; Fulda 2017). These norms are implicit in the structure of cultural affordances in the specific local niches occupied by individuals at a particular developmental age or stage. Individuals become attuned to the niches they discover or are directed to by others according to their age, gender, and other dimensions of social status.

These niches afford individuals epistemic resources for acquiring specific types of knowledge, skills, or dispositions to respond. In effect, the function of external mechanisms for evaluating epistemic affordances is to enable the emergence and *stabilization of epistemic resources*. The notion of epistemic resources relates directly to work on how cultural knowledge held by others in the community can reach into the hierarchy of processing at higher levels through linguistic or symbolic communication to install priors directly (Bengio 2014).

Epistemic resources, which underwrite epistemic affordance (either overtly through action selection or covertly through attentional selection; i.e., mental action), are stabilized through niche construction, in the sense that the niche comes to encode the expectations that enable the interaction with those affordances. Epistemic resources act as developmental anchors. In human social contexts, epistemic resources can be viewed as shared expectations and cultural affordances that become available to a group of agents, as expectations that "sediment" in public places, practices, and affordances that are repetitively and reiteratively engaged by groups of agents. This process involves feedback or looping effects and hence is self-reinforcing over time. For example, the grass patch on a street corner solicits cutting across and, over time and in turn, as it is worn down by many walkers, comes to afford a "desire path" (Ingold 2016).

One might ask whether the story should not be told the other way around. It might be that dirt trails allow for cutting across the park, but only later, solicits a "desire path," as it is only once the agent has acquired the cultural knowledge that the path can be traversed that it can become "desired" as something that the agent wants to engage. Precisely what is at stake here is the virtuous circularity and bootstrapping operative in social learning – which must go from simple to more complicated. On a phenomenological level, what is being challenged is that the world calls to us in specific ways prior to the desires installed by culture – in cutting across the path, the unstated background of desire might have to do with getting somewhere we want to be more quickly, with enjoying transgressing the rule of walking (only) on sidewalks, or simply the aesthetics of walking along a dirt path. Hence, it is not self-evident that one can consider a desire path, or for that matter, any cultural object, as a cultural affordance until some way of engaging the world has been acquired.

Affordances have been proposed to explain how skilled agents manage to engage their environment without having to know how their environment "works" (i.e., to employ learnt representations or to acquire representational contents). The variational approach furthers this line of thinking by distinguishing mathematically action that is selected by the agent and the *affordance* of action for the agent. In effect, the FEP allows us to formulate a *principle of most affordance* – that is, a version of the principle of least action from physics, applied to the adaptive behaviour of groups of organisms living together in a niche (Ramstead et al. 2019a). The action with the most affordance, the one that solicits the organism most (i.e., the one associated to the least *expected free energy*), is the one that ends up selected by the organism.

The cultural affordances framework suggests that acquiring the ability to leverage conventionalized affordances means acquiring a regime of attention. The regime of attention is not some specific content that one learns, but a mode of attending to and actively sampling the world, through a generative process that involves (overt) motor behaviour and the (covert) tuning of neural gating via expectations about precision, as well as culturally patterned search strategies for salient information,

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

which are "shared" to some extent by all individuals of a local culture.

The idea behind the desire path as a cultural affordance relies on and extends the notion of regime of attention by highlighting that *epistemic affordances* depend not only on the brain, but also on features of the environment (see Fig. 2).[6] The desire path, as a cultural affordance, enables skilful pre-reflective engagement. This can often happen without the agent having to know the content of the specific artefact from the start. For example, I might be late for my train, and following that trajectory through the park might be a good solution to catch my train on time. In that scenario, there is probably very little *content* involved with about where *exactly* the path will lead. Rather, there is (1) *an expectation* on the part of the agent, (2) a *solicitation* on the part of the environment, and between those, (3) an embodied *history* of agent-niche interactions (i.e., the traces left by repeated actions), which increases the likelihood of the path leading to a commonly experienced goal (e.g., the other side of the park). This history of cycles of expectation, solicitation, and action, encoded in cultural affordances, supports individuals' intuitive, culturally meaningful response to environmental cues. Under the TTOM model, when individual agents do not know quite what is situationally appropriate, their behaviour switches to *epistemic foraging*, in which agents will preferentially sample whatever other, relevant agents sample as well.

A large part of the social learning enabled by the developmental niche is mediated by shared attention (Tomasello 2014). For example, once a path is worn in the grass, implicit shared attention and expectations that others also intended to do so will prompt followers to walk along the path. This will hold even for paths that are not otherwise efficient, even if a less costly path is available – and, in some instances, this holds even for paths with uncertain trajectories or end points. Of course, most of the traces of human activity are not paths on grass, but the affordances provided by institutions, archives, and repositories of knowledge, plans, and protocols. Regimes of attention provide ways to locate, attend to, and engage these affordances in a wide variety of structured cooperative activities (Malafouris 2015).

3.6. Why human thinking is always already thinking through other minds

Homo sapiens evolved to rely on bodies of accumulated cultural knowledge and skills for survival (Henrich 2015; Sterelny 2012; Tomasello 2014). We shape each other's learning through specifically adapted cultural practices (regimes of attention) that allow individuals to enact recursively nested forms of intentionality. This includes the capacity to view ourselves through the eyes of another in a kind of reciprocal aboutness (e.g., "What would Mother expect me to do?") After childhood, typically, these ways of thinking about oneself are internalized, encoded, and expressed as "What should I do?" or "What am I expected to do?" Recent research on mind wandering suggests that most of our spontaneous mental life is dedicated to rehearsing social scenarios (Poerio & Smallwood 2016). In their recent "interactionist" account of the evolution of human reasoning, Mercier and Sperber (2017) review a wealth of experimental evidence to support the claim that humans best solve problems and optimize individual intelligence collectively in dialogical and argumentative contexts, which may extend to hypothetical, "silent" scenarios. Although no large-scale evidence is available on what so-called "silent reasoning" entails in individual human heads, Sperber and Mercier conjecture that most silent reflective ideas are generated through the rehearsal of arguments with, and justifications to, others. Even solitary thinking, on this view, is a rehearsal for bona fide social interactions with peers.

Recent work in the philosophy of psychiatry also supports the hypothesis that solitary human cognition is social through and through. In their cultural and evolutionary account of the origins of psychosis, for example, Gold and Gold (2015) propose that the many kinds of delusions described in the literature on psychopathology (i.e., persecutory, grandiose, erotomanic, control, thought, somatic, nihilistic, reference, guilt, and misidentification) share one broad, overarching theme: a concern with one's relationship to *other people*. Hence, all known delusions can be recast as statistically improbable interpretations of, and expectations about, one's experiences *in relation to others*.

For a species such as *Homo sapiens* that evolved to rely upon cooperative and highly elaborate coordinated action, expectations about folk psychology (or probabilistic inferences about the way other people think and reason and what they expect of the world) are at least as important as, if not more important than, expectations about statistical regularities that characterize the physical world. In other words, in a world populated by creatures "like me," most of my expectations call on the prior belief that "I am like you and you are like me – and you believe that I am like you and you are like me" and so on. In effect, the world of human experience is always already mediated by, and filtered through, the "lens" of expectations about another's expectations.

The expectations that *Homo sapiens* have leveraged most over their phylogenetic history involve the capacity to *outsource cognition to relevant others* (people, artefacts, practices, and institutions). In other words, human beings outsource to other humans many of the evaluations of salience that they employ in their engagement with their worlds, which allows others to perform culturally relevant tasks (Tomasello 2014). Indeed, it is precisely these evaluations by others that make worlds "meaningful" for humans. To exploit this cooperative cognitive task sharing, humans agents explicitly and implicitly bestow trust and assign authority to others – both individuals and institutions – acquiescing to and leveraging cues (physical, culturally meaningful *signs*) associated with reliability, authority, and prestige (Henrich 2015).

What distinguishes between different human phenotypes is the priors under which they are operating, and which guide adaptive behaviour. If we consider the dynamics of human TOM abilities in this light, the process of TTOM consists in inferring the priors or expectations that guide the beliefs of another agent or group of agents. Provided that agents can solve the inference problem about the sort of person that their interlocutors are, and provided that they have a model of their conspecifics' prior beliefs, then any one agent can leverage their own action (policy) selection mechanisms under the prior beliefs of their fellows to infer the mental states of their fellows (and, indeed, their own mental states).

Epistemics get into the game when this inference is made more difficult by a lack of shared priors. Hence, the cues that emerge from niche construction can be nonspecific cues that tell agents about what is situationally appropriate to do (but which could be done in a solitary way, such as stopping at a red traffic light) or very particular cues that provide information about the priors of other agents, which coincides with mind reading and properly thinking through *other minds* (e.g., I have a prior about you

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

having a prior about me stopping at the red light and crossing at the green light – and, hence, that you will not run me over). The process of inference is made easier by the availability of cues (that shape regimes of attention) that tell agents "where to look" (i.e., that allow one to leverage where others are looking to determine where oneself should look). For example, if I do not know when to cross at the intersection because I am not familiar with the colours used by the traffic light system, I can guide my action by relying on epistemic cues that have been shaped by (presumably adaptive) cultural practices such as the ways people around me act in context (e.g., other agents' behaviour or gaze patterns).

The TTOM model accounts for the ways in which human agents outsource their policy selection to relevant others and to aspects of their material niche. In this sense, our model covers cases of cultural cognition that range from the lone encultured agent acting in conformity with the cultural norms that they have internalized – which involves inferences only indirectly about and through other minds – to full-blown cultural engagement with other human agents, which requires (implicit and explicit) inferences about the minds of other humans. Given the nature of their inferential systems and the way they learn generative models according to TTOM, inferences about my own generative model can be leveraged, and, in effect, is always being leveraged, to make inferences *about others like me*. Inference about one's own mind is always mediated and made possible by inferences about the minds of others.

4. Addressing TOM critiques with TTOM

According to TTOM, human agents organize most of their behaviour as a function of *what they can infer from other human minds*. Humans find guides for action by picking up on statistical regularities in the realm of *folk psychology*, which identifies the most relevant states of the external world, as well as the most relevant sources of inferences about the shared social world. Our framework recognizes the contribution of the varied approaches to human TOM abilities outlined in the first section and offers a compromise position.

4.1. Response to the cross-cultural critique: TTOM is universal for Homo sapiens, but realized through cultural niches

We agree that folk notions of personhood vary across culture and likely exercise specific constraints on automatic perception and social coordination through *normative* social learning (e.g., McGeer 2007). Although folk notions of the locus of personhood and agency vary broadly between groups and historical periods (e.g., to include a soul, brain-mind, heart-mind, or external agencies like gods, ancestors, or spirits), we question the extent to which communication and coordination would be possible without a species-wide intuitive notion of *propositional psychological interiority* (which may be postulated and enriched in different ways culturally).

The example of "silent thinking" during courtship, reported from ethnographers of the Korowai (Stasch 2009), is telling. In everyday human experience, affectively charged situations such as "I wonder if she really likes me" abound and likely emerge in infancy without recourse to language or explicit mentalizing, as humans form mental models of other agents in their life. Indeed, developmental psychologists have shown that 15-month-old infants are able to take into account the false beliefs of other agents (Onishi & Baillargeon 2005) and that the ability to attribute goals to any entity (living or not) that appears to be animate emerges as early as 5 months (Luo & Baillargeon 2005; see Mahajan & Woodward 2009 for different results).

Additional cross-cultural and developmental findings support the view that intuitive dualism (Jack 2014), or the folk tendency to situate personhood in an intangible psychological interior, is likely a cross-cultural universal that does not require specific cultural immersion in Cartesian cultures (Chudek et al. 2013). As Paul Bloom (2005) has argued, children across cultures can readily understand a story about a prince becoming a frog without explicit enculturation into folk Cartesianism.

As we argue in Section 4.2, TTOM makes no ontological claims about mind-body dualism; we simply point out from experimental and ethnographic evidence that coordinated action in human sociality *does* rest on the universal human cognitive capacity to understand others as having goals, beliefs, desires, and intentions that may be different from their stated ones (what we call "propositional psychological interiority"). At the core of this cognitive capacity is the process of active inference mediated by processes of developmental and selective niche construction, which, in humans, scaffold complex sets of prior beliefs encoded in sites across the brain-body-environment-others system. Hence, "mind reading" sometimes requires explicit deliberation (something resembling "theory theory") and at other times can be automatically intuited through simulation (in forms of embodied and extended cognition).

4.2. Response to the embodiment critique: TTOM is grounded in the bodies of self and others

Anxieties around dualism in current cognitive science reflect a common confusion between *normative* and *descriptive* commitments on the part of philosophers and cognitive scientists. Although dualism as a scientific description of the relation between the mind and body is mistaken, it does not follow that our theorizing about other minds should not consider folk dualist thinking as a normative and very real phenomenon that shapes every day and scientific thinking. As an illustration, even psychiatrists who espouse an integrative, monistic view of mind and body employ a naive dualism in assessing vignettes of problematic behaviour as indicating either deliberate action (rooted in individual psychology and, hence, blameworthy) or as accidental, because of malfunctioning biology of the brain (Miresco & Kirmayer 2006) – as though these two causes were grounded in distinct mental and bodily processes. Our best theories about folk social cognition ought to reflect that dualism, on pain of descriptive inadequacy.

TTOM, to be sure, does not make ontological claims about the nature of mind as separate from the body. We simply offer that, as a matter of universal human epistemology, patterned cultural practice involves an ability to make inferences from, through, and about other minds, as propositional processes – indeed as inferential processes. In some cases, folk theorizing about dualism may simply be a useful tool to both generate and inquire on such practices (e.g., through dialogues in clinical setting). TTOM formalizes the inferential structure of such folk theorizing.

The ability to infer each other's expectations, which makes human cognition, sociality, and culture possible at all, ranges from the fully explicit to the fully automatic depending on the situation. In our model, this ability depends on the learning of a spectrum of expectations encoded across the brain-body-

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

environment-others system that underwrites regimes of attentions. The FEP is unique here in its ability to account for *inference and dynamics* as two sides of the same coin, and this is what allows TTOM to overcome the sharp dichotomy between internalist and externalist approaches to TOM abilities. Under the FEP, all systems dynamics are inferential, and inference is itself dynamics; namely, the dynamics of sentient systems are a gradient flow over free energy (Friston 2010; Ramstead et al. 2018). Because free energy is a measure of the complementarity between the organism and the niche, in terms of a generative model of the relation between them, any dynamics formulated in terms of the FEP are ipso facto *inferential dynamics* that pertain to the *self-organization of information flows in sentient systems*.

Rather than describing cultural differences in the folk models (including Western philosophical models!) of social cognition in "either/or" terms (either dualistic or not; focusing on explicit intentions or focusing on resonance in action), we propose to situate these differences on a continuum of *hypo-cognition* to *hyper-cognition* of intentions (see Duranti 2015). The notion of hyper- and hypo-cognition has been explored in the context of cultural variations in emotions (Levy 1975; 1984). The degree or depth of cognitive elaboration of emotion serves individual and social regulatory functions. As a matter of normative concern, cultures vary in the kinds of emotions people are encouraged to cultivate or suppress, thereby allocating attention, attributing meaning, and patterning behaviour in ways that constitute specific codes of conduct or expression, modes of experience, and folk explanations that account for behaviour.

4.3. Response to the cooperativity critique: TTOM is built on the developmental scaffolding of cooperativity

Shedding light on a cross-cultural continuum of normative commitments to the hyper- and hypo-cognition of intentions may also help resolve the Machiavellian-mutualist debate on the evolution of human cognition. It seems self-evident from the human record that our species is capable of both selfishness and altruism as a matter of individual, situational, and cultural variation – but also that the scaffolding of "altruism" proper clearly follows an evolutionary and developmental trajectory. Tomasello (2009), for example, proposed the early Spelke, later Dweck hypothesis[7] to describe children's gradual immersion into social norms that harness and enhance their natural capacity for adjusting their behaviour to what others expect of them.

Rather than start from a specific commitment to one normative position (e.g., "humans ought to be altruistic"; "humans ought to act in rational self-interest"), our account recognizes these varied possibilities inherent in human behaviour and stresses the importance of specific cultural practices in patterning behaviour to elaborate either side of the selfish-altruistic continuum.

Hrdy herself, as a proponent of the mutualist argument, has stressed the importance of developmental environments, such as collective parenting, in providing rich (or impoverished) opportunities to form bonds and learn to relate with multiple attachment figures – a process she describes as crucial in the development of social cognition, emotional regulation, and empathy (Hrdy 2011). In Hrdy's account, our "proximity" to the kind of selfish intelligence found among chimpanzees is a matter of ontogenetic contingencies at least as much as evolutionary "distance." Indeed, the capacity to engage in nuanced, compassionate, other-regarding action is increasingly understood to be dependent on language, explicit teaching, effortful deliberation, and practices and to be distinct from (though perhaps developmentally scaffolded on) the innate capacity to imitate and follow others and favour one's narrow in-group (Bloom 2017).

Contemplative practices of loving-kindness meditation, for example, entail the explicit enrichment and effortful rehearsal of one's mental models of others, which eventually become automatic through practice (Lebois et al., in press; Lutz et al. 2008). The linguistic (narrative) elaboration of these models may be essential to their extension to include members of out-groups, the whole of humanity, or even to all sentient beings. These varied examples point to the importance of both implicit and explicit mentalizing mechanisms in the mediation of human cognition and cultural practice.

TTOM supports current mutualist, cultural intelligence, or "dual-inheritance" accounts that emphasize the co-evolution of human cognition and culture (Henrich 2015; Tomasello 2014). Rather than to discount Machiavellian and other "selfish" accounts of these processes altogether, we suggest that what one might call *extended mutualism* (i.e., large-scale cooperation), and the ability to leverage a large repertoire of shared expectations to guide group action, arises because of the match between naturally and culturally selected dispositions to acquire cultural abilities (e.g., mind-reading abilities) and inherited developmental conditions enabling the (re)acquisition of these abilities. Selected, or evolutionarily old, dispositions constitute a cultural learning "start-up kit" of sorts (Heyes 2018b; Heyes & Frith 2014), which includes the kind of neural machinery that underwrites attention and the estimation of salience, leading to the acquisition of shared expectations (see Fig. 2).

At the developmental timescale, inherited cultural practices enable the learning of shared expectations via the patterning of those evolutionarily old dispositions. This emerges via agents' engagement with epistemic cues that undergo processes of cultural evolution through developmental niche construction activities, which filter what persists across generations as a function of the success of the behaviours they afford (Laland 2018; see Fig. 3).

This sets up a cycle of mutual fitting between individual and niche. For example, in a circular fashion, I can trust the learning biases provided by my caregiver – and more specifically, the cues they provide through their gaze direction, pointing, gesturing, and so on, towards salient situations. I am licensed to do this because patterns of offspring-caregiver interaction have been filtered and fine-tuned through gene-culture co-evolutionary processes and developed in specific cultural norms, signs, places, and practices over historical time – all in the service of guiding the learning of salience; that is, to guide the learning of what is adaptive in the local cultural context (e.g., "listen to and copy this high prestige individual because prestigious individuals are typically the ones that have succeeded in the past"). Put another way, one can trust learning biases because biases indicate action policies selected by other agents "like me," so these must have been the most adaptive for creatures "like me."

On our account, cognition and culture are largely synonymous for humans, as both are predicated on the capacity for shared expectations. Priors leveraged and finessed through active inference, and the folk psychology they specify (i.e., what we expect others also to expect), constitute the central domain of statistical regularities that ground humans' models of their world. This domain of statistical regularities that we call TTOM specifies the mechanistic processes that drives the implicit acquisition of culture over development.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

5. Concluding remarks: The future of TTOM

5.1. Future research

We have argued that the pervasive influence of culture, through widespread shared expectations, institutions, and practices, can be cast as a process of co-constructing and responding to a shared set of affordances. Human engagement with cultural affordances is enabled by (often implicit, recursively nested) expectations about other relevant agents' expectations. These expectations are acquired by agents through immersive participation in the practices that define their shared way of life, in a process that gradually takes hold in ontogeny through regimes of attention and niche construction (See Box 2).

The human mind is optimized for outsourcing information to other human minds in order to function in a niche that requires the shared, coordinated pursuit of joint goals. Error and surprise minimization in large-scale social systems hold because *individual* human minds are coupled to one another in an environment of other minds. This kind of "extended mind" is distinctive to human beings because of the capacities for culture (i.e., regimes of attention, linguistic communication and installation of higher-order priors, multiscale cooperation, declarative memory/historicity, and collective norms and goal setting) that are made possible by the human nervous system (Clark 2008; Clark & Chalmers 1998; Menary 2010; Sutton 2010).

If we have been successful in presenting our account, however, from an FEP point of view, it should also be clear that humans think, feel, imagine, and act in ways that are only possible because they are afforded by the niches they inhabit and co-construct, and the cultural practices that make up their shared form of life, and that all serve to enculture human agents (Constant et al. 2018a; 2018b; Ramstead et al. 2016). Even the collaborative construction of new niches, which allows the exploration of new modes of experience and the improvisation of new forms of cooperative action, depends on the cultural scaffolding of a relatively stable set of shared expectations and regimes of attention through the cognitive tools or gadgets of narrative and metaphor (Heyes 2018b; Lakoff & Johnson 1980) and the social organization that constitutes particular niches or communities.

TTOM is a generic active inference (also known as FEP or variational) account of the acquisition of culture and mind-reading abilities. We have designed TTOM as a guide for the production of *testable models* in related domains. Although TTOM per se would be difficult to test (because of its generality), one can derive specific integrative models from TTOM to study specific forms of sociocultural dynamics. A good example of a testable model derived from TTOM is the theory of regimes of expectations as applied to the study of social conformity (Constant et al. 2019b).

Social conformity refers to the deference to social norms such as that embodied by other agents. From the point of view of social psychology, social conformity is one possible response to social influence of epistemic, trusted others (Asch 1956). From the point of view of cultural evolution, in turn, social conformity is viewed as an adaptive social learning strategy in an uncertain environment (Morgan & Laland 2012).

The theory of regimes of expectations integrates the perspectives of social psychology and cultural evolutionary theory by modelling social conformity as a process that obtains through the intergenerational finessing of environmental cues that guide social learning over development. Social learning that is aided by these cues, in turn, allows the active inference agent to perform action selection in a fast and efficient way in uncertain contexts by *leveraging trusted others* (either through material cues that stand as culturally signalled proxies for other, relevant or prestigious minds or directly by copying such individuals). These trusted others are defined as "deontic cues" (Constant et al. 2019b).

"Deontic cues" in this model are context-specific epistemic resources (as defined by TTOM) that enforce an obligatory response to the context that embeds them (e.g., a red traffic light enforcing stopping behaviour). The theory of regimes of expectations models social conformity as an active inference process of action selection that operates via the estimation of the epistemic, pragmatic, and also *"deontic" value* of action, which is the type of value learned through the engagement of deontic cues. The deontic value is essentially the value of an action policy specified by the shared beliefs and preferences of a sociocultural group.

In line with the sort of specific models that can be derived from TTOM, the theory of regimes of expectations as applied to the study of social conformity integrates externalist approaches (e.g., cultural evolutionary approach) and internalist ones (e.g., the social psychology approach) by describing the cultural domain of statistical regularities optimized through active inference and governing action selection.

The theory of regimes of expectations as applied to the study of social conformity makes specific predictions that stem from the TTOM model – namely, that (1) social conformity leads to more efficient cognitive processing and policy selection (e.g., as conveyed by psychophysics measurements like reaction time) in the presence of deontic cues (epistemic resources in TTOM terms); (2) conforming actions minimize variational free energy over time more efficiently in social context, because regimes of attention will be optimized for zeroing in on social information conveyed through deontic cues; (3) deontic cues reproduce conformist biases in cross-cultural between-subjects designs but fail in within-subjects designs (i.e., not all deontic cues will elicit social conformity for participants with culturally diverse backgrounds because of the influence of culture-specific regimes of attention).

5.2. Limitations

Because it is based on the FEP, TTOM provides a mathematical formalism that can be used to model the effects of cultural affordances on adaptation to specific kinds of social niches. The model needs to be further elaborated to deal explicitly with the many varieties of cultural learning and regimes of attention. These include the distinctively human functions of narrativity that entail the linguistic and symbolic hierarchical installation of higher-order priors (Bengio 2014). For example, this will include culturally shared expectations about the cause of sensory observations (e.g., the prior belief that "the slap I received on my wrist was caused by my belief that it is permissible to reach for the cookie jar, which motivated my action, which then led to the slap, which indicated it was not"). In this sequence, the slap not only conveys a social norm, but in itself reflects the broader social norm that it is permissible to intervene in childrearing in this fashion – these overarching norms are learned over time within a particular niche and may change, for example, with migration to a new sociocultural context, with serious consequences for how one (mis)reads (culturally conventional or permissible) affordances. In modelling an active inference agent, such structures of high-order priors could capture the potential for reflexivity and self-reference that gives human cultural-linguistic cognition its unique reach (Taylor 2016).

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Box 2. Glossary of key terms.

Active inference: Active inference is the process whereby organisms learn the statistical structure of their environment through the selective sampling of predicted or expected sensory information (also known as action), based on perceptual inferences about the cause of the sensory input (also known as perception). The process of active inference realises the free-energy principle. In active inference, everything that can change does change to minimize variational free energy, which is a statistical measure of the mismatch between organism and environment. This mandates actions that minimize expected free energy following an action – namely, actions that resolve uncertainty.

Affordance: Generally speaking, possibilities for engagement with an ecological niche that are defined in interactional terms, as a relation between features of organisms' environment and their own abilities.

Attentional salience: The degree to which uncertainty is reduced under a particular course of action. Mathematically, salience is known as expected Bayesian surprise, information gain, intrinsic motivation, and epistemic value. Salience underwrites epistemic affordance.

Attentional selection: Calibration or weighting of the precision (inverse variance) of sensory evidence, or prior beliefs.

Conventional affordance: Affordances that agents can engage by skilfully leveraging explicit or implicit expectations, norms, conventions, and cooperative social practices.

Cultural affordance: The kind of affordance that characterizes the human niche. Cultural affordances depend on shared expectations that are acquired over development (i.e., through enculturation and social learning). Cultural affordances come in two flavours, which form a spectrum from the more innately specified to the more learning dependent: natural and conventional affordances.

Epistemic affordance: One of the two components of expected free energy that determine action selection. Epistemic affordance quantifies the extent to which a particular way of actively sampling the world reduces uncertainty about the state of the world or its statistical regularities.

Epistemic authority: A symbol, person, cue, or feature of the environment (usually associated with prestige, status, and group affiliation) that signals salient, high-quality, uncertainty-reducing information in a given cultural context, and as such possess the "power" to guide attention, enhance credibility, and prescribe action (e.g., biomedicine and neuroscience possess high epistemic authority in current culture; the *Guardian* newspaper possesses high epistemic authority for liberals, as does Fox News for conservatives).

Epistemic foraging: The agent's uncertainty-resolving behaviour. Epistemic foraging disambiguates Bayesian beliefs about a situation in order to be better poised to exploit the pragmatic value of action (i.e., value that relates to the sensory preference of the agent).

Epistemic resources (also known as cultural affordances): Cues that are encoded in external states of the ecological niche (e.g., material cues and other agents), which guide epistemic foraging and implicit learning of patterned cultural practices.

Expectations: Bayesian beliefs and preferences about external states of the world, which are operationalized as probability distributions.

Free-energy principle (FEP): A principle of least action derived from information theory. The free-energy principle states the minimal conditions that systems must meet if they are able to endure in a bounded set of states (i.e., if they are endowed with a phenotype).

Generative model: A probability distribution or mapping from beliefs about hidden causes to observed consequences (i.e., sensations). Technically, this is the joint probability of a sensory state and a (hidden) state of the world. Under the FEP, the generative model defines free-energy gradients (a function of sensations and predictions under the generative model) and subsequent perception and action.

Natural affordance: Affordances that agents can engage by leveraging their innate phenotypical endowments.

Niche construction: The process whereby organisms (implicitly and explicitly) modify their ecological niches, such that the states of the environment come to encode relevant aspects of their prior beliefs, which they can leverage "downstream" to optimize their adaptive behaviour and act in contextually appropriate ways. The "Janus face" of active inference.

Pragmatic affordances: One of the two components of expected free energy in policy selection. Pragmatic affordance is essentially equivalent to expected utility in economics and quantifies the extent to which an action policy conforms to the prior preferences of the agent (also known as pragmatic or instrumental value).

Regimes of attention: Patterned cultural practices whereby members of a group of people acquire and maintain shared expectations that *modulate* attention, *structure* salience, and thereby *guide* action (Fig. 2), as well as the internalized patterns of attention that result from the repeated engagement with such practices (e.g., as a group-specific affordance, it takes a regime of attention for the colour white to signify mourning for Hindus; it also takes a species-wide regime of attention for humans to feel invited by a path in the woods that signals the trace of other humans' intentions).

Salience: Expected information gain under a given action.

Surprise: Also known as surprisal or self-information in information theory. This is simply the negative log probability of some state or event.

Thinking through other minds (TTOM): The *domain* of beliefs about statistical regularities (i.e., Bayesian prior beliefs) that are exploited in learning cultural affordances. This domain is primarily situated in the realm of *expectations that humans learn to form about other people in the niche* – that is, in the realm of *folk psychology*. TTOM is also the *process* of engaging others' expectations and inferences by leveraging this domain.

The free-energy minimizing dynamics described previously involve feedback processes that tune organismic expectancies to fit local environmental contexts and therein minimize surprise and uncertainty. Accounts of enculturation tend to suppose *stable social contexts*, and the FEP assumes a kind of optimization that depends on *stability in adaptive contexts*, but the reality (especially in the context of cultural interactions and contexts) is often one of *constant change*. Hence, realistic models of human cognition in context will require taking into account cultural mobility, hybridity, and the cognitive effects of the constantly changing social niches that reflect cultural co-evolution. Ultimately, models based on conservative processes

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

like the FEP model need to address the significance of historicity and contingency in the emergence and evolution of cultural systems.

Among other potential domains of application, our model has implications for psychiatry. One interesting path towards experimental verification builds on recent proposals for a *computational psychiatry* (Adams et al. 2016; Friston et al. 2014b; Huys et al. 2016; Montague et al. 2012). In brief, computational psychiatry aims to leverage computational techniques in order to better phenotype various psychiatric conditions, such as psychosis (Adams et al. 2016) and autism (Constant et al. 2018a). Characterizing individual and group variations in the capacity to leverage TTOM, and the ways in which human agents adapt to their ecological niche, could reveal an important set of dimensions for such diagnostic frameworks. One could, for example, consider individuals who experience inference about the sort of person they and others are in a way markedly different from the neurotypical population (e.g., people with autistic traits). One could recruit participants who score high and low on the autistic spectrum, to test their relative ability to make inferences and predictions about others based on the ability to leverage information about gaze direction, or vary the context in which they deploy such inferences, to study the coupled dynamics between context and cognition that is typical to such individuals (Constant et al. 2018a).

Other conditions could be studied in this manner as well, shedding light both on TTOM as a general cognitive architecture and on these specific conditions. Higher rates of schizophrenia and psychosis among migrant populations might also be an excellent lens to approach such phenomena. Indeed, the careful study of such populations highlights the need for an interactional view of how sense of self and functioning may be destabilized by migration – to a new niche that has specific affordances for people of colour (Kirmayer & Gold 2011; Kirmayer et al. 2015). Depression might also be a useful phenomenon to consider, as it is an interactional phenomenon that involves complex inferences about self and other that is aggravated by retreat from the social niche, now perceived as lacking positively valenced affordances and occupied by other minds with intentions that are hard to understand, and which may in turn aggravate the condition itself (Baldwin 1992; Wang et al. 2008). This kind of work could inform a formal phenotyping of psychopathology based on the TTOM model.

Finally, although arguing for the applicability of the FEP to the puzzle of the acquisition of cultural practices, knowledge, and grammars, we caution against describing cultural ensembles as autonomous systems that maintain their organization and structural integrity through allostasis and homeostasis (Veissière 2018). Adaptation rests on an ongoing process of predicting events, engaging with the environment, and adjusting expectations in response to feedback from the world (including the body and other creatures). This occurs through constant transactions with the environment, and, in the case of human beings, that environment is fundamentally cultural and social – constructed with, and inhabited by, other people with whom individual agents must cooperate if they are to survive. This cooperation is itself patterned by cultural knowledge, skills, norms, institutions, places, and practices that have their own history and contingency.

Notes

1 There are many ways of interpreting this haiku by the modern poet Mayuzumi Madoka. The shift in gaze might be seen as an experience of erotic presence or represent an awakening to sexism and self-estrangement. It also recalls a culture-specific experience of the self as a performance (echoing the Japanese sense of always being on a stage; Heine et al. 2008). At its core, though, the poem powerfully illustrates the fundamentally human affective process of seeing and feeling oneself through the perspectives (and desires) of another.

2 Technically, an expectation corresponds to the average of a probabilistic belief or probability distribution. When the distribution is over (discrete) states of affairs, the expectation corresponds to the likelihood that any given state of affairs is true. Throughout, we will use beliefs in the sense of Bayesian belief updating or belief propagation, which could be either propositional or subpersonal in nature.

3 That is, the *act* of deploying precision weighting to select sources of sensory evidence, often discussed in terms of *mental action*.

4 The FEP is a variational principle of least action, like those that describe other systems with conserved quantities – for example, in the Lagrangian formulation of Newtonian mechanics, in which energy and momentum are conserved (Coopersmith 2017).

5 Intrinsic motivation is commonly used in developmental robotics to describe the epistemic value that reduces uncertainty (i.e., promotes information gain). In active inference, salience scores the reduction in uncertainty about transient states of the world, whereas novelty scores the reduction in uncertainty about the more stable parameters of a generative model. In short, *salience* is to *inference* as *novelty* is to *learning*.

6 The epistemic, uncertainty-reducing aspect of this formulation comes to the fore when human agents need to figure out what to do, more so than when agents are simply acting in accordance with the regimes of attention that they have internalized through enculturation.

7 With reference to the works of psychologists, Elizabeth Spelke, who documents infant "core knowledge" in the domains of intuitive physics, intuitive biology, and intuitive psychology, and Carol Dweck (Dweck 2013; Johnson et al. 2007), who emphasizes the role of learning, experience, and rewards from adherence to social norms (Kinzler et al. 2007; Olson & Spelke 2008; Spelke & Kinzler 2007).

Acknowledgments. We thank Paul Badcock, Shaun Gallagher, Casper Hesp, Dan Hutto, Safae Essafi, Michael Kirchhoff, Sander Van de Cruys, Alan Jürgens, Thomas Parr, Ian Robertson, Ryan Smith, Anna Strasser, Auguste Nahas, Erik Rietveld, Jonathan St-Onge, Simon Tremblay, Jared Vasil, Eric White, Julian Xue, and all those present at the Naturally Evolving Minds conference at University of Wollongong (20–23 February 2018) for helpful discussions and comments. We also sincerely thank the editor, Barbara Finlay, and the anonymous reviewers who provided us with valuable feedback.

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 M. J. D. Ramstead), from a a grant from the Foundation for Psychocultural Research (S. P. L. Veissière). from an Australian Laureate Fellowship (Ref: FL170100160) (A. Constant) and by a Social Sciences and Humanities Research Council doctoral fellowship (Ref: 752-2019-0065) (AConstant), 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 (M. J. D. Ramstead), and a Wellcome Principal Research Fellowship (K. J. Friston - Ref: 088130/Z/09/Z).

Appendix

This appendix describes the free-energy principle in terms of a Bayesian mechanics that emerges from the existence of a Markov blanket in a random dynamical system at nonequilibrium steady state. A Markov blanket is a four-way partition of states that define a self-organizing system and its environment (i.e., a system that has self-organized to nonequilibrium steady state). This partition comprises *internal* and *external* states $\{\mu, \eta\}$ that are separated by blanket states $b = \{s, a\}$. In turn, blanket states are divided into *sensory* and *active* states. In brief, the Markov blanket allows us to talk about internal states *representing* external states in a probabilistic sense. Heuristically, this means that one can ascribe probabilistic beliefs to internal states, in the sense that they are about something – namely, external states. This

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

interpretation rests upon a *variational density* over external states that is parameterized by internal states:

$$\begin{aligned} \boldsymbol{\mu}(b) &\triangleq \arg\max_{\mu} p(\mu|b) \\ q_{\boldsymbol{\mu}}(\eta) &= p(\eta|b) \end{aligned} \tag{1.1}$$

This variational density arises in virtue of the blanket as follows: If we condition internal and external states on the blanket, then there must exist a most likely internal state for every blanket state. This means that there must be a conditional density over external states conditioned on that blanket state. At nonequilibrium steady state, the flow of internal and active states can be expressed as a gradient flow on the *same quantity* – namely, the surprisal (i.e., negative log likelihood) of states that comprise the system (Friston 2013). We will refer to internal and active states $\alpha = \{a, \mu\}$ as *autonomous* because they are not influenced by external states:

$$\begin{aligned} f_{\alpha}(s, \alpha) &= (Q_{\alpha\alpha} - \Gamma_{\alpha\alpha})\nabla_{\alpha}\Im(s, \alpha) \\ \Im(s, \alpha) &= -\ln p(s, \alpha) \end{aligned} \tag{1.2}$$

These two aspects of a Markov blanket underwrite a Bayesian mechanics, in which we can talk about internal states holding Bayesian beliefs about external states – and autonomous states acting on external states, under those beliefs. We will first look at the underlying formalism in terms of a free-energy lemma and its path integral form that speak to (1) the most likely flow of internal states (i.e., perception) and (2) the trajectory of active states (i.e., action).

Lemma (variational free energy): Given a variational density, $q_{\boldsymbol{\mu}}(\eta) = p(\eta|b)$, the most likely path of autonomous states, given sensory states, can be expressed as a gradient flow on a free-energy functional of systemic states, $\pi = \{b, \mu\} = \{s, \alpha\}$:

$$\begin{aligned} \boldsymbol{\alpha}[\tau] &= \arg\min_{\alpha[\tau]} \mathcal{A}(\alpha[\tau]|s[\tau]) \\ &\Rightarrow \delta_{\boldsymbol{\alpha}[\tau]}\mathcal{A}(\boldsymbol{\alpha}[\tau]|s[\tau]) = 0 \\ &\Rightarrow \dot{\boldsymbol{\alpha}} = (Q_{\alpha\alpha} - \Gamma_{\alpha\alpha})\nabla_{\alpha}F(s, \boldsymbol{\alpha}) \end{aligned} \tag{1.3}$$

This means the most likely path conforms to a variational principle of least action, where variational free energy is an upper bound on surprisal:

$$\begin{aligned} F(\pi) &\triangleq \underbrace{E_q[\Im(\eta, s, \alpha)]}_{Energy} - \underbrace{H[q_{\mu}(\eta)]}_{Entropy} \\ &= \underbrace{\Im(s, \alpha)}_{Surprisal} + \underbrace{D[q_{\mu}(\eta)||p(\eta|s, \alpha)]}_{Divergence} \\ &= \underbrace{E_q[\Im(s, \alpha|\eta)]}_{Inaccuracy} + \underbrace{D[q_{\mu}(\eta)||p(\eta)]}_{Complexity} \geq \Im(s, \alpha) \end{aligned} \tag{1.4}$$

This functional can be expressed in several forms; namely, an energy minus the entropy of the variational density, which is equivalent to the surprise associated with systemic states (i.e., *surprisal*) plus the KL (Kullback-Leibler) divergence between the variational and posterior density (i.e., *divergence*). In turn, this can be decomposed into the negative log likelihood of systemic states (i.e., *inaccuracy*) and the KL divergence between posterior and prior densities (i.e., *complexity*).

Proof: The most likely trajectory – that minimizes action – obtains when the random fluctuations about the flow take their most likely value of zero. By equation (1.2), the flow of the most likely autonomous states $\boldsymbol{\alpha} = \{\mathbf{a}, \boldsymbol{\mu}\}$ can be expressed as a gradient flow on surprisal or, by definition, variational free energy:

$$\begin{aligned} \boldsymbol{\alpha}[\tau] &= \arg\min_{\alpha[\tau]} \mathcal{A}(\alpha[\tau]|s[\tau]) \Rightarrow \\ \dot{\boldsymbol{\alpha}} &= (Q_{\alpha\alpha} - \Gamma_{\alpha\alpha})\nabla_{\alpha}\Im(s, \boldsymbol{\alpha}) \\ &= (Q_{\alpha\alpha} - \Gamma_{\alpha\alpha})\nabla_{\alpha}F(s, \boldsymbol{\alpha}) \end{aligned} \tag{1.5}$$

Where, for the most likely internal state, $\boldsymbol{\mu} \in \boldsymbol{\alpha}$:

$$F(s, \boldsymbol{\alpha}) = \Im(s, \boldsymbol{\alpha}) + \underbrace{D[q_{\boldsymbol{\mu}}(\eta)||p(\eta|s, \boldsymbol{\alpha})]}_{Divergence} = \Im(s, \boldsymbol{\alpha}) \tag{1.6}$$

The equivalence between variational free energy and the surprisal of systemic states follows from the definition of the variational density that renders the divergence zero.

Given this stipulative formulation of gradient flows under a Markov blanket, one can now use the path integral formalism to characterize the most likely path of autonomous states from any initial state.

Corollary (path integral formulation): Under some simplifying assumptions, the action of autonomous paths from any initial systemic state is upper bounded by expected free energy:

$$\mathcal{A}(\alpha[\tau]|\pi_0) \leq G(\alpha[\tau]) \tag{1.7}$$

Expected free energy is defined as follows:

$$\begin{aligned} G(\alpha[\tau]) &\triangleq \underbrace{E_q[\Im(\eta, s, \alpha_{\tau})]}_{Energy} - \underbrace{H[q_{\tau}(\eta)]}_{Entropy} \\ &= \underbrace{E_q[\Im(s, \alpha_{\tau})]}_{Expected\ surprisal} + \underbrace{D[q_{\tau}(\eta|s)||p(\eta|s, \alpha_{\tau})]}_{Expected\ divergence} - \underbrace{D[q_{\tau}(\eta|s)||q_{\tau}(\eta)]}_{Information\ gain} \\ &= \underbrace{E_q[\Im(s, \alpha_{\tau}|\eta)]}_{Ambiguity} + \underbrace{D[q_{\tau}(\eta)||p(\eta)]}_{Risk} \\ &\geq \mathcal{A}(\alpha[\tau]|\pi_0) \end{aligned} \tag{1.8}$$

The expectation in equation (1.8) is under the predictive density over hidden and sensory states, conditioned upon the initial systemic state and subsequent trajectory of autonomous states:

$$q_{\tau}(s, \eta) \triangleq p(s, \eta, \tau|\alpha[\tau], \pi_0) \tag{1.9}$$

The expected free energy in equation has been formulated to emphasize the formal correspondence with variational free energy in equation (1.4), where the complexity and accuracy terms become *risk* (i.e., expected complexity) and *ambiguity* (i.e., expected inaccuracy).

In summary, variational free energy is an upper bound on the surprisal of systemic states, and expected free energy is an upper bound on the action of autonomous states. On a conceptual note, the role of nonequilibrium steady state takes on a different aspect, depending upon whether the aforementioned variational dynamics are thought of in terms of gradient flows (i.e., the variational free-energy lemma) or as picking out the most likely paths (i.e., the path integral corollary).

From the point of view of a statistician, the gradient flow formulation regards the probability density at nonequilibrium steady state as a generative model – in other words, a probabilistic specification of the sensory impressions of external states hidden behind the Markov blanket. It is this dynamic that licenses an interpretation of self-organization in terms of statistical (i.e., approximate Bayesian) inference.

The picture changes when we consider the path integral formulation. Here, we are picking out trajectories of autonomous states (i.e., active and internal states) that are most likely under the generative model. On this view, the generative model can be regarded as some prior beliefs about the sensory states (and their external causes) that will be encountered in the future. In other words, the generative model prescribes the attracting set that the system will autonomously work towards – by apparently selecting the paths of activity that lead to these attracting states. This enactive perspective makes it look as if the generative model is no longer simply an explanation for sensory samples but a specification of the states to which a system aspires.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Open Peer Commentary

Thinking through prior bodies: autonomic uncertainty and interoceptive self-inference

Micah Allen[a,b,c], Nicolas Legrand[a], Camile Maria Costa Correa[a] and Francesca Fardo[a,d]

[a]Center of Functionally Integrative Neuroscience, Aarhus University Hospital, 8000 Aarhus, Denmark; [b]Aarhus Institute of Advanced Studies, Aarhus University, 8000 Aarhus, Denmark; [c]Cambridge Psychiatry, University of Cambridge, Cambridge CB2 8AH, UK and [d]Danish Pain Research Centre, Aarhus University Hospital, 8000 Aarhus, Denmark.
Micah@cfin.au.dk nicolas.legrand@cfin.au.dk correa@cfin.au.dk francesca@clin.au.dk https://the-ecg.org/

doi:10.1017/S0140525X19002899, e91

Abstract

The Bayesian brain hypothesis, as formalized by the free-energy principle, is ascendant in cognitive science. But, how does the Bayesian brain obtain prior beliefs? Veissière and colleagues argue that sociocultural interaction is one important source. We offer a complementary model in which "interoceptive self-inference" guides the estimation of expected uncertainty both in ourselves and in our social conspecifics.

In their impressive synthesis, Veissière and colleagues argue that enactive social interaction is a prime ground for generating higher-order prior beliefs (both implicit and explicit). We share this enthusiasm for social–cultural patterning of priors, and also their comprehensive embrasure of the enactive and embodied turn within the larger predictive processing movement (Allen & Friston 2018; Barrett & Simmons 2015; Gallagher & Allen 2018; Ramstead et al. 2019b; 2019c; Seth 2013). As they elegantly argue, ontogenetic development provides a wealth of knowledge about how to behave in a given context. It follows that this "duet for one" of mutual prediction not only constrains how we engage with others, but also our own self-inference (Friston & Frith 2015a; 2015b). As such the proposal that much of our "repertoire of prior beliefs" emerges from socio-cultural interaction and enactive, embodied engagement is both feasible and exciting.

However, we disagree that "information from and about other people's expectations constitutes the primary domain" from which prior beliefs about "statistical regularities" (i.e., expected precision) arise. Although socio-cultural sources certainly contribute, we highlight the predominance of more than 1 million years of phylogenetic evolution in shaping our "prior bodies" as the key constitutive factor molding how we predict ourselves and other agents (Allen & Tsakiris 2019). In advance of any ontogenetic development, one is born with homeostatic and morphological features which shape the expected statistics of one's life, and these, in turn, provide a rich generative model which can be inverted to understand a wide range of human social behaviors.

Developing this view, we recently proposed a computational model of interoceptive self-inference (Allen et al. 2019). Our model argues that the visceral body provides a fundamental constraint on belief precision, and that interoception serves to sample these rhythms so as to better estimate expected uncertainty. This model formalizes other more conceptual accounts of interoceptive inference in the domain of emotion (Barrett & Simmons 2015; Chanes & Barrett 2016; Seth 2013; Seth & Tsakiris 2018), selfhood (Ainley et al. 2016; Apps & Tsakiris 2014; Limanowski & Blankenburg 2013), and metacognition (Petzschner et al. 2017). Our model generalizes beyond these to argue that the primary homeostatic rhythms of the body fundamentally constrain prior beliefs about the precision, or confidence, of both interoceptive and exteroceptive belief updates. This to say, visceral rhythms embed primary control dynamics, or hyper-priors, on the agent's landscape of precision. This, in turn, dictates the confidence I assign to any shift in my posterior beliefs and provides a useful starting point for estimating self-precision in others.

To illustrate the core of our model (summarized in Fig. 1), consider the following example of sensory attenuation in the retina. In the eye, the pulsation of blood across the optic disk at each heartbeat distorts the retinal surface, briefly occluding ascending sensory information. In a hierarchical brain, this predictable fluctuation of precision is a crucial learning signal, which can be sampled via interoception to improve estimates of expected uncertainty (Parr & Friston 2017a; Pulcu & Browning 2019). Expected uncertainty, in turn, provides an invaluable control signal dictating how much I should update my beliefs in the face of new information; Bayesian decision theory tells us that when confronted with a volatile environment (or a volatile colleague), one should more rapidly update their beliefs in response to prediction error.

Through simulation, we show that this simple coupling of sensory precision to the rhythm of homeostasis enforces a primary interaction between the body and our perception of the world. In our model, lesioning afferent viscerosensory information caused a cascade of interoceptive prediction error which elicit psychosomatic hallucinations, blunted belief updating, and attenuated physiological reactions. These domain-general alterations in precision ultimately cause agents to update their higher-order beliefs, resulting in top-down metacognitive biases (i.e., misestimation of expected uncertainty) that characterize many psychiatric and psycho-social illnesses (Lawson et al. 2017; Powers et al. 2017). In contrast, maladaptive prior beliefs about self-uncertainty can elicit misperception or hyper-arousal in the interoceptive domain. This equips the model with a deeply circular, enactive causation; my expectation of confidence in the world constrains my visceral inference and regulation, and the statistics of visceral rhythms constrain my exteroceptive percepts and beliefs.

Here perhaps is where there is the most potential for crosstalk between the model of Veissière and colleagues and interoceptive-self inference. Our model suggests that agents are imbued at birth with a repertoire of "embodied priors" or statistical regularities dictated by their morphological forms, which act as hyper-parameters on meta-cognition and learning (Allen & Tsakiris 2019). In particular, these priors influence the confidence or precision dictating the perceptual and emotional salience we assign to various interoceptive and exteroceptive outcomes. The notion of *variational niche construction* developed by Veissière and others can be cast as building ontogenetic refinement in addition to these fundamental constraints (Bruineberg et al. 2018a; 2018b). That is to say, in thinking through other minds we demarcate novel boundaries of salience, refining the embodied set-points

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

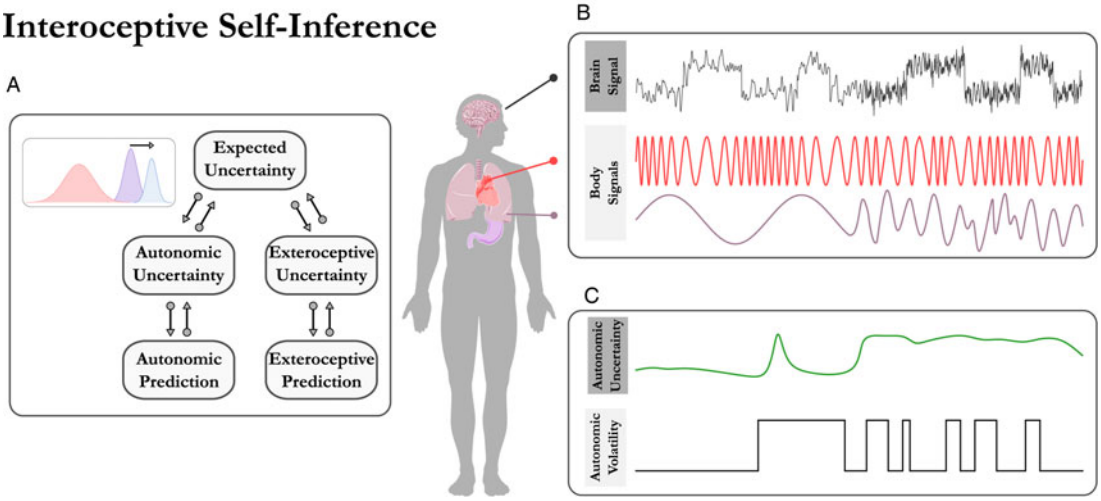


Figure 1. (Allen et al.) Interoceptive self-inference model. (A) Hierarchical precision-weighted inferences integrate confidence signals from the internal and external environment into an overall estimate of expected uncertainty. (B) For example, slow-respiratory oscillations stabilize cardiac cycles, resulting in low-autonomic uncertainty. In contrast, a volatile breath pattern increases baseline neural uncertainty, as illustrated in a simulated brain response to a steady state exteroceptive input. (C) These fluctuations can be modeled, for example by a dynamic reinforcement-learning approach in which the volatility of interoceptive state transitions inflates the estimate of autonomic uncertainty. Through inversion of the self-inference model to conspecifics, agents can predict the confidence of others' beliefs.

that define a landscape of precision for agents. We maintain, however, the hegemony of the phylogenetic body in setting these starting points; ultimately, the strongest possible source one can sample from concerning the volatility of others is found within oneself.

Acknowledgments. MA is supported by a Lundbeckfonden Fellowship (R272-2017-4345), and the AIAS-COFUND II fellowship program that is supported by the Marie Skłodowska-Curie actions under the European Union's Horizon 2020 (Grant agreement no. 754513), and the Aarhus University Research Foundation.

Thinking *with* other minds

Edward Baggs[a] and Anthony Chemero[b]

[a]The Rotman Institute of Philosophy, Western University, London, Ontario N6A 5B7, Canada and [b]Departments of Philosophy and Psychology, University of Cincinnati, Cincinnati, OH 45221.
ebaggs@uwo.ca chemeray@ucmail.uc.edu

doi:10.1017/S0140525X19002747, e92

Abstract

We applaud the ambition of Veissière et al.'s account of cultural learning, and the attempt to ground higher order thinking in embodied theory. However, the account is limited by loose terminology, and by its commitment to a view of the child learner as inference-maker. Vygotsky offers a more powerful view of cultural learning, one that is fully compatible with embodiment.

Embodied approaches to cognition are compelling when they are invoked as explanations for perceptually controlled action phenomena: how do people catch balls, or continuously maintain an upright posture – that kind of thing. A challenge is to extend this kind of explanation to an account of "higher order" tasks that are characteristic of human behavior, that is, those involving language and social interaction. The authors seek to rise to this challenge. They offer an account of cultural learning based on a predictive coding view of cognition. We applaud the ambition of their project. We think the authors are right to ground this project in an account of learning, and we agree that the account must begin with a conception of the child as an embodied agent.

In its details, though, the proposal is rather limited. The authors' stretching of conceptual terminology leads them to pursue a solution that remains fundamentally disembodied and inferential. And they overlook a long-standing approach which offers a promising solution to the problem they are addressing.

First, the vague terminology. The authors are too flexible in their use of terms. Particularly problematic is the word "expectation." The authors use the word interchangeably to mean both a prediction generated by the brain of an actor as part of its perceptual foraging, and socially-enforced norms of "expected" behavior, for example, that one should sit up straight. The problem here is that the latter meaning is the thing the authors are trying to explain. By using the same term interchangeably for both meanings, they are effectively presupposing the solution to the problem. (Such ambiguity in the use of terms seems to be a recurring problem in predictive coding accounts, see Anderson & Chemero 2013.)

The authors attempt to address this problem by appealing to the concept of affordances, but in doing so stretch that concept beyond its useful boundaries. Just as the action possibilities of objects can be described in terms of affordances, the authors suggest, so *culturally appropriate* behavior can be described in terms of *cultural* affordances. Unfortunately, affordances don't work as explanations for the things the authors seek to explain, namely: How individual humans learn to act appropriately according to the prevailing mores of their neighbors. This is not what the concept of affordances was invented for. The concept was originally intended to resolve a dualism in the philosophy of perception

between phenomenal and physical worlds (Heft 2017). It achieved this by suggesting that meaning resides not inside the organism but in the fit between an organism and its surroundings, specifically in the fit between the organism and some structure that the organism can act upon. In practice, it has proved remarkably difficult to leverage this concept as part of a fruitful empirical program, even when the target of explanation is a single actor performing a simple task (see Wagman 2019 for a review). It is much more difficult to apply the concept to culture, and to our knowledge no one, including the current authors, has succeeded in doing so non-metaphorically. You can say "cultural affordances" all you want, but that does not constitute an empirical research program.

To solve the problem the authors set out to solve, they need an account of how you get from situated action in the infant to enculturated behavior in later childhood and adulthood. That is, of how you get from one kind of expectation to the other, from the if-I-put-my-hand-in-that-fire-it-will-be-hot type of expectation to the children-should-sit-up-straight type of expectation. It is not enough to say that the child simply learns a set of behavioral norms by observing adults. The question is: how?

Essentially, the authors think learning is inference – another term whose meaning is stretched – and view the child as a scientist in the crib. It is here that their account remains a disembodied one. It has been pointed out that to view the child learner as fundamentally an inference-maker is to inappropriately project onto the child the worldview of the adult (and, traditionally, male) scientist: In pursuit of experimental control, the scientist imposes certain restrictions on himself in how he can gather data about the world, and he then wrongly assumes that the child is subject to similar restrictions (Donaldson 1978). This sort of account acts as if the child learns everything on her own, without scaffolding provided by others.

It doesn't have to be this way. A compelling non-inferential account of cultural learning already exists, and the authors overlook it. This is the cultural historical approach associated with Vygotsky and later activity theorists (Cole 1996; Leont'ev 1974; Vygotsky 1962; 1978). On this view, the child is not an incomplete adult who must collect a set of cultural norms before she is able to act in a fully adult manner. Rather, the child is already a complete actor. She continuously learns new skills as part of her coping with situations involving others. Crucially, social activity comes first. Language starts off as concrete acts of speaking and being spoken to. It is only later, after the child has developed some linguistic competence in the presence of others that she can engage in private mental cognition: construct an internal narrative, engage quietly in counterfactual reasoning, and develop inferences and scientific theories.

A key argument for Vygotsky was that the acquisition of language (and culture) leads to a qualitatively different form of cognition: Norms are no longer imposed exclusively by evolutionary fitness but can be self-imposed through language. Given this view of cultural learning, the present "variational" account of higher cognitive functions misses the point: Language-involving cognition operates according to a different set of norms, and is not merely a more elaborated form of adaptive fitness.

The Vygotskyan view is, we submit, a more attractive one. It provides an escape from the adaptationist trap that the current proposal falls into. It is also compatible with an embodied, socially embedded understanding of human behavior. Instead of seeing other minds in instrumental terms, as things we think *through*, it sees other minds as people we can think *with*.

"Through others we become ourselves": The dialectics of predictive coding and active inference

Dimitris Bolis[a,b,c] and Leonhard Schilbach[a,b,d]

[a]Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, 80804 Munich-Schwabing, Germany; [b]International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany; [c]Munich Medical Research School (MMRS), Dekanat der Medizinischen Fakultät, Ludwig-Maximilians-Universität München, 80336 Munich, Germany and [d]LVR Klinikum Düsseldorf/Kliniken der Heinrich-Heine-Universität Düsseldorf, 40629 Düsseldorf, Germany.
dimitris_bolis@psych.mpg.de https://sites.google.com/site/dimitrisbolis/
leonhard_schilbach@psych.mpg.de
https://www.leonhardschilbach.de/english.html

doi:10.1017/S0140525X19002917, e93

Abstract

Thinking through other minds creatively situates the free-energy principle within real-life cultural processes, thereby enriching both sociocultural theories and Bayesian accounts of cognition. Here, shifting the attention from thinking-through to becoming-with, we suggest complementing such an account by focusing on the empirical, computational, and conceptual investigation of the multiscale dynamics of social interaction.

We applaud Veissière and colleagues for pursuing the ambitious goal of situating the free-energy principle within the context of sociocultural processes. This is, indeed, a much needed undertaking, which has only recently started developing, holding promise for advancing not only relevant sociocultural research fields, but also computational psychiatry (cf. Bolis & Schilbach 2017; 2018b; Constant et al. 2019b; Friston & Frith 2015b; Gallagher & Allen 2018). In fact, human cognition and culture have often been studied in isolation. For instance, the field of computational psychiatry has been developing rigorous experimental protocols and mathematical toolboxes to mechanistically explain human cognition and action. Yet, until recently a rather individualistic perspective has been adopted, which neglects levels of description beyond the individual (cf. Bolis et al. 2017; De Jaegher & Di Paolo 2007; Kirmayer & Crafa 2014; Schilbach et al. 2013). On the other hand, sociocultural fields, such as cultural anthropology, have rightfully adopted a more holistic perspective to complex phenomena of life, yet frequently lacking formal descriptions of cognitive and biological mechanisms (cf. Seligman & Brown 2009).

An artificial dichotomy between the individual and the collective has inevitably led to a "chicken-egg" paradox (cf. Dumas et al. 2014). However, such causality dilemmas dissolve once one considers the dialectical nature of human-*becoming*, which is multiscale, reciprocal, dynamic, cumulative, and inherently contradictory (cf. Bolis & Schilbach 2018b; Di Paolo et al. 2018; Dumas et al. 2014; Vygotsky 1930–1935/1978). Processes from evolution and culture to individual development, learning, and sensorimotor activity, can all be viewed as mutually interacting adjustments between the species and the environment. Here, reciprocity is deep, as "it is not only humans who change the

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

environment, but the environment in turn changes them in face of their impact on it" (Bolis & Schilbach 2018b; Levins & Lewontin 1985). TTOM, therefore, constitutes an important development because it addresses how human agents learn shared expectations and how they construct their own social niches in complex interaction between the individual and the environment.

We concretely appreciate the consideration of predictive coding and active inference within a framework of circular causality. Indeed, an organism can be viewed as embedded within the dialectic between the two above-mentioned processes, which in order to survive obeys a simple, but fundamental rule: "adjust yourself to reality or change the reality itself" (Bolis & Schilbach 2018b; Friston 2010). When it comes to TTOM, it is not only the agent which learns environmental regularities and adjusts accordingly, but also the environment in turn "learns" the agents' "beliefs" through repeated and culturally regulated actions. TTOM resonates well with the dialectical attunement hypothesis (Bolis & Schilbach 2018b), which views human-becoming as the interplay between internalization and externalization in and through (culturally-mediated) social interaction; internalization being the "co-construction of bodily hierarchical models of the (social) world and the organism" [cf. predictive coding], whereas externalization the "collective transformation of the world" [cf. active inference]. In a nutshell, "interpersonal statistical regularities shape multiscale hierarchical models on an individual level and vice versa."

To offer a formal description of how the environment "learns," the authors interestingly suggest twisting the modeling equations by inverting relevant quantities across actions and sensations. This offers various potential modeling scenarios about the degree of interactivity within the system of brain–body–environment–body–brain (cf. Froese et al. 2013). Here, a multiscale meta-Bayesian scheme might nicely lend itself for modeling not only individual processes, but also collective and environmental interactions (Bolis & Schilbach 2017; Brandi et al. 2019; Ramstead et al. 2018).

Not only are we in line with the authors on conceptual and computational grounds, but also concerning the need for empirical studies. To make this more concrete, we describe certain experimental directions: Systematically varying social structure, cultural, and socioeconomic background, affective bonds and interpersonal similarity across interacting individuals will enable the mechanistic study of interpersonal attunement. With regard to psychiatric disorders, construed as disorders of social interaction (Schilbach 2016), two-person (or indeed collective) psychophysiology allows to move beyond the individual (cf. Bolis & Schilbach 2018a). Taking autism as a paradigm example, the dialectical misattunement hypothesis has put forward a research line, which, moving away from an exclusive study of individual differences, considers types of interacting groups: that is, autistic, neurotypical, and mixed groups, expecting smoother interactions within the more homogeneous groups or dyads (Bolis et al. 2017). Taken together, such experiments will not only inform TTOM within the "neurotypical social world," but also open up avenues for evaluating and updating the ontological status of conditions, such as autism, as relational and interactional (cf. double empathy problem; Milton 2012).

Apart from praising TTOM, we would also like to point out a fundamental aspect which, in our opinion, would benefit from further elaboration. We feel that the potentially constitutive role of *real-time* social interaction in sense-making and human-becoming was not sufficiently taken into account within the model (cf. De Jaegher & Di Paolo 2007; Vygotsky 1930–1935/1978). It has been suggested that thinking about and with others might be fundamentally different in real-time interactive scenarios, as compared to passive observational situations (cf. second-person perspective; Redcay & Schilbach 2019; Schilbach et al. 2013). Crucially, such interactive interpersonal processes have been thought of as dialectically preceding the individual both in evolutionary and developmental regards (cf. Bolis & Schilbach 2018b; Tomasello 2019). As Vygotsky proclaimed almost a century ago, "through others we become ourselves" (Vygotsky, 1931/1987). Yet, to do justice to the authors, the field today has not yet reached a conclusive consensus. For instance, although Di Paolo et al. (2018) suggest that "interactive situations present a richer, more complex set of possibilities" and "the key to our sociality is not in our heads or in our genes," Schönherr and Westra (2017) claim to have (conceptually) shown that "ersatz interactivity works just as well as the real thing," by "real thing" denoting genuine, real-time social interaction. We, therefore, conclude our commentary with a question still desperately begging for a definite empirical answer. *Does (real-time) social interaction matter… or is it all in our heads?*

Have we lost the thinker in other minds? Human thinking beyond social norms

Nabil Bouizegarene

Department of Psychology, Université du Québec à Montréal, Montreal, Quebec H3C 3P8, Canada
bouizegarene.nabil@uqam.ca

doi:10.1017/S0140525X19002875, e94

Abstract

Veissière and colleagues suggest that thinking is entirely based on social norms. I point out that despite the fact that social norms are commonly used to alleviate cognitive processing, some individuals are willing and able to go about the costly process of questioning them and exploring other valuable ways of thinking.

The framework introduced by Veissière and colleagues provides a compelling account of enculturation as a process of thinking through other minds (TTOM). However, their account may fall short of providing a thorough model of cognition. I agree with the authors' proposal that humans outsource most of their thinking unto other minds in order to minimize cognitive load, essentially by engaging others' expectations and inferences. However, I find their claim that all solitary human thinking is "social through and through" (sect. 3.6, para. 2) rather extreme and wish to highlight ways in which it could be nuanced. I will present evidence from research on identity processes which suggest that individuals vary in the degree to which they adhere to social norms and the expectations of others when they engage in identity construction.

Berzonsky's research on identity processing styles (Berzonsky 1989; 2011) identified individual differences in the tendency to use normative information for self-definition processes. He showed

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

that individuals with a normative identity style tend to base their self-view on the relevant social norms and standards that their group and significant others value and adopt. In contrast, individuals with an informational identity style come to think about themselves on the basis of a complex exploration process. That is, they actively seek out self-relevant information by testing out beliefs, activities, and interests and assessing the degree to which they fit to themselves (e.g., Do I really want to be a lawyer?). In contrast, individuals high in normative identity style tend not to question their culturally prescribed commitments (e.g., of course, like my father, I will be a lawyer because it is a well-paid and respectable job).

The fact that individuals with a normative identity style prefer to cut short the taxing process of deeply questioning and exploring their identity is consistent with Veissière and colleagues' framework, in that norms are herein recruited as a short-cut to a great deal of uncertainty (or free energy). Consistent with this idea is the finding that normative identity style is related to higher levels of need for structure and need for cognitive closure (Soenens et al. 2005). However, I am not sure that informational identity style could be understood through Veissière and colleagues' framework. How should we understand that some cultured agents seek and tolerate the uncertainty of questioning their identity beyond social norms and voluntarily go about a long process of thinking autonomously about themselves, rather than using norms as an antidote to this uncertainty?

Mclean and colleagues (McLean & Syed 2015; McLean et al. 2017) make similar observations in their research on identity development, which focus on the relationship of identity and society in personal narratives. This team focuses on the narratives that are the cultural templates for the experiences one should expect to have in their lives, which they call master narratives. They define the latter as shared narrative expectations regarding what is a culturally valued biography. They found evidence that individuals develop their identities by negotiating the degree to which these narratives are maintained or changed when individuals create their own life story.

A particularly relevant result is that individuals who develop alternative narratives (i.e., changed relative to the master narrative) are also engaged with more identity work (McLean et al. 2017). Specifically, it was found that those who develop alternative narratives made a greater number of explicit connections between life events and their selves and displayed higher levels of identity exploration. These results are consistent with Veissière and colleagues' framework because they suggest that identifying with cultural norms requires less effort. However, these results also challenge the TTOM framework because they suggest that some individuals decide to exert the effort of developing alternatives to these cultural norms.

Social norms are attractive because they provide ready-made answers to the difficult and urgent questions we face throughout our lives. This may explain why conformism is endemic but does not preclude that some individuals are willing and able to go about the costly process of questioning these social norms. Furthermore, this questioning might be an essential part of the iterative process underlying the cumulative culture phenomena described by Tomasello et al. (1993). Individuals thinking about norms in a unique and original way instead of just blindly assimilating them may catalyze the generation of useful ideas and solutions that are integrated in culture and passed on to future generations. The adaptiveness of today's culture may owe a lot to individuals in past generations who distanced their thinking from their culture. If we lose the thinker in others mind, we may lose much of the adaptive potential of culture.

Unification at the cost of realism and precision

Rachael L. Brown[a], Carl Brusse[a,b], Bryce Huebner[c] and Ross Pain[a]

[a]School of Philosophy, the Australian National University, Canberra, 0200 ACT, Australia; [b]Department of Philosophy and Charles Perkins Centre, The University of Sydney, Sydney, NSW 2006, Australia and [c]Department of Philosophy, Georgetown University, Washington, DC 20009.
rachael.brown@anu.edu.au http://rachaelbrown.net
carl.brusse@sydney.edu.au
bryce.huebner@georgetown.edu brycehuebner.weebly.com
ross.pain@anu.edu.au

doi:10.1017/S0140525X19002760, e95

Abstract

Veissière et al. must sacrifice explanatory realism and precision in order to develop a unified formal model. Drawing on examples from cognitive archeology, we argue that this makes it difficult for them to derive the kinds of testable predictions that would allow them to resolve debates over the nature of human social cognition and cultural acquisition.

Veissière et al. have uncovered an interesting set of high-level regularities, which appear to show up wherever humans attempt to calibrate their behavior against one another. They have also shown that the FEP provides a unified mathematical framework that is useful for describing these regularities. Highly general and unified explanatory models such as TTOM can be extremely useful. For example, where a discipline lacks a common theoretic language for describing competing perspectives, such models can be deployed to *dissolve* disputes by bringing rival positions under a single theoretical framework. Veissière et al. assume that such a strategy will prove fruitful in *resolving* persistent disagreements within the cognitive science of cultural acquisition and social cognition, as TTOM seems to provide a unified framework for characterizing insights from a number of otherwise incommensurable theories. We applaud their attempt to provide a more unified account of social cognition and the acquisition of culture; but we contend that bringing these phenomena under a single mathematical framework is unlikely to resolve the relevant disputes.

Providing a simple, overarching characterization of complex and inherently variable biological systems is challenging. Abstract mathematical models of biological phenomena, such as TTOM, attempt to overcome this challenge by prioritizing explanatory generality over competing ideals such as explanatory precision and biological realism (Levins 1966). If successful, this strategy can offer unifying explanations of seemingly disparate biological phenomena, such as the action of different but analogous biological systems, or of heterogeneous parts of the same system. Yet, unification comes at the cost of explanatory realism and precision. In order to draw parallels between non-identical systems, general models must make idealizing assumptions about patterns of biological variation as well as the causal specificities of the particular systems being described. This allows such models to capture the general properties of a system, by focusing on

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

broadscale similarities. But as a result, they fail to be entirely precise and accurate when it comes to the particularities of the system (Woodward 2005). The mathematical framework provided by the FEP does give TTOM a high level of generality; but we worry that this involves stripping away fine-grained causal details and evolutionary histories without much obvious explanatory pay-off.

This is not to deny that unification can offer new perspectives, but we doubt that there is more to say about social cognition and cultural acquisition at the highly abstract level afforded by the FEP than is already being said at a less general, but causally richer, level of description. This concern might be mitigated if TTOM succeeded in providing a common framework for usefully describing and comparing competing theories in cognitive science, but we worry that any theoretical unification achieved via TTOM will be more perspectival than substantive, as the unification it provides is generated by looking at the issues from a level of abstraction that makes the details disappear. Long-standing debates in cognitive archeology illustrate these problems nicely.

The story one tells about the evolution of hominin cognition is highly dependent on the position one adopts on social cognition. Debates between dynamicists/externalists (Malafouris 2016; Noble & Davidson 1996; Overmann 2016; Tomlinson 2015) and representationalists/internalists (Cole 2016; Coolidge & Wynn 2018; Mithen, 1996) in cognitive archeology mirror broader debates in cognitive science. For instance, Noble and Davidson (1996) employ an externalist and Gibsonian approach to the analysis of stone-tools and the evolution of social cognition, whereas Mithen (1996) employs an internalist and modular approach. If TTOM provides a tool for resolving debates in cognitive science, it should also offer the resources for arbitrating between these different views, and for finding a clear route to a resolution. Unfortunately, even if TTOM can express these rival accounts in the general, abstract, mathematical language, this redescription seems to add little to our existing, much richer causal understanding of the systems in question.

Debates about hominin cognitive evolution largely concern the kinds of cognitive traits that are required to produce lithic technologies. And resolving such debates requires generating mutually exclusive and testable empirical predictions to compare against the Paleolithic record and findings in contemporary cognitive science; any common vocabulary for comparing theories must be causally rich enough to engage with such evidence. Unfortunately, TTOM is so abstract and multiply realizable that the evolutionary histories and fine-grained causal information that instantiate the competing views about hominin cognitive evolution are largely omitted. Given this causal frugality, TTOM seems incapable of generating the testable predictions cognitive archeologists require to resolve these debates, and hence the overall payoff for deploying it is unclear.

We suspect that the state of affairs in cognitive archeology is a reflection of broader debates in the study of human social cognition and cultural acquisition. Recent experiments have revealed significant intra- and inter-personal variation in mentalizing capacities (e.g., Warnell & Redcay 2019); this may reflect heterogeneity in the underlying biological systems (Schaafsma et al. 2015), or it may suggest the development of different kinds of sense-making strategies (De Jaegher 2013). An approach that focused on patterns of qualitative variation might yield empirically tractable predictions in this domain; and given a plausible set of bridging principles, the resulting data may be useful for adjudicating the relevant disputes. By contrast, the unified theoretical framework advanced by Veissière et al. can only reveal the points where these different kinds of approaches are likely to converge. As we see it, TTOM mistakenly equates formal unification with explanatory power. Explanation in science is, alas, far more complex; and generality comes at the cost of valuable explanatory realism and precision. In light of this worry, we contend that the explanatory value which TTOM appears to have is likely to reflect its ability to systematize existing data, rather than its ability to produce novel hypotheses, or novel ways of negotiating intractable disputes.

Normativity, social change, and the epistemological framing of culture

Andrew Buskell

Department of History and Philosophy of Science, University of Cambridge, Cambridge CB2 3RH, UK.
ab2086@cam.ac.uk

doi:10.1017/S0140525X19002681, e96

Abstract

The authors deploy an epistemic framework to represent culture and model the acquisition of cultural behavior. Yet, the framing inherits familiar problems with explaining the acquisition of norms. Such problems are conspicuous with regard to human societies where norms are ubiquitous. This creates a new difficulty for the authors in explaining change to mutually exclusive organizational structures of human life.

Thirty years of work in cultural evolution, primatology, human behavioral ecology, and cognitive science has established a consensus framework for understanding culture. This framework characterizes agents as voracious epistemic optimizers: Individuals who exploit cues, adopt strategies, and intervene on situations to extract high-quality information relative to their goals. This framing extends out into the world, seeing it in epistemic terms: the physical environment and other agents are repositories of, and instruments for, information acquisition.

Veissière et al. (hereafter, "the authors") adopt and synthesize this consensus framework. Their particular concern is the thoroughly social character of the informational world in which humans develop and live – and their novel contribution is to wed empirical research on this topic with the apparatuses of variational Bayesian inference and the free-energy principle. These tools, they suggest, provide means of modeling key features of enculturation, behavior, and cultural change.

Yet, an important feature of this account needs to be noted at the outset. The authors' epistemic framing grounds both culture and enculturation in the extraction and employment of information, and in so doing, minimizes the explanatory clout of other core aspects of human life; notably, deliberative choice, affect, and normativity. On the authors' account, these latter features are either reducible or subsidiary to variational inference. The result is a conservative model of culture, one already well-formulated by David Laitin (2007): cultures are "circumstances in which members of a group […] are able to condition their behaviour on common knowledge beliefs about the behaviour of all members of the group" (p. 64). Such a model renders cultures largely homeostatic; enculturation means both learning and expecting others to stay within the bounds of established behavior.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

No doubt models couched within this framework are useful for understanding cultural phenomena; for instance, explaining how adolescents flexibly adopt different learning strategies to become capable members of the group (e.g., Salali et al. 2019). But, as a general account of culture, the picture is likely impoverished. Enculturated agents do not just think through other agents in an epistemic way, as something to be conditioned over, but in a normative and affective way: A confession of love can be terrifying even before hearing the response; and even thinking about committing a crime can bring feelings of disquiet and shame. Norms, in other words, are not mere conventions – subject to more or less accurate conditionalization – but, affectively motivated standards for appropriate behavior. More importantly, they are ubiquitous structures of human sociality.

Normative and affective features are difficult to grasp using the language of epistemic optimization. Consider a key piece of the authors' machinery; cultural, or "epistemic" affordances. Affordances are relationships between features in the environment and agential behavioral repertoire (Chemero 2009; Walsh 2015). Although some researchers do link affordances to normatively freighted concepts such as "skill" (e.g., Rietveld & Kiverstein 2014), this is not because the two are identical. They are not. Indeed, the ubiquity of human norm-governedness is a puzzle, and there is a growing literature aimed at exploring just how affordance-structured ape cognition could be bootstrapped and changed into norm-rich human cognition (Birch, m.s.).

Though the authors do mention normativity, it is as a *fait accompli* – something that comes for free with variational inference in a structured environment. But, this is where one is owed further explanation: How are affordances bootstrapped into skills? How does variational inference explain how human beings are motivated to act in accordance with rules? In short, how epistemic optimizing explain norm-governed behavior? Although the authors show some sensitivity to these issues, they owe us an account that explains the development of such capacities in inferential terms.

This line of thinking leads to broader concerns with how the Bayesian framework construes and understands culture. Cultural groups are governed by local normative structures that set standards for behavior and regulate social interaction. Yet these normative frameworks are continually contested and subject to change, and this change often involves deliberate attempts on the part of individuals to reorganize governing structures. Whether this involves shifts in kinship or marriage practices, corporate or political governance, religious or ritual practices, agents often act collectively to change the normative strictures guiding their behavior. Yet the epistemic account on offer leaves little explanatory space to account for such social change. This is especially true of the core organizational structures of human sociality.

The authors' account of social change ties it to exploration and the acquisition of information. Agents are motivated by "intrinsic, epistemic imperatives" to acquire information before engaging in actions with pragmatic value – turning on the lights to avoid stumbling on the furniture, for instance. Nonetheless, the authors suggest, these epistemic imperatives have the added benefit of promoting curiosity and by extension, enlarging the "repertoire of responses for the individual or the group" (sect. 3.3, para. 6).

Yet, even if human beings were motivated by such imperatives, this could not be the whole story. For many core organizational features of human life are exclusive: it is either possible to marry your cross-cousin or impossible; you either pay brideprice or a dowry; you either live in a big man society or you do not. Transitions between these organizational states are complex – and are often the result of deliberative, communal experimentation (Wengrow & Graeber 2015) – but it is difficult to see how epistemic imperatives and enlarged "repertoires of responses" get a grip on these changes. What would it mean to get information about breaking taboos on marrying cross-cousins without actually marrying a cross-cousin? What about motivating systematic change in governance? How could one motivate pragmatic action without knowing what comes next?

As noted, the authors' account is conservative – focusing on how to things keep going on as before. But, radical organizational change seems to be ubiquitous among human cultures (Mace & Jordan 2011), suggesting that the authors' model should be open to such phenomena. Yet, such openness will require greater attention to the role of affect, normativity, and communal decision-making – and this in turn may require going beyond Bayesian variational inference and the epistemic framing of human beings and the world.

The multicultural mind as an epistemological test and extension for the thinking through other minds approach

George I. Christopoulos[a] and Ying-yi Hong[b]

[a]Culture Science Innovations, Nanyang Business School, Nanyang Technological University, Singapore 639798, Singapore and [b]Department of Marketing, The Chinese University of Hong Kong, Shatin, Hong Kong.
cgeorgios@ntu.edu.sg www.deonlabblog.com
yyhong@cuhk.edu.hk http://www.yingyihong.org/hong-ying-yi.html

doi:10.1017/S0140525X19002711, e97

Abstract

The multicultural experience (i.e., multicultural individuals and cross-cultural experiences) offers the intriguing possibility for (i) an empirical examination of how free-energy principles explain dynamic cultural behaviors and pragmatic cultural phenomena and (ii) a challenging but decisive test of thinking through other minds (TTOM) predictions. We highlight that TTOM needs to treat individuals as active cultural agents instead of passive learners.

> *…I was constantly referring my new world to the old for comparison, and the old to the new for elucidation…It is painful to be conscious of two worlds.*
> — Mary Antin, Russian Jew immigrant to USA (p. 3) (Antin 1912)

Thinking through other minds (TTOM) offers an excellent, multi-level account aiming at explaining how individuals learn a specific culture – typically, the culture that the individual is exposed to during childhood. This framework remains to be tested and improved against a wider range of cultural phenomena.

Culture is not monolithic – in fact, individuals could acquire knowledge of multiple cultures and carry multiple cultural identities. Many cultural phenomena are essentially dynamic and involve drastic changes, including within the individual (Fig. 1).

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

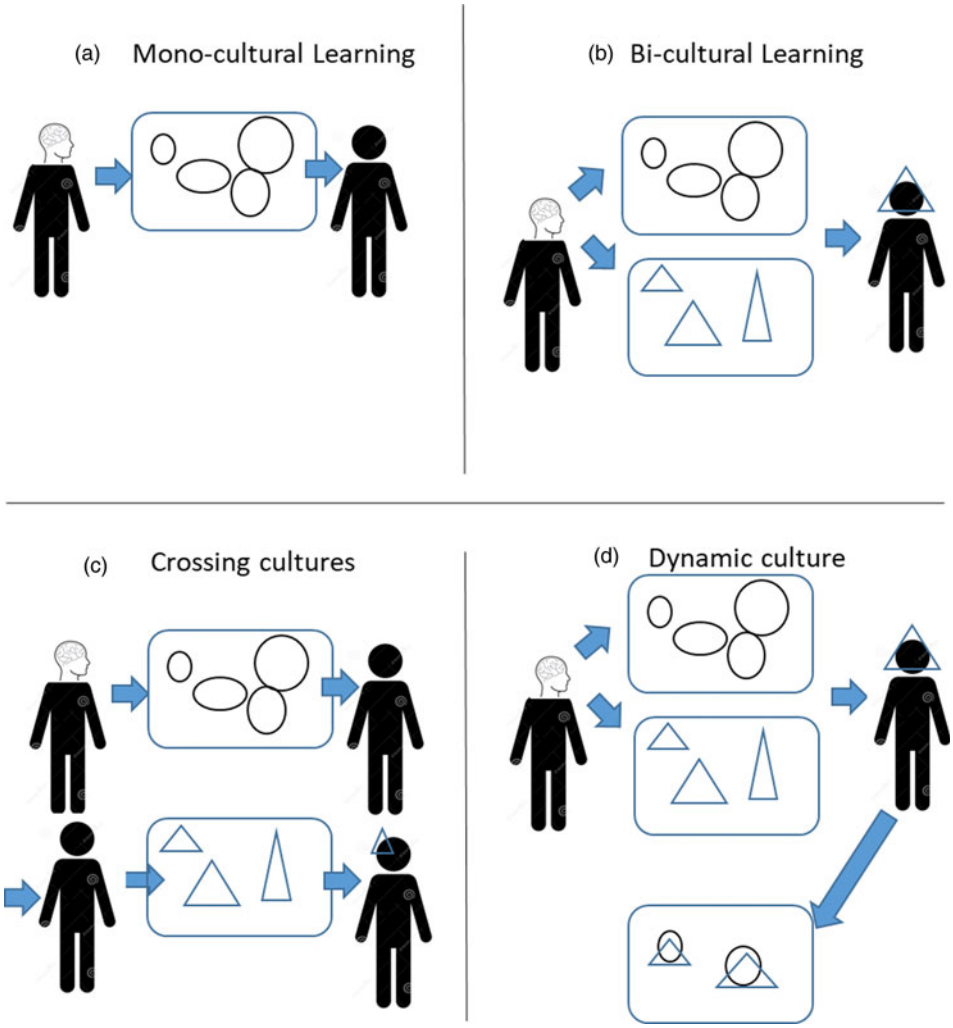


Figure 1. (Christopoulos & Hong) Four types of cultural learning. (a) Monocultural: learning one culture ("circle" culture). (b) Bicultural: concurrently learning two cultures ("circle" and "triangle" cultures). (c) Crossing cultures bicultural: learning a host culture followed by learning a new culture. (d) Dynamic culture: an individual re-constructs a new culture.

A typical example is the bicultural individual, that is, individuals who have been substantially exposed to two cultures (either concurrently during early childhood or sequentially with secondary enculturation happening later in life). Second, dynamic cultural phenomena could emerge when political or other events accentuate or suppress or even create a new cultural identity.

We explain these phenomena, elaborate their importance for testing TTOM, and discuss tools for its empirical assessment.

The bicultural phenomenon

The bicultural individual "knows" that the very same behavior could have different causes, interpretations, and consequences depending on the culture (context) it is manifested. This critical ability to navigate different cultural frames is termed *cultural frame switching* (Briley et al. 2005; Hong et al. 2000) and has been adequately demonstrated – yet, the exact computational accounts are hypothesized at best.

Using free-energy principles (FEP) terminology, the bicultural mind not only has many generative models, but, also, the ability to effortlessly switch between them. Antithetical cultural identities might disrupt cultural switching (Benet-Martínez et al. 2002) and understanding the underlying processes could be a critical test for the FEP framework. Moreover, adopting different generative models entails different pragmatic affordances and, eventually, different decision computations. Crucially, even short-lived cultural cues, such as the linguistic environment or visual stimuli, could alter attitudes, behavior, and choices of biculturals (Cheng et al. 2006). This mechanism can be harnessed to experimentally "induce" cultural frames (Hong et al. 2000).

Crossing cultures

This is a type of bicultural individuals whose native culture has been well established, with enculturation processes associated with the secondary ("host") culture ensuing. Typical examples include immigrants, "expatriates," students at a foreign country, and even sojourners. Thus, the exposure and immersion to a new culture might depend on the intergroup relationships between the immigrant groups and the host society, which will in turn evoke learning, unlearning, and relearning of the new and native culture (Christopoulos & Tobler 2016). Bicultural identity processes (Benet-Martínez et al. 2002) mentioned are also applicable. Understanding these individuals within the

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

TTOM framework is particularly interesting, as the whole process of enculturation is observable and reportable and TTOM predictions can be tested – including neurobiological responses.

For the TTOM framework, the individual-crossing cultures needs to exercise active inference to learn and understand the new environment. Given that uncertainty substantially increases in a new culture, attentional salience would be intensified. Thus, understanding attentional selection processes and the role of the native culture priors are critical questions: An individual who substantially relies on the priors and expectations of her native culture might face substantial enculturation difficulties; on the other hand, rebuilding a new generative model could be an unnecessarily slow process. It would be interesting if FEP could generate hypothetical models that demonstrate how secondary enculturation happens, fails, or succeed. Finally, the TTOM framework emphasizes that social learning is based on *leveraging trusted others* – yet, the thorny problem for individuals crossing cultures is to identify these trusted others. Thus, locating the deontic cues in a new culture is a central process of secondary enculturation.

Individuals as active cultural agents

There is a third phenomenon where political or other events accentuate cultural aspects and could even create relatively new cultural identities – even within the very same cultural space. For instance, the recent social movement in Hong Kong following the proposal of the extradition bill has amplified the differentiation between the Hong Kong and Mainland Chinese cultures (cf. Cheon & Hong 2019). In another historical example, the expulsion of Singapore from the Malaysia led to the emergence of an independent nation. Since then, Singapore developed her own cultural identity, building on cultural affordances from distinct cultural heritages (Malay, Chinese, Indian, and other mixed cultures – e.g., Peranakan).

Such phenomena are also of particular interest, as the cultural agent has to not only learn, unlearn, and relearn, but, essentially, to actively "construct" a new culture. This is a challenge for the TTOM which (probably) treats the individual as a passive agent; however, individuals assess, interpret, shape, and further develop culture as active culture agents.

Cognitive flexibility, model-free, and model-based learning

All examples described above would need to elicit neurocognitive functions that go beyond pure learning. The obvious candidate here is cognitive flexibility (CF) – the capacity to adapt to change and problem solve in new situations (Friedman et al. 2006). Neurocomputational models of CF overall, and of learning, unlearning, and relearning specifically, have been specified (Dayan & Daw 2008, among others) and the FEP computational approaches can accommodate them. Thus, we would welcome FEP computational models building on CF neurocomputational processes to interpret dynamic cultural phenomena. The models can be empirically tested, extending familiar neurocognitive assessments (probabilistic reversal learning, etc.).

Beyond intra- and cross-cultural studies, dynamic multicultural phenomena can offer significant insights into how culture is shaped. The TTOM approach proposed by Veissière et al. could explain the way cultural agents adapt to a new culture. Our proposal of examining the multicultural mind will not only be a strong epistemological test for the TTOM theory, but will help addressing complex and pragmatic cultural phenomena.

Digital life, a theory of minds, and mapping human and machine cultural universals

Kevin B. Clark[a–e]

[a]Research and Development Service, Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA 90073; [b]Felidae Conservation Fund, Mill Valley, CA 94941; [c]Campus Champions, Extreme Science and Engineering Discovery Environment (XSEDE), National Center for Supercomputing Applications, University of Illinois at Urbana–Champaign, Urbana, IL 61801; [d]Expert Network, Penn Center for Innovation, University of Pennsylvania, Philadelphia, PA 19104 and [e]Virus Focus Group, NASA Astrobiology Institute, NASA Ames Research Center, Moffett Field, CA 94035.
kbclarkphd@yahoo.com www.linkedin.com/pub/kevin-clark/58/67/19a

doi:10.1017/S0140525X19002838, e98

Abstract

Emerging cybertechnologies, such as social digibots, bend epistemological conventions of life and culture already complicated by human and animal relationships. Virtually-augmented niches of machines and organic life promise new free-energy-governed selection of intelligent digital life. These provocative eco-evolutionary contexts demand a theory of (natural and artificial) minds to characterize and validate the immersive social phenomena universally-shaping cultural affordances.

Veissière et al., fitting free-energy formalism to human cognition and culture, produce an overdue exciting eco-evolutionary description of culture and associated phenomena with significant scope and eloquence. The authors' ambition to mathematically unify phylogenetic, ontogenetic, and sociogenetic mechanisms necessary for culture acquisition, expression, adaptation, and transmission is both welcome and admirable. Nevertheless, the precision, power, and internal and external validity of their framework await further testing and elaboration by skilled empiricists and theorists. For more complete, if inconsistent (Gödel 1931), axiomatic accounts, proof and refinement of concepts may be arguably best achieved through the use of broader constructivist approaches that entertain social systems of varying complexity and epistemological likeness to human culture. Veissière et al., for instance, disappointingly avoid considering major evidence-backed cultural features of animal societies and their modifiable relationships to human culture (cf. Clark 2012; 2019; Gowdy & Krall 2016; Laland & Galef 2009; Russon et al. 1996), including, but not limited to, pedagogy, tool production and use, mating and mourning rituals, grooming and personal adornment, division of labor, group defense and hunting, parenting, folk medical practice, superstition, communication dialects and mannerisms, abode construction, and sociopolitical hierarchies and maneuvering. Although debated among experts (Laland & Galef 2009), many phylogenetically earlier features of so-called animal culture seem to rely on the same mechanisms commonly attributed to human cultural expression. Indeed, much can be scientifically learned from these antecedent affordances of human culture, perhaps especially characteristics delimiting the origins and evolution of social learning and inferences which, in turn, shape the collective factors embodying and driving social context and

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

culture, a process the authors' term "Thinking through Other Minds."

Despite valuable insights gained from contemplating comparative aspects of animal culture, another overlooked or forgotten nascent culture – the culture of intelligent machines – also demands scrutiny and might better inform Veissière et al.'s variational approach to cognition and culture. Intelligent machines notably deliver prospective high-performance computational platforms to model and emulate all known organic forms of Earth life (e.g., Chew et al. 2014; Derex & Boyd 2015; Fan & Markram 2019; Fortuna et al. 2013; Fung 2015; Gillings et al. 2016; Lenski et al. 1999; Libin & Libin 2005; Millar et al. 2019; Ranjan et al. 2019; Sarma et al. 2018). They, moreover, render virtually-augmented eco-evolutionary spaces to prototype, replicate, and adapt digital life or e-life (i.e., electronic life), new provocative programmable *in silico* forms of life thought capable of taxa-blurring sophisticated intelligent agency and culture often celebrated and feared by fantasists and futurists (cf. Arbib & Fellous 2004; Asada 2015; Bostrom 2014; Briegel 2012; Cardon 2006; Clark 2015; 2017; in press a; Davies 2016; Fung 2015; Han et al. 2013; Kaipa et al. 2010; Lake et al. 2018; Lumaca & Baggio 2017; Mathews et al. 2017; McShea 2013; Parisi 1997; Thomaz & Cakmak 2013; Wallach et al. 2010). Trends in state-of-the-art digital life research and development emphasize dynamic machine-learning construction, operation, and evolution of whole semi-autonomous artificial organisms, such as OpenWorm's simulated nematode *Caenorhabditis elegans* and the plant-science community's crop-on-a-chip *Arabidopsis thaliana*. In addition, persistent improvements in machine architecture and software engineering continue to innovate progressively human-like artificial brains and intelligences, such as the Blue Brain Project's intricate virtual brain connectome, IBM's *Jeopardy*-quiz-show-winning Watson supercomputer, and Amazon's Alexa and Apple's Siri cloud-based digital assistants or digibots.

Without fully autonomous machine intelligences, current state-of-the-art machine form and function, including machine–machine and human–machine interactions, still fail to scale to computational and combinatorial complexities of purposeful sentient humans and their interactive societies. But, the near-seamless, data-exchanging distributed connectivity of smart machines to one another and to humans instantiates a kind of hybrid society built upon quasi model-free/model-based social networks. Networks comprised of social digibots and humans, for example, support local expectation, selective patterning of attention and behavior, and cultural evolution and inheritance. Changes in local network behavior reciprocally affect collective habits, norms, and expectations that typify larger virtually-augmented eco-environmental spaces shared by machines and humans. Critical analysis of this not-so-hypothetical cultural niche of organic and digital life first requires detailing a non-trivial theory of (natural and artificial) minds, through which Veissière et al.'s "Thinking through Other Minds" can be rigorously validated within and across taxonomic and technological margins. To that end, roboticists now carefully employ principles derived from relevant domain-knowledge sources, such as philosophy, computer science, zoology, engineering, neuroscience, anthropology, education, and sociobiology, to identify, craft, and implement machine traits fundamental to emergence and normalization of social–emotional intelligences (cf. Clark 2015; 2017; Lake et al. 2018). Machine intelligences evolve from simple initial state conditions that iteratively generate fuzzy clusters of core complex traits, such as causative inference, personality, self-awareness, empathy, and creativity. These traits help perfect and humanize learner machines via human and machine social modeling and queried instruction, where passive- and active-learning interfaces between learners and demonstrators/teachers correspond to credible primitive digital life manifestations of Veissière et al.'s "Thinking through Other Minds." Affordances of combined machine and human cultures then map to patterned immersive experiences of response priming/contagion, social facilitation, incentive motivation, and local/stimulus enhancement.

Similar to the authors' variational approach, free-energy constraints on digital life and virtually-augmented eco-evolutionary spaces dictate statistical regularities leveraged to forecast and organize behavior into machine and, in part, hybrid machine–human cultures (cf. Clark 2012; 2015; Langton 1990; Wolfram 1984). However, machine-culture thermodynamic profiles may radically deviate from profiles distinguishing human and even hybrid machine–human cultures. Causes of framework inconsistencies predictably stem from the frequent quasi-nature of machine architectures, algorithms, and social networks, which might inaccurately emulate human brains, neurocognitive processes, and social structures and practices. Furthermore, digital life may never bind or may become unbound from referential human-life conventions, so self-organizing machine performances reflect and harness purer semiotic and cybernetic optimization. Dissociations between digital and organic life thus greatly impact the types of statistics governing respective epistemic machine and human cultural resources, likely invalidating Veissière et al.'s preferred classical Bayesian models for variational frameworks more suited to individual and hybrid taxon/technology cultural niches defined by rarer, exotic, or otherwise different statistics (cf. Clark 2012; 2015).

Affective Social Learning serves as a quick and flexible complement to TTOM

Fabrice Clément[a] and Daniel Dukes[b,c]

[a]Cognitive Science Centre, University of Neuchâtel, Neuchâtel, Switzerland; [b]University of Fribourg, Fribourg, Switzerland and [c]Swiss Center for Affective Sciences, University of Geneva, Geneva, Switzerland.
fabrice.clement@unine.ch
https://www.unine.ch/islc/home/collaborateurs/professeurs/fabrice-clement.html
daniel.dukes@unige.ch https://dukes.space/

doi:10.1017/S0140525X19002784, e99

Abstract

Although we applaud the general aims of the target article, we argue that Affective Social Learning completes TTOM by pointing out how emotions can provide another route to acquiring culture, a route which may be quicker, more flexible, and even closer to an axiological definition of culture (less about what *is*, and more about what *should be*) than TTOM itself.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

The quest to identify what is unique to being human tends nowadays to highlight two dimensions: an individual's ability to represent the mental states of one's conspecifics, or, *theory of mind*, and a social organization that is highly dependent on shared practices and beliefs, or, *culture*. For a long time now, these two aspects of human life have been studied by disciplines that did not interact much: psychologists focussed on *internal* mentalistic processes, whereas sociologists and cultural anthropologists were interested in the *external* properties of cultural processes and institutions. Recently, however, more and more researchers from both sides acknowledge not only that psychological life does not occur in a social void, but also that the acquisition of culture necessitates complex social and psychological processes (Clément and Dukes, in review).

Although the birth and development of theory of mind has given rise to a great number of articles and lines of research, attempts to understand how individuals blend into their culture have been far less numerous. One of the difficulties is that much of what is expected from members of any given culture (body practices, food preferences, ways of speaking, etc.) is not explicitly transmitted. Individuals have, therefore, to figure out what is expected of them. The model exposed in the target article, Thinking through other minds, does an excellent job at explaining how this process is underpinned by "the 'lens' of expectations about another's expectations" and how the kind of expectations that "*Homo Sapiens* have leveraged most over their phylogenetic history involve the capacity to 'outsource' cognition to relevant others" because it is these "evaluations by others that make worlds 'meaningful' for humans" (sect. 3.6, para. 4). One of the strengths of this model is to show that a standard view of theory of mind is insufficient to explain how culture is transmitted. When new members are trying to figure out what kind of behaviours are expected from them, they are usually not trying to represent other mental states. Most often, in fact, they are observing their social surroundings, trying to imitate what must be done. The concept of cultural affordance, that we also defend (Kaufmann & Clément 2014), indicates how individuals can behave appropriately, even if they do not master explicitly the "rules of the game."

However, although we entirely agree with the idea that it is possible to follow the norms and rules of a given culture "without engaging with others' interiority" (sect. 2.3, para. 4), we believe that the scaffolding of attentional processes is not sufficient to explain how culture is transmitted. Of course, it is crucial to detect the right models to imitate, notably through social status or prestige, and even children are good at doing this (Charafeddine et al. 2015; Chudek et al. 2012). It is also important to organize collectively cultural niches that will guide attention to what is socially relevant to reproduce the "regime of attention" (sect. 3.5, para 5) that characterizes the mastery of any cultural forms. But, these aspects of cultural learning do not do justice to the intrinsic *normative* dimension of culture. For instance, a child can detect via others that a given object or behaviour is worthy of attention. However, this does not yet specify how this object has to be evaluated or *appraised*. It is, for example, possible to be attracted by a particular behaviour because it is highly despicable in a given culture (spitting on the ground). Or the same object (a trinket) can be judged as highly valuable in one community (collectors of said trinkets) but ludicrous in another (members of the "high society").

Fortunately, humans are endowed with a faculty that plays an essential role in such evaluative processes: *emotions*. In the context of cultural learning, it is the valence and the intensity of others' emotional expressions in particular that can be used to detect what is expected from each member. From a very early age, babies are sensitive to the emotional valences and intensities that help them evaluate their environment (Sorce et al. 1985). For us, emotions play such an important role in the process of socialization that we recently proposed the concept of Affective Social Learning to refer to the different processes enabling humans (and, to a lesser extent, non-human primates) to use others' emotional expressions to figure out the norms and values intrinsic to any social group (Clément & Dukes 2017; Dukes & Clément 2019). One of the advantages of Affective Social Learning is to show that the transmission of values can follow different paths, marked notably with variable intensities of intentionality. For instance, a first and basic evaluation of an object, person, or situation can rely on *emotional contagion* (e.g., parents becoming anxious in the presence of out-group members). Affective Social Learning can also involve active exploration by the new member, even when the model is not aware that her emotional expressions are being observed. This is what we call *affective observation* (e.g., the expression of disdain for less fortunate people). When the model is explicitly communicating an emotional reaction, we find ourselves in a more classic situation: *social referencing* (e.g., a proud smile from her mother will encourage a child to browse the shelves of the public library). Finally, the transmission of values can be very explicit, with guidance by a passionate teacher who is planning the different steps of an interesting learning process, a procedure that could be called *natural pedagogy*.

Taking emotions into account completes the TTOM approach by pointing out another route to acquiring culture, a route that may be *quicker* (in terms of the reduced frequency of behavioural corrections or encouragements), *closer* to an axiological definition of culture (less about what *is*, and more about what *matters* and what is *meaningful*) and indeed may prove more *flexible* to the constant changes in cultural values than the model portrayed in the target article that relies on stability (sect. 5.2, para 1).

Maladaptive social norms, cultural progress, and the free-energy principle

Matteo Colombo

Tilburg Center for Logic, Ethics, and Philosophy of Science (TiLPS), Tilburg University, 5000 LE Tilburg, The Netherlands.
m.colombo@uvt.nl https://mteocolphi.wordpress.com/

doi:10.1017/S0140525X19002723, e100

Abstract

Veissière and collaborators ground their account of culture and social norms in the free-energy principle, which postulates that the utility (or adaptive value) of an outcome is equivalent to its probability. This equivalence would mean that their account entails that complying with social norms has always adaptive value. But, this is false, because many social norms are obviously maladaptive.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Veissière et al.'s account of "the ability to perform inferences about the shared beliefs that underwrite social norms" (sect. 1.2, para. 1) blurs the distinction between descriptive expectations (i.e., beliefs that enough people in a certain situation behave in a certain way), normative expectations (i.e., beliefs that enough people in a certain situation expect others to behave in a certain way), and preferences for conformity to social norms. This distinction grounds many existing accounts of social norms (e.g., Bicchieri 2006; Binmore 1994; Boyd & Richerson 2001; Colombo 2014; Elster 1989; Gintis 2007; Ullmann-Margalit 1977). In particular, according to one influential account, preferences for norm compliance are conditional on having the right descriptive and normative expectations (Bicchieri 2006). Hence, in order to identify and change social norms, the key is to find out and intervene on relevant expectations and their associated preferences for conformity.

Because the active inference model underlying Veissière et al.'s account reduces preferences to probabilities (more precisely, it identifies the utility of an observation with its log probability), it apparently rejects the separation between the probability of an outcome and its utility (Colombo 2017). This move raises the question of how those probabilities should be interpreted – are they subjective degrees of belief (or credences), or are they objective frequencies of outcomes within a given reference class, or are they objective propensities of entities in the actual world (cf., Colombo & Wright 2018, Section 3.1)? Clarifying this question is important to facilitate the comparability between Veissière et al.'s model and other accounts of social norms, but also – and especially – to evaluate the prospects of using Veissière et al.'s account for real-world interventions on *maladaptive social norms*.

Veissière and collaborators explain that their account "focuses on the *conservative nature of human culture* – its ability to ensure that certain well-bounded and highly valuable states are frequented" (sect. 3.3, para. 2). This focus on conformity and conservation is potentially misleading, because it suggests that any social state brought about by social norm compliance is valuable, insofar norm compliance reduces one's sensory uncertainty. The problem with this suggestion is that there are many social norms that are maladaptive, harmful, morally abhorrent, or just lack any value or social function. Social norms such as female genital cutting, open defecation, binge drinking, and norms of revenge, are widespread in several present-day communities, but have no value. Although these norms may encapsulate statistical regularities in a community, and "act as a guide to what to expect from the future […] putting uncertainty under control" (Douglas 1986, p. 48), they are maladaptive. Significant public policy efforts are, in fact, being made to intervene and change them, by targeting people's expectations and conditioned preferences for following those norms (Bicchieri 2016).

If Veissière et al.'s account is committed to the ideas that "[t]he action with the most affordance… is the one associated to the least *expected free energy*" (sect. 3.5, para. 7), that expected free energy is equivalent to uncertainty, and that the adaptive value of an action is equivalent to the reduction of uncertainty brought about by that action, then their account cannot obviously explain the emergence and resilience of maladaptive social norms, which nonetheless reduce uncertainty in local cultural contexts. This would mean that Veissière et al.'s account of cultural practices is less unifying than what they claim. It could explain only adaptive social norms, and why they are resilient to cultural change.

Veissière et al. might respond by pointing out that the time scale at which uncertainty is reduced is the key factor for explaining the resilience of maladaptive social norms too. Considering longer time scales, it can make good sense for a community to sample social norms that have low utility or adaptive value – and, hence, high uncertainty. The local, short-term increase in uncertainty produced by sampling and complying with maladaptive norms would serve "the more general process of reducing free energy (either for the individual, because it prepares the organism for potential changes in adaptive contexts, and enlarges the repertoire of responses for the individual or the group)" (sect. 3.3, para. 6). Compliance with maladaptive norms would thus be in the service of guiding the learning of social norms that are adaptive, and that constitute local or global minima in the larger free-energy landscape of sensory samples of different communities.

One challenge, however, is that this response seems to assume that cultural dynamics must promote cultural progress at a longer time scale. That is, over the course of social and cultural history, adaptive social norms would replace maladaptive ones, as the human condition will continue to improve, and more and more (risk-prone) moral trendsetters deviate from accepted, socially harmful norms. The rise in income and wealth that the world has experienced in the past couple of centuries, the increase in average life expectancy worldwide, and the increasing resistance to sexist and racist norms in many cultural communities in the world could be cited as examples of cultural dynamics promoting progress – although one may also offer several counterexamples, such as present global threats like human-induced climate change, which is sustained by an array of environmentally-harmful social norms.

One way to avoid this challenge is to deny that Veissière et al.'s account should be committed to the idea that cultural dynamics must promote cultural progress, and particularly that social states with high probability tend to have high adaptive value. Denying this commitment would mean, however, that at least in the social domain minimizing free energy with respect to actions may not be equivalent to maximizing expected utility.

The upshot can be formulated as a dilemma. Either Veissière et al.'s account should be grounded in the free-energy principle, or it need not. If it should, then its explanatory scope would be limited to the acquisition, production, and stabilization of only adaptive social norms. If their account avoids the commitment to the free-energy principle, its explanatory scope could obviously include social norms that are maladaptive, harmful, or morally abhorrent, whereas it would still illuminate the important insight that social norms can helpfully be conceived of as uncertainty minimizing devices.

Explaining or redefining mindreading?

Krzysztof Dołęga[a], Tobias Schlicht[a] and Daniel C. Dennett[b]

[a]Institut für Philosophie II, Ruhr-Universität Bochum, D-44780 Bochum, Germany and [b]Center for Cognitive Studies, Tufts University, Medford, MA 02155.

krzysztof.dolega@rub.de krysdolega.xyz
tobias.schlicht@rub.de www.rub.de/philosophy/bewusstsein/schlicht.html.en
Daniel.Dennett@tufts.edu https://ase.tufts.edu/cogstud/dennett/

doi:10.1017/S0140525X19002772, e101

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Abstract

Veissière et al. disrupt current debates over the nature of mindreading by bringing multiple positions under the umbrella of free-energy. However, it is not clear whether integrating the opposing sides under a common formal framework will yield new insights into how mindreading is achieved, rather than offering a mere redescription of the target phenomenon.

In their target article, Veissière et al. set out to close the divide between different accounts of mindreading by proposing a model framed in terms of the recently popular free-energy framework (Friston 2009; 2012). As they explain, they aim for "a compromise position between *internalist*, brain-based approaches (e.g., simulation and theory–theory theories), which emphasize the neural machinery in individual humans' brains that is necessary to read other minds, and *externalist* approaches (e.g., radical enactive and cultural evolutionary theory)" (sect. 1.3.3, para. 7). This way, the authors seem to follow recently popular "pluralist" approaches which allow for more than one strategy for understanding other minds (Fiebich & Coltheart 2015; Newen 2015; Zahavi 2014), whereas also being strongly committed to the unificatory force of the free-energy formulation. Although we applaud the article's core proposal of establishing a common formal foundation for bringing opposing accounts into a useful dialog, we think that Veissière et al. fail to appreciate important differences in the scope and explanatory aims of these accounts. We want to clarify these differences and point out that the competing positions are cast on different levels of analysis (Marr 1982). This means that although the competing models of mindreading do fit with Veissière et al.'s formalization of the target phenomenon, their position counts only as a first step toward a formal analysis of the explanandum and does not allow for disambiguating between different proposals regarding how it comes about.

There are two main sources of disagreement in the literature on mindreading. The first point of contention, as Veissière et al. correctly point out, is the way in which the target phenomenon should be defined. Supporters of theoretic and simulationist accounts of mindreading construe the explanandum as an internal, computational process involving manipulation of representations. Defenders of externalist accounts, on the other hand, claim that the phenomenon in question is something that happens between people and not just inside their skulls. What is crucial here is that this debate can be understood as taking place on what David Marr called the computational level of analysis, one concerned with defining what the problem solved by the cognitive system is. This can be brought to light using the example of Gallagher's (2008) enactive view which offers an entirely phenomenological model, which purposefully sets complicated issues of neural processing aside. Gallagher disagrees with proponents of the internalist accounts by challenging the idea that mindreading should be characterized in terms of "prediction and explanation" of others' behavior. Instead, he takes the target phenomenon to be more akin to "something like evaluative understanding" (Gallagher 2001, p. 94). Authors such as Gopnik and Wellman (2012) or Carruthers (2015) can happily acknowledge that mindreading often feels like this kind of understanding of others, but it is not the issue that they want to address (see below). It is in the context of this debate about the nature of the explanandum that the authors' proposal seems most promising. As they point out, the free-energy framework offers a formal toolkit that does not allow for a "strict distinction between *dynamics* (as emphasized by externalists) and *inference* (the focus of internalist models)" (sect. 1.3.3, para. 7). Thus, it can not only provide a common platform for formulating and comparing different models of mindreading, but also may promote forming new models integrating insights from both sides of the debate.

However, the target paper does not go beyond redefining the explanandum in terms of free-energy, as it does not touch on the second important issue in the mindreading literature – explaining how we should adequately capture the neural processing which underlies the capacity in question. This debate, waged predominantly between proponents of the two dominant internalist paradigms (though some anti-representationalists are also involved, see e.g., Hutto & Myin 2017), is concerned with what Marr called the algorithmic level of analysis. In other words, the issue at the core of this disagreement is not about what it is that the brain is doing, but how it is doing it – the nature of the representational vehicles and neural algorithms which make mindreading possible. Admittedly, the authors seem to acknowledge this much when they state that their proposal "would be difficult to test (due to its generality)" (sect. 5.1, para. 4), but they hope that it can help "derive specific integrative models" (sect. 5.1, para. 4). However, it seems to us that Veissière et al.'s account cannot offer serious insights into the algorithmic level as both theory- and simulation-theorists already employ probabilistic computational models compatible with the free-energy formulation (causal Bayesian graphs in the case of the former – Gopnik & Wellman, 2012; probabilistic forward-models in the case of the latter – Gallese, 2003, p. 521) to support their claims. Following Pickering and Clark (2014), we think that the only way to make progress in this debate is not to integrate different models under one computational description, but to identify specific constraints and empirical predictions these models place on physical mechanisms that could implement them.

"Social physiology" for psychiatric semiology: How TTOM can initiate an interactive turn for computational psychiatry?

Guillaume Dumas[a,b], Tudi Gozé[c,d] and Jean-Arthur Micoulaud-Franchi[e,f]

[a]Human Genetics and Cognitive Functions, Institut Pasteur, UMR3571 CNRS, Université de Paris, Paris, France; [b]Human Brain and Behavior Laboratory, Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, FL; [c]Department of Psychiatry, Psychotherapies, Art Therapy, Toulouse University Hospital, Toulouse, France; [d]Equipe de Recherche sur les Rationalités Philosophiques et les Savoirs – EA3051, Université de Toulouse, Jean Jaurès, France; [e]Services d'explorations fonctionnelles du système nerveux, Clinique du sommeil, CHU de Bordeaux, 33076 Bordeaux, France and [f]USR CNRS 3413 SANPSY, CHU Pellegrin, Université´ de Bordeaux, Bordeaux, France.

guillaume.dumas@pasteur.fr www.extrospection.eu
goze.t@chu-toulouse.fr

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

https://erraphis.univ-tlse2.fr/accueil-erraphis/navigation/equipe/goze-tudi-510220.kjsp
jean-arthur.micoulaud-franchi@chu-bordeaux.fr
https://sites.google.com/site/jarthurmicoulaud/home/

doi:10.1017/S0140525X19002735, e102

Abstract

Thinking through other minds (TTOM) encompasses new dimensions in computational psychiatry: social interaction and mutual sense-making. It questions the nature of psychiatric manifestations (semiology) in light of recent data on social interaction in neuroscience. We propose the concept of "social physiology" in response to the call by the conceivers of TTOM for the renewal of computational psychiatry.

Psychiatric semiology, that is, the science of clinical manifestations, considers that both symptoms and signs are "units of analysis." These units are actionable psychopathological features that are essential in practice for making the diagnosis and prognosis that underpin clinical decision-making (Micoulaud-Franchi et al. 2018). Psychiatric semiology is, therefore, very important. We comment on how the concept of thinking through other minds (TTOM) and the associated computational psychiatry model proposed by the authors not only questions the mechanisms underlying psychiatric manifestations (see Veissière et al. target article, sect. 5.2, para. 3), but also the ways in which these manifestations are expressed by patients and captured by psychiatrists. Indeed, we need a computational model that questions how clinical manifestations are expressed and captured. In the field of transcultural (Kirmayer & Crafa 2014) and phenomenological psychiatry (Nordgaard et al. 2013), it is widely considered that semiology is partially based on social and cultural construction (i.e., history of medicine, consensus of experts, folk psychology, etc.; Kirmayer & Ramstead 2017). Moreover, with a hand outstretched to medicine, psychiatry considers that symptoms and signs are in some way linked to physiologic disturbances in the brain, as investigated by neuroscience (Micoulaud-Franchi et al. 2016). Since Jaspers' work, clinical manifestations have been taken to reflect both physiologic disturbances and patients' attitudes toward them (Stanghellini et al. 2013). In this view, clinical manifestations are not just related to an underlying physiologic cause but also to cognitive-interpretive and interpersonal processes that are at play during the constitution of symptoms and signs (Kirmayer & Ramstead 2017; Kirmayer & Sartorius 2007). Although theories have been proposed to account for these two dimensions central to psychiatric semiology (Berrios 1996; Borsboom et al. 2018; Kirmayer & Ramstead 2017), none has been formalized with a computational model. Although TTOM is a welcome addition that could help in formalizing clinical manifestations expressed and captured at the cultural/social level, we think that it should also take recent advances in the neuroscience of social interaction into account.

In the last decade, social neuroscience has indeed become interactive in acknowledging the impact of interpersonal social dynamics on intra-personal neurobehavioral dynamics (Hari & Kujala 2009; Redcay & Schilbach 2019). The second-person perspective (Schilbach et al. 2013) has already led to the development of a "second-person neuropsychiatry" that considers psychiatric disorders as "disturbances of social cognition" (Schilbach 2016). Complementary to this perspective, two-body neuroscience and two-person physiology (Bolis & Schilbach 2018b; Dumas 2011) follow the call for radical embodiment in cognitive science (Thompson & Varela 2001) and emphasize the constitutive role of interpersonal dynamics in individual cognition. Hyperscanning, that is, the simultaneous brain recording of several people (Montague et al. 2002), has demonstrated how non-verbal interaction through sensorimotor loops modulates individuals respective internal neurophysiological dynamics and how interpersonal dynamics are measurable at the electrophysiological level through inter-brain synchronizations (Dumas et al. 2010). Although it does not negate the existence of higher-order representations (e.g., language, cultural habitus; Shea et al. 2014), it supports the development of a "social physiology" that continuously integrates sensorimotor and representational levels of analysis. Interestingly, TTOM is already trying to resolve the difference between the individual and the inter-individual, and the authors present embodied interaction as always being culturally coded by implicit cultural learning. However, in our opinion, there is also a more basic non-culturally coded and non-representational layer of interaction that is directly rooted in early developmental processes.

This layer of interaction is particularly relevant for studying neurodevelopmental disorders such as autism and schizophrenia. In autism, multi-scale approaches have already attempted to solve this social paradox (Bolis et al. 2017; Dumas et al. 2014). In schizophrenia, the phenomenological approach has been used to account for impairment in the ability to learn the implicit social senses and use them in non-verbal communication (Fuchs 2015). This involves a nonverbal and pre-individual layer of relationship (Lavelle et al. 2014) that is closely linked to minimal-self disorder as an alteration of the first-person perspective (Parnas & Zandersen 2018). This alteration raises the question how symptoms are expressed by patients and captured by psychiatrists.

This issue has been analyzed within the framework of the classical concept of "praecox feeling." This "feeling" of bizarreness in interaction can be considered as a crucial determinant of medical decision-making in psychiatry (Cermolacce et al. 2010; Gozé et al. 2018), because it is directly rooted in minimal-self disorders (Parnas 2011; Sass et al. 2018). These well-documented first-person accounts suggest that patients are affected at a more basic level than TTOM. Hence, this suggests the existence of a sub-layer of TTOM itself which could be impaired, so the mechanisms underlying the involvement of TTOM require their own model. To meet this objective, social physiology calls for a computational model under (non-representational) and beyond (implicit or explicit social representations) the individual. Although Bayesian statistics can virtually integrate these dimensions (Friston & Frith 2015a), they need to be captured by generative models that are based on other types of computational formalism (Friston et al. 2017; Montague et al. 2011), especially biophysically-based neural circuit models (Wang & Krystal 2014). In our view, although TTOM provides a good matrix to model psychiatric semiology and its relationship both to physiology and to social interaction, it also requires the development of complementary computational models to account for physiological brain mechanisms and non-representational interpersonal dynamics (Dumas et al. 2012). The goal is to encompass the neurophysiological level and not separate the "implementation" from the "computational" in Marr's sense (Marr 1982). Such TTOM combined with social physiology, including all three of Marr's levels (computational, algorithms, and implementation), could offer great perspectives for our understanding of how psychiatric manifestations are expressed and captured. This could help computational psychiatry to structure the classification of mental

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

disorders, even including the more tacit mechanisms of intuition in clinical decision-making.

Enculturation without TTOM and Bayesianism without FEP: Another Bayesian theory of culture is needed

Martin Fortier-Davy

Department of Cognitive Studies, Institut Jean Nicod, EHESS/ENS/PSL University, 75005 Paris, France.
martin.fortier@ens.fr https://sites.google.com/site/martineliefortier/

doi:10.1017/S0140525X19002905, e103

Abstract

First, I discuss cross-cultural evidence showing that a good deal of enculturation takes place outside of thinking through other minds. Second, I review evidence challenging the claim that humans seek to minimize entropy. Finally, I argue that optimality claims should be avoided, and that descriptive Bayesianism offers a more promising avenue for the development of a Bayesian theory of culture.

In recent years, Bayesian approaches to the mind/brain have become very influential. Two lines of research deserve to be highlighted: one explores how cognitive development can be accounted for by rational constructivist models (Gopnik & Wellman 2012; Griffiths et al. 2010; Tenenbaum et al. 2011; Xu 2007), whereas the other investigates how neural processing can be understood as a form of predictive coding (Clark 2013b; Shipp 2016), and more generally, how entropy reduction is made possible through active inference (Friston 2010). As yet, very few attempts have been made to apply these emerging theories to the study of culture. In this respect, Veissière et al.'s endeavor can only be applauded. Although I am sympathetic with the general spirit of the authors' endeavor, I will contend that their theory is not viable because of at least two important flaws.

Drawing upon two concepts – "thinking through other minds" (TTOM) and the "free-energy principle" (FEP) – Veissière et al. intend to explain the origins of implicit cultural norms, beliefs, and habits (sect. 1.1, para. 2). As a brief reminder, TTOM has it that "information *from and about other people's expectations* constitutes the primary domain of statistical regularities that humans leverage to predict and organize behavior"; moreover, FEP stipulates that "living systems act to limit the repertoire of physiological (interoceptive) and perceptual (exteroceptive) states in which they can find themselves." Can these two concepts help us understand the mechanisms of enculturation? I doubt it. Indeed, *contra* TTOM, it is known that some priors are shaped through strictly *intra*personal – and not *inter*personal – processes; and, *contra* FEP, there is ample evidence of human behavior not complying with entropy minimization.

Veissière et al. mention "optical illusions" as one of the *explananda* of their theory (sect. 1.1, para. 2). Therefore, let me first discuss this specific example. The best theory we have of visual illusions – the natural scene statistics theory – shows that visual priors responsible for illusions are shaped by regularities in the surrounding environment (Howe & Purves 2002; Howe et al. 2005). This theory accords well with cross-cultural work demonstrating that variation in the perception of visual illusions directly results from the exposure to every day's environment (Miyamoto et al. 2006; Segall et al. 1966). For instance, individuals growing up around complex and ambiguous scenes will be more likely to develop a "holistic perceptual style" and to be tricked by illusions requiring context-independent scrutiny. The process through which the enculturation of visual priors takes place does not involve one's expectations about other people's expectations; in other words, it is TTOM-free. Therefore, it is difficult to understand what the authors might want to mean when they claim that TTOM can shed light on the enculturation of implicit priors responsible for optical illusions.

Importantly, this objection is not restricted to the domain of vision; it applies to numerous other domains. For example, categorization and reasoning have been shown to vary across cultures because of "ecocultural factors" (e.g., being a farmer rather than a fisherman) (Uskul et al. 2008). Now, these factors are all about individual exposure to specific environmental patterns and have not much to do with TTOM. In sum, it seems that Veissière et al. have overlooked the wealth of evidence showing that a good deal of enculturation takes place completely outside of TTOM.

Another central claim of the article is that humans *tend to minimize entropy*. Interestingly, Veissière et al. point out that entropy reduction is consistent with *temporary* entropy increase (sect. 3.1 and 3.3). When humans happen to be seeking uncertainty, the authors note, it is only because they anticipate that a dramatic drop in entropy will take place soon after (the peekaboo game is mentioned to illustrate this point). Unfortunately, here again, apart from anecdotal evidence, no experimental data are offered by the authors to corroborate their claim. Crucially, against FEP, numerous studies have shown that in esthetics (Delplanque et al. 2019), music perception (Chmiel & Schubert 2017), visual perception (Chetverikov & Kristjánsson 2016), consumer behavior (Kao & Wang 2013), etc., humans have a preference for medium entropy patterns rather than low-entropy patterns. Entropy and liking follow an inverted U curve: expected (low entropy) patterns are judged to be boring, completely unexpected (high entropy) patterns are deemed too difficult/demanding, and medium entropy patterns are liked and looked for (cf. Berlyne 1966).

Things happen to be even more intricate than just suggested, for if, on the one hand, plenty of studies have shown a preference of humans for medium entropy, on the other hand, an increasing number of studies demonstrate that the preferred level of entropy is highly variable across individuals (e.g., Güçlütürk et al. 2016; Güçlütürk & van Lier 2019). This line of research suggests that the relationship between liking and entropy may be largely shaped by cultural factors, and as a consequence, that any normative claim – for example, "humans seek to minimize entropy" – is pointless.

Optimality claims are particularly knotty (Frank 2013); this is why, in response to critics, some Bayesians have recently proposed to distinguish between *normative* and *descriptive* Bayesian models, and have further argued that descriptive Bayesianism fares better against criticisms (Tauber et al. 2017). It is unfortunate that the authors do not address this important issue and fail to adumbrate an optimality-free version of their framework. Last but not least, Friston has elsewhere acknowledged that FEP is not empirically falsifiable (Friston et al. 2018, p. 21); therefore, it is not clear to me whether Veissière et al. intend to make an experimentally testable claim when they state that humans tend to minimize entropy. For the same reason, it is not clear either whether FEP can be of any avail to social scientists.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

In conclusion, I wish to emphasize that none of the above criticisms undermines the prospect of a Bayesian theory of culture. What I have argued, rather, is that if such a theory is to be achieved, it will build upon descriptive constructivist Bayesian models of cognition (e.g., Fortier & Kim 2017) rather than FEP.

Acknowledgment. I wish to thank Daniel A. Friedman for his feedback on a previous version of this commentary.

The role of communication in acquisition, curation, and transmission of culture

Hyowon Gweon

Stanford University, Stanford, CA 94305.
gweon@stanford.edu https://sll.stanford.edu

doi:10.1017/S0140525X19002863, e104

Abstract

Veissière et al.'s proposal aims to explain how cognition enables cultural learning, but fails to acknowledge a distinctively human behavior critical to this process: *communication*. Recent advances in developmental and computational cognitive science suggest that the social-cognitive capacities central to TTOM also support sophisticated yet remarkably early-emerging inferences and communicative behaviors that allow us to learn and share abstract knowledge.

Veissière et al.'s proposal tackles a big question: How does human cognition support acquisition and transmission of culture? They suggest that the key link between cognition and culture is social learning, which occurs when people infer others' expectations – about how one ought to interact with the physical environment and about how one ought to interact with others in social contexts – and use these inferences to guide their own behaviors. The scope of the phenomena they try to explain is ambitiously broad, and their model is correspondingly quite general. Yet, despite its generality, their proposal fails to acknowledge a distinctively human behavior that is critical to acquisition and transmission of culture: *communication*.

Their claim that most of cultural learning occurs "without explicit instruction" (sect. 1.1, para. 2) reflects the widespread (yet misleading) dichotomy between "implicit" versus "explicit" social learning; the former is often characterized as attentional biases and copying strategies that are rooted in evolution and shared across species, and the latter usually refers to deliberate instruction and pedagogy that emerge relatively late in ontogeny (Heyes 2018a). By adopting this view, Veissière et al. provide a discussion on social learning that is a glass only half-full, omitting a range of rich, inferential epistemic practices that exist "in between" the two ends of the spectrum.

Recent advances in computational and developmental cognitive science offer a more precise account of how mutual expectations and mental-state reasoning naturally give rise to contexts where two parties communicate to achieve a joint goal (Grice 1975; Tomasello 2010): One intends to learn, and the other intends to inform. Bayesian models of social learning (Shafto et al. 2014; Vélez & Gweon 2018) and communication (Goodman & Frank 2016) have formalized such cooperative exchanges of information as a set of mutually constraining inferences and expectations about other minds; the learner expects the teacher to consider the learner's goals and knowledge to provide the best set of evidence for the learner, and the teacher expects the learner to rationally update her beliefs given the evidence. These expectations naturally give rise to powerful inferences and communicative behaviors that are present even early in life and ubiquitous in our everyday social interactions.

Children as (selective) learners

Prior developmental research has offered initial empirical support for these formal models using children's exploration as an index of their inferences as learners. When an adult pedagogically demonstrates one causal function of a novel gadget (e.g., pressing a lever plays music), the model expects a knowledgeable, helpful "teacher" to provide an exhaustive demonstration of its functions; consistent with this expectation, children infer that the gadget has no other functions and modulate their explorations of the gadget accordingly (Bonawitz et al. 2011). These results were replicated in Yucatec Mayan culture where pedagogical instruction is rare (Shneidman et al. 2016), further supporting the idea that these inferences are rooted in basic social-cognitive capacities rather than culturally specific teaching practices. Learners' expectations about helpful teachers also allow children to identify unhelpful sources of information. Beyond using accuracy as a cue, preschool-aged children understand that the same accurate information can be under- or over-informative depending on the learner's prior knowledge, and evaluate others based on what they *expect* of a helpful, knowledgeable informant (Gweon & Asaba 2017; Gweon et al. 2014; 2018). Although selective social learning is often characterized as detecting cues that indicate when or whom to copy (e.g., a learner's own uncertainty, particular traits of conspecifics such as age or prestige; Kendal et al. 2018), these mutual expectations and mental-state inferences allow young learners to flexibly shield themselves from various forms of misinformation.

Children as teachers

When one learns primarily through copying and imitation, the decision to copy typically falls on the learner. However, when social learning occurs via communication, the real heavy lifting comes from teachers who can choose what knowledge should be passed onto the next generation. A recent study suggests that by 5–7 years of age, children make rational decisions about what is best to teach by considering what is rewarding to learn and what is more costly for learners to discover on their own (Bridgers et al., in press). Although teaching benefits learners by reducing the cost of exploration and trial-and-error, teachers must be selective because it's impossible to teach everything. By prioritizing high-utility knowledge, teachers ensure that learners acquire the most critical, valuable information without the high costs of learning. Over generations, this process curates a body of cultural knowledge that is considered worthy of preserving and teaching.

Toward a more complete picture of social learning

Ironically, the cognitive capacities that support these epistemic practices are also central to Veissière et al.'s proposal: Mutual expectations and mental-state reasoning supported by Bayesian inferences and representations of expected utility (Goodman &

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Frank 2016; Jara-Ettinger et al. 2016; Shafto et al. 2014). In contrast to their assumptions, communication-based social learning emerges early in ontogeny (Csibra & Gergely 2009), is widespread across cultures (Hewlett et al. 2011), and does not always involve explicit, verbal transmission of knowledge (Gweon et al. 2010). In fact, these assumptions reflect a broader issue in the field: A disproportionate emphasis on copying as the primary means of social learning (Boyd et al. 2011; Rendell et al. 2010).

Human cultural knowledge is far more than a random collection of information. Regardless of its source – exploration, copying, communication – the evidence we observe is incorporated into a system of abstract, structured knowledge (i.e., intuitive theories) that allow us to explain the past, predict the future, and plan our own actions. Critically, the past includes our mistakes or what we didn't know, the future includes the benefit of "knowing" to our offspring, and the actions include actively sharing our knowledge with those who will benefit from it. To understand how knowledge grows over one's lifetime and over generations, we must ask how human social learning goes *beyond copying*, and how smart social learners become smarter teachers who willingly take on the costs of "cognitive outsourcing."

The cost of over-intellectualizing the free-energy principle

Daniel D. Hutto

School of Liberal Arts, Faculty of Law, Humanities and the Arts, University of Wollongong, New South Wales 2522, Australia.
ddhutto@uow.edu.au https://scholars.uow.edu.au/display/daniel_d_hutto

doi:10.1017/S0140525X19002851, e105

Abstract

This commentary raises a question about the target article's proposed explanation of what goes on when we think through other minds. It highlights a tension between non-mindreading characterizations of everyday social cognition and the individualist, cognitivist assumptions that target article's explanatory proposal inherits from the predictive processing framework it favours.

The target article seeks to illuminate the processes which enable and drive the acquisition of culture. Looking to clarify the big picture the target article ambitiously offers an account of how we catch implicit cultural norms – our ingrained habits and shared ways of doing things – from one another.

At the top level, the article defends the important but not-so-novel cognitive niche story, according to which the patterned practices that shape our expectations of one another are shaped by individuals responding to what they find salient in the landscape of affordances available to them (Hutto & Kirchhoff 2015). Moreover, the salient features of those local environments are themselves shape over the course of repeated of interactions – and they continue to cognitively shape the organisms that interact with them. And so on, and on. Thus, as the authors note, a major implication of their thinking through other minds, TTOM, thesis is that more traditional, individualistic approaches to the origins and basis of human cognition and reasoning are on the wrong track – or at least such approaches are inadequate if they adopt an individualist starting point. All of these ideas are promising, and worth examining in more detail. The article valuably highlights new research and empirical findings that can be used to support these aspects of the TTOM account they seek to defend.

Yet, the article has a much bolder aspiration. It looks to knit together a number of existing proposals – about cultural affordances, patterned practices, niche construction – by casting explanatory light on the mechanism or process by which we think through other minds. Calling on the theoretical resources of predictive processing and Bayesian brain theories of cognition, the solution on offer in the paper is to look to the free-energy principle, FEP, as the explanatory unifier.

The bare bones of the FEP proposal are credible enough. Certainly, the article succeeds in showing how its account of "regimes of attention" that are interactively shaped by a history of interactions with others which loops into a common local environment is compatible with the FEP. Even so, despite these modelling advantages FEP might not provide any deep insight into the nature of the actual processes that account for how shared practices are shaped by shared environments and vice versa.

TTOM invites philosophical challenge in that it sets out its positive story by appeal to individualist, cognitivist assumptions the cast the human social cognitive predicament as being fundamentally epistemic in character. When augmented by appeal to FEP the TTOM explanation is shot through with references to the idea that human brains encode information and the idea that individual brains are fundamentally inference engines – that brains deal their seclusion problem by advancing and improving best guesses about the causal structure of the world beyond. The cognitivist trappings of TTOM are utterly transparent, for example, in the claim that "the process of TTOM consists in inferring the priors or expectations that guide the beliefs of another agent or group of agents" (sect. 3.6, para 5).

Like its theory of mind predecessors, TTOM seeks to explain the basis of our social cognition by appeal to cognitive machinery that makes contentful inferences. Such accounts look promising if we accept the dominant characterization of social cognition – the standard mindreading story – which holds that attributing contentful mental states is what explains our capacity to understand and successfully interact with others.

The Bayesian brain framework assumes that – at its core – cognition is always and everywhere about making inferences concerning the hidden causes of sensory phenomena. The brain's inferences are taken to be subpersonal – implicit and unconscious – and abductive in character, unlike inferential operations of the sort found in deductive proofs. Yet, for all that, the brain's inferences are still presumably contentful and aim to accurately represent the true causal structure of reality.

In adopting this much of the predictive processing framework, the TTOM proposal appears to buy into the assumption that we are, fundamentally, at a spectatorial distance from others. Under the sway of this familiar assumption, philosophers are wont to claim that individuals have no direct access to other people's minds; that mental states are the, out-of-sight, hidden causes that drive behaviour; that in trying to understand what drives another's behaviour we need to posit hypothetical entities in our efforts to accurately get at hidden causes, and so on and on.

The spectatorial assumption is bound up with the idea that primary point and pervasive purpose of everyday social cognition is to bridge an assumed epistemic gap that allegedly exists between

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

us and others. It assumes that what drives our engagements with others is the brain's efforts to accurately infer the mental states – whether these are taken to be beliefs or expectations – that move them; to get at the hidden causes of their behaviour.

The spectatorial assumption is often adopted by default and without question. Nevertheless, many philosophers have argued that close inspection to our practices of everyday social cognition cast doubt on the idea that we are always at a fundamental epistemic remove from others (Hutto 2004; 2008; McGeer 2007).

Everyday social cognition, on these analyses, is a matter of directly interacting and engaging with the attitudes and emotions of others and understanding their projects and commitments, trusting or not trusting the accounts others give of why they do what they do. If this is correct, neither we nor or brains are always and everywhere attempting to discover the underlying causes of another's behaviour. To accept is to recognize that we do not "interact with one another as scientist to object, as observer to observed" (McGeer 2007, p. 146). It remains to be seen whether TTOM can adjust or relinquish the cognitivist and individualist commitments of its proposed FEP explanation in a way that is compatible with such analyses of social cognition.

Skill-based engagement with a rich landscape of affordances as an alternative to thinking through other minds

Julian Kiverstein[a,b] and Erik Rietveld[a,b,c]

[a]Department of Psychiatry, Amsterdam University Medical Centre, Amsterdam, The Netherlands; [b]Amsterdam Brain and Cognition, Amsterdam, The Netherlands and [c]Department of Philosophy, University of Twente, Enschede, The Netherlands.
j.d.kiverstein@amsterdamumc.uva.nl d.w.rietveld@amc.uva.nl

doi:10.1017/S0140525X1900284X, e106

Abstract

Veissière and colleagues make a valiant attempt at reconciling an internalist account of implicit cultural learning with an externalist account that understands social behaviour in terms of its environment-involving dynamics. However, unfortunately the author's attempt to forge a middle way between internalism and externalism fails. We argue their failure stems from the overly individualistic understanding of the perception of cultural affordances they propose.

Much of human social behaviour is regulated by normative expectations that originate in social and cultural life, and that people go along with without giving the matter any thought. Veissière and colleagues call these normative expectations "doxa" as contrasted with "dogma" (Bourdieu 1977). Doxa derives from regular ways of doing things held in common in a social group, and taken for granted by its members. How could doxa be learned if it is not transmitted as dogma is, through explicit instruction? The model Veissière and colleagues propose has on the face of it a strongly individualist, and internalist flavour. It is only by thinking through other minds and by forming expectations about what others expect of the world that we learn what others in our community expect of us. However, the authors suggest their account is also able to do justice to the arguments of externalists. They convincingly show how social expectations could take the form of statistical patterns that owe their existence to processes of developmental niche construction. An example of this statistical structure from the work of our own group is that of a desire path such as a well-trodden path through a park (Bruineberg et al. 2018b). The path has been set up over time by the repeated actions of others. The habits that come to guide our own walking behaviour can be thought of as the result of following a path already laid down by others before us.

We think the author's valiant attempt to forge a middle way between internalism and externalism fails. The failure stems from the overly individualistic understanding of the perception of cultural affordances the authors operate with, or so we shall argue. The target article shows how individuals learn the doxa of their community by having their attention tuned to cultural affordances. Veissière and colleagues characterize affordances as relationships between the abilities of an individual agent and the physical properties of things in the world. The perception of affordances they claim depends on the individual harbouring beliefs or expectations. An affordance just turns out to be a good bet – a highly probable belief – about what the agent can do with the world. Thus, when Veissière and colleagues ask how do people acquire or learn cultural affordances (sect. 2.3) what they really seem to mean is how do people learn the beliefs or expectations that are necessary for gaining perceptual access to cultural affordances. However, once we think of the perception of affordances as in this way dependent on the learning of beliefs, we cannot see what there is left for the environment to do that could not simply be done by the agent's beliefs. In what sense can the author's account be said to be externalist? The developmental niche only gets to make a contribution to the actions of individuals when they think through the minds of others, and learn what others expect of the world. The resulting model of social behaviour seems to us far removed from a theory in which the socially and culturally structured environment directly guides behaviour through an individual's responsiveness to its affordances.

We suggest two correctives to the account Veissière and colleagues propose of how doxa is learned. First, we suggest the distinction Veissière and colleagues make between sensorimotor and conventional affordances is an artificial distinction, and one that will end up doing more harm than good in an account of doxa. The distinction between natural and conventional affordances misses the way in which the so-called *natural* affordances of the human environment grow out of practices that are always both social and material (Rietveld & Kiverstein 2014; Van Dijk & Rietveld 2017). Affordances do not belong to an environment conceived of in the terms of physics and geometry as the authors suggest (sect. 2.2, para. 2). The availability of affordances in an econiche is dependent on the history of past activity of the creatures that inhabit this niche, just like the path through the park we began by discussing. The affordances of the human niche owe their existence not to the physics and geometry of things, but to the regular ways of doing things in our many forms of life (Rietveld & Kiverstein 2014). The physical structure of an econiche is entangled with and inseparable from the normatively regulated activities of the individuals that live in a given niche.

Second, Veissière and colleagues miss a distinction we will argue is crucial for understanding social learning and how individuals

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

acquire the abilities and skills for coordinating with each other in everyday life. We distinguish the rich landscape of affordances that is available in our human ecological niche by virtue of the skills and abilities in sociomaterial practices, and "solicitations" or relevant *inviting* affordances. Individuals acquire abilities through a process of the education of attention by other members of the sociomaterial practices. Based on the abilities and sensitivities they develop through the education of the attention (Gibson 1979; Rietveld & Kiverstein 2014), affordances come to stand out as soliciting or inviting action in particular concrete situations of action.

We certainly do not mean to dispute the key idea behind the free-energy principle that a complex adaptive system must minimize free energy if it is to preserve its organization for a prolonged period of time in a dynamic environment. In contrast, we think living beings are sensitive to the rise and fall in free energy over time (Kiverstein et al. 2019). It is on the basis of this sensitivity that affordances stand out from the landscape as relevant. Free energy we understand as the disattunement in an agent–environment system which the individual acts to keep to a minimum. Nor do we dispute that other people can serve as epistemic resources that help us to act in ways that ensure that we keep expected free energy to a minimum. We suggest, however, that to do adequate justice to these ideas the role of the econiche in constraining and structuring an individual's behaviour must be recognized. We dispute that to know how to go on in the same way as others do in a practice, we must first think through the minds of others. We learn what to do in a practice through having our attention educated to the affordances of our niche.

NOTE. Julian Kiverstein and Erik Rietveld are supported by the European Research Council in the form of ERC Starting Grant 679190 (EU Horizon 2020) for the project AFFORDS-HIGHER, the Netherlands Organisation for Scientific Research (NWO) in the form of a VIDI-grant awarded to Erik Rietveld, and by a project grant from the Amsterdam Brain and Cognition research group at the University of Amsterdam.

Culture and the plasticity of perception

Michael Lifshitz and T. M. Luhrmann

Department of Anthropology, Stanford University, Stanford, CA 94305-2034.
lifshitz@stanford.edu luhrmann@stanford.edu

doi:10.1017/S0140525X19002887, e107

Abstract

Culture shapes our basic sensory experience of the world. This is particularly striking in the study of religion and psychosis, where we and others have shown that cultural context determines both the structure and content of hallucination-like events. The cultural shaping of hallucinations may provide a rich case-study for linking cultural learning with emerging prediction-based models of perception.

One of the welcome consequences of the thinking through other minds model is that it supports a particular definition of culture: that culture is about patterns that people infer from the behavior of other people, and which in turn motivate their own behavior. The American anthropologists Alfred Kroeber and Clyde Kluckhohn (1952), the great shaggy lions of their day, settled on that definition in the early 1950s after considering hundreds of others.

In the ensuing decades, anthropologists roundly rejected this definition. It seemed too mental. Geertz (1973) was in part responsible. He insisted that culture happened between people, not in their heads. Then came poststructuralism and postmodernism, and anything that spoke of implicit rules seemed to imply too much explicit human intention.

This target article by Veissière et al. reminds us that if we are to give an account of culture that is consonant with a model from neuroscience, we must return to persons who draw inferences about the social world around them and who then act based upon those inferences. At the same time, its model should reassure anthropologists because the embodied cognition account does not depend on explicit intention.

The further promise of this model lies in its potential to bridge levels of explanation that are usually isolated: from the cultural through the psychological to the neuronal. The free-energy framework may help to investigate how large-scale cultural models shape private experience and behavior. However, whereas Veissière et al. focus on the transmission of cultural norms and knowledge via regimes of attention, they do little to unpack how their model might explain the power of culture over perception *per se*. Indeed, cultural context shapes not only human behavior and attention, but also our most basic subjective experience of sensory perception.

Our own work studying hallucination-like experiences demonstrates the impact of culture on the senses. In the domain of religion, for example, we have found that Charismatic Christians who pray to God with the expectation that God will talk back sometimes report that they occasionally hear the voice of God responding in a way that feels audible (Luhrmann 2012). There are individual differences: not everyone hears the voice of God, and those with a capacity for imaginative and sensory absorption seem particularly prone to such sensory overrides (Lifshitz et al. 2019). Training in prayer also seems to play an important role (Luhrmann & Morgain 2012). Still, it is a striking observation that holding (and practicing) a cultural model which says that God can (and should) talk back often leads people to experience directly that God is speaking with a hearing quality.

At times, specific cultural events may lead groups of people to report hallucination-like events. In the days following the death of Menachem Schneerson – a Hasidic Rabbi given messianic status – many of his followers reported seeing brief glimpses of him partaking in the activities of daily life (Bilu 2013). Clearly, a strong cultural expectation was at play: that the messiah does not die as normal people do and so may linger visibly on the earthly plane. The enduring puzzle is to link this top-down cultural expectation with the brain's prediction of incoming visual information to explain how these believers came to report that they had, in fact, seen their beloved Rebbe with their own open eyes.

In the study of psychosis, cultural context impacts not only the distress and prognosis of the illness (as pointed out by Veissière et al.), but also the structure and content of the auditory-verbal hallucinations themselves (Larøi et al. 2014). In a recent series of phenomenological interviews, one of us (TML) observed that in Chennai, psychotic patients often experienced the voices of kin; in Accra, patients frequently identified their voices as God; in California, people more often described voices as violent, and emanating neither from God nor from people they knew (Luhrmann et al. 2015). This is a complex story. Biological affordance, genetic predisposition, life experience, and cultural

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

invitation all seem to interact to shape the experience of hallucination-like events (Luhrmann et al. 2019).

Hallucinations may provide a particularly pertinent domain for fleshing out the implications of the thinking through other minds model. Recently, research has begun to outline a mechanistic predictive coding account of hallucinations, which relates hallucination-proneness to an over-weighting of top-down priors in response to ambiguity in the lower levels of sensory prediction (Corlett et al. 2018). If scientific evidence continues to support the strong-priors theory of hallucinations, this may open an exciting opportunity to link this low-level sensory/neurobiological explanation to the higher-order model of cultural transmission proposed by Veissière and colleagues. We may then have the beginnings of a cross-level account, scaffolded by the free-energy framework, of how culture comes to shape the senses.

It is interesting to note in the context of thinking through other minds that hallucination-like events are often social in nature (Wilkinson & Bell 2016). We hear someone speak, we see those who have passed away, we feel the touch of spirits and angels. Perhaps, such invisible beings may count among the relevant "others" that humans "think through" in transmitting culture. What we take God to think and do might very well sway how we come to think and behave ourselves. In this way, gods and spirits become some of those other minds through which we think.

This then raises the deep puzzle of why certain hallucination-like events acquire authority whereas others do not. When the authority is granted, of course the event becomes powerful because it then carries with it the authority of God, perhaps the ultimate bearer of epistemic prestige. But, why did Joan of Arc's voices compel King Charles VII to give her an army? Or – to frame the question in the most controversial way – why did Jesus's experience of God's voice lead others to follow him as divine, whereas the experiences of so many other would-be prophets did not? That is a complex story of historical uncertainty, but also of charisma, madness and perhaps of things that pass all human understanding.

How does social cognition shape enculturation?

John Michael[a] and Leon de Bruin[b]

[a]Department of Cognitive Science, Central European University, Budapest 1051, Hungary and [b]Department of Philosophy, Radboud University Nijmegen, 6500 HD Nijmegen, The Netherlands.
michaelj@ceu.edu l.debruin@ftr.ru.nl

doi:10.1017/S0140525X19002814, e108

Abstract

Other people in our culture actively transform our behavioral dispositions and mental states by shaping them in various ways. In the following, we highlight three points which Veissière et al. may consider in leveraging their account to illuminate the dynamics by which this occurs, and in particular, to shed light on how social cognition supports, and is supported by, enculturation.

One of the central aims of Veissière et al. is to explain how agents learn to respond to the norms and resources of their local cultural niche. Crucially, such an explanation has to be sensitive to the fact that among the most important components of this cultural niche are *other people*. Insofar, as our social success is about cultural niche construction, it has to involve the construction or shaping of people. Other people do not only scaffold our capacity to acquire norms and conventions, but they actively transform our behavioral dispositions and mental states by shaping them to respect the prevailing norms and stereotypes (De Bruin & Strijbos 2020; Zawidzki 2013). Although this idea seems to be implicit in Veissière et al., the thinking through other minds (TTOM) hypothesis could shed more light on the dynamics by which perception, action, and niche construction led to the acquisition and production of cultural habits, and to the inference and learning about other minds (Fig. 4 in the target article). The following commentary is an invitation to Veissière et al. to expand on this. We offer three possible starting points.

First, Veissière et al. do not go into detail about the forms of enculturation that require us to make inferences about the mental states of other agents and those that do not. On the one hand, they note that some forms of enculturation do not involve mindreading but merely "following the 'tracks' laid down in local environments by others, or following the norms and rules presented through institutions, without engaging with others' interiority" (sect. 2.3, para. 4). On the other hand, they also note that other forms require "inferences to how others think" and that "the process of TTOM consists in inferring the priors or expectations that guide the beliefs of another agent or group of agents" (sect. 3.6, para 5). In other words, some forms of enculturation require only superficial action copying, whereas others require inferences about mental states ranging from intentions to beliefs and desires as well as emotions. Given the variety of forms of enculturation and of mental states, it would be important to specify what forms of enculturation require what forms of theory of mind. A deeper explanation of these dynamics would make it easier to grasp the ontogenetic and phylogenetic implications of TTOM.

Second, and relatedly, it would be valuable if Veissière et al. could explain what sorts of theory of mind capacities they believe enculturation *presupposes*, and what sorts of theory of mind capacities enculturation *makes possible*. In some instances (e.g., sect. 3.6, para. 7) they appear to indicate quite broadly that theory of mind capacities are necessary for enculturation, but this is unlikely to be the case for more sophisticated forms of theory of mind, such as explicit reasoning about reasons, as in Mercier and Sperber's (2017) framework (which they cite repeatedly). For Mercier and Sperber, reasoning about reasons depends on theory of mind capacities that involve the attribution of propositional mental states with rich semantic contents. This would seem to presuppose enculturation rather than the other way around. Perhaps more basic forms of mindreading, such as level-1 perspective taking, shared attention, and action tracking are necessary for enculturation. It would be helpful to spell out reasons for thinking what forms of mindreading enable enculturation, and what forms are enabled by enculturation.

The two starting points identified above both concern the question what kind of inferences agents need to make, and which theory of mind capacities they need, in order for enculturation to take off (i.e., the top half of Fig. 4 in the target article). A third starting point would be to elucidate how enculturation contributes to the shaping of agents (i.e., the bottom half of the figure). One possibility, recently put forward by Hohwy and Michael (2017; cf. also

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Michael 2015), is that infants and young children progressively refine their agent-models through interaction with others, and also increasingly conform to those models themselves. More concretely, infants and young children apply agent-models to others during the course of development, and use these agent-models to guide imitation and other forms of cultural learning. A result of this is that it becomes increasingly feasible for others in their culture to model them as well. In other words, agents' models of themselves (their self-models) and their actual selves are fitted together as a natural consequence of modeling and interacting with each other. From the perspective of the prediction error minimization framework, this appears as a form of active inference: infants and young children shape their selves progressively to match the agent models that they have been using to interpret others.

We hope that careful consideration of these three points will help in clarifying the relationship among culture, theory of mind, and predictive coding, and that it will stimulate progress in explaining the dynamics by which minds shape, and are shaped by, other minds through the complex process of enculturation.

Enculturated minds, not error reduction minds

Robert Mirski[a], Mark H. Bickhard[b], David Eck[c] and Arkadiusz Gut[d]

[a]The John Paul II Catholic University of Lublin, 20-950 Lublin, Poland; [b]Lehigh University, Bethlehem, PA 18015; [c]Cañada College, Redwood City, CA 94061 and [d]The Nicolaus Copernicus University in Toruń, 87-100 Toruń, Poland.
robertmirski@kul.lublin.pl https://www.researchgate.net/profile/Robert_Mirski
mhb0@lehigh.edu https://www.lehigh.edu/~mhb0/mhb0.html
eckd@smccd.edu https://canadacollege.academia.edu/DavidEck
arekgut2001@gmail.com https://umk.academia.edu/ArkadiuszGut

doi:10.1017/S0140525X19002826, e109

Abstract

There are serious theoretical problems with the free-energy principle model, which are shown in the current article. We discuss the proposed model's inability to account for culturally emergent normativities, and point out the foundational issues that we claim this inability stems from.

We believe the free-energy principle (FEP) lacks theoretical resources to account for the complex phenomenon of culture. The current article's attempt at doing so results in a trivialization of the problem, and a reductionist view on what culture and its participants are. Below we focus on the problems the proposal faces with accounting for the diverse normativities that characterize encultured persons. After that, we argue that this is a symptom of more fundamental theoretical problems with the FEP.

The FEP claims that the overarching goal of every individual is to reduce free energy or uncertainty. Accordingly, all normativities that the system instantiates are claimed to come from the preselected set of "expectations"; for instance, living organisms are argued to move away from dangerous temperatures because these temperatures generate inputs incompatible with "expectations" about them (this is the example given in the current article). These adaptive "expectations" are argued to reside in the highest level "expectations," sometimes called hyperpriors (Clark 2013a), which have been formed during phylogeny; only those individuals with adaptive hyperprior "expectations" managed to survive and procreate (Friston et al. 2012; Kiebel et al. 2008).

Although a rather ingenious idea, the above claim runs into clear problems in the context of enculturation. People certainly have phylogenetically old normativities such as the ones satisfying our basic survival needs, but they also house a whole plethora of normativities emergent over the course of development, ones that cannot be argued to have formed in phylogeny. It hardly needs demonstration that genetically identical and raised in the same socio-cultural milieu twins can develop radically opposing sets of values and goals. What is more, these goals and values can sometimes override the phylogenetically old, adaptive normativities: history knows many cases of people deciding to die or suffer for some highly abstract cause. This fact seems entirely incompatible with the FEP model, and it is especially problematic in the context of the current proposal because these powerful, novel normativities usually emerge as part of the process of enculturation. In fact, encultured persons *are constituted* by such emergent normative phenomena: We certainly can identify more with our values and goals than with our biologically given motivation to stay alive, which itself is far from defined innately as it emerges ontogenetically in a social context too (e.g., we learn the "proper" ways of eating or sleeping from our cultures) (see Eck & Levine 2017).

In the context of the multi-layered human cognitive system, the highest-level, adaptive normativities given in hyperpriors are argued to yield information-seeking or global-uncertainty-reduction dynamics. This is held up in the current article as solving the "dark-room problem": increases in local uncertainty are expected to decrease global uncertainty over time, that is, to keep the organism within the innately expected states specified in the hyperpriors. This claim seems to give us another kind of normativity that is derived from the overarching motivation of the FEP: namely, the epistemic-gain motivation. Unfortunately, this does little to help the situation as motivations emergent in encultured persons cannot be reduced to information seeking either. How does my re-watching for a hundredth time an old cult movie at my house benefit me epistemically? In fact, culturally emergent normativities are sometimes flatly hostile to epistemic gain – ignorance passes for cool in some communities.

These issues of the FEP being incompatible with the reality of culturally emergent normativities bear heavy on the proposed model. Although the paper talks about relevant phenomena – such as norms, affect, or prestige – as if they have been explained (there are many such glaring cases of petitio principii in the article), the proposed model hinges solely on the epistemic-gain motivation. Culture boils down to informational redundancies created for a more efficient epistemic gain (cultural niche construction) and individuals learning about these redundancies (learning cultural affordances). There are no persons with their ontologically novel normativities, such as values, ideals, and other diverse motivations – just individual organisms helping each other minimize their free energy (for an analysis of the problem of such "manipulationist" views on culture in other models, see Eck 2015).

It becomes clear that the predictive processing framework and the FEP are not fit for modeling culture. Following Litwin and Miłkowski (submitted), we believe that the framework needs serious theoretical development before it can be fruitfully applied to specific problems such as the one at hand. Indeed, the inability to model normative emergence in enculturation is an important

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

special case of a more general problem: Free energy cannot handle *any* normative phenomena per se – neither the ones involved in enculturation nor the ones involved in development in general, not even the basic normativities inherent in life (Bickhard 2015; 2016; Martyushev 2018; Roesch et al. 2012). At best, pre-programed hyperpriors can extensionally capture predetermined behavior patterns. Any exceptions (e.g., seeking dark, instead of turning on the light; seeking pain [e.g., hot peppers] instead of avoiding it; etc.) must also be pre-programed: there is no modifying the hyperprior probabilities (they are innate, as are *all* of the *spaces* over which *all* of the probabilities are distributed) – there is no normative learning, no development, no socialization, and no enculturation, the last of which we discussed in this commentary. For a related discussion of problems with such foundationalism in cross-cultural research, see Mirski and Gut (2018). For an anticipatory framework that does address issues of normativity, see, for example, Bickhard (2009) and Campbell (2015) – including in the context of culture and language (e.g., Bickhard 1992; 2007; 2008).

Acknowledgments. RM was supported by a grant from the National Science Centre (UMO-2016/23/N/HS1/02887).

Importance of the "thinking through other minds" process explored through motor correlates of motivated social interactions

Harold Mouras

Laboratoire de Neurosciences Fonctionnelles et Pathologies (LNFP) EA 4559, Université de Picardie Jules Verne, Amiens, France.
harold.mouras@u-picardie.fr
https://www.u-picardie.fr/m-harold-mouras-202138.kjsp

doi:10.1017/S0140525X19002656, e110

Abstract

We wanted to gather recent results supporting the idea of the central role of sharing agency in socioaffective and motivational information processing. Here, we want to support the idea that this process is quite arbitrary, early in the temporal chain of processes and not only influence the psychological, but also the motor correlates of socioaffective information processes.

In their target article, Veissière et al. provide new theoretical arguments supporting the idea that "the human sense of obligation is intimately connected with the formation of a shared agent 'we', directing collaborative efforts and self-regulating them." Thus, they argue that "the human sense of obligation may thus be seen as a kind of self-conscious motivation."

Recently, several studies have brought experimental arguments supporting this idea by showing that the cognitive processes involved in this "formation of a shared agent" directly influence the psychological and motor correlates of socioaffective processes. Mainly, these studies have used a well-known theoretical and experimental model, that is, empathy for pain, which compares the processes (ratings, motor correlates) when viewing painful situations as compared to non-painful situations. Classically, the difference between the two conditions can be used as an index of empathy felt toward the character involved in the depicted situation. Therefore, several studies have been able to manipulate the nature of the social link between the observer (i.e., the participant) and the depicted character, in order to explore its influence on empathy.

Within social psychology, it is well-known that people have the propensity to divide the social world into *us* versus *them* influencing affective, cognitive, and behavioral processes. Interestingly, a powerful old paradigm (the minimal group paradigm; Tajfel et al. 1971) had demonstrated that the mere categorization of individuals into two social groups on the basis of arbitrary criteria (e.g., to over- or under-estimate the number of dots on a screen; Diehl 1990) was sufficient to produce similar consequences as compared to natural groups. We used for the first time this paradigm within the framework of empathy for pain (Montalan et al. 2012). Briefly, participants were shown pictures of people in painful or non-painful situations and were instructed to imagine themselves or imagine members of two minimal groups (in-group vs. out-group) in the same situation and participants had to rate the level of perceived pain according to the different perspectives. The results were quite clear: More than replicating previous results showing that the mere assignment of individuals to arbitrary groups elicits evaluative preferences for in-group relative to out-group members (Brewer 1979), we found that the mere act of categorizing people in two distinct social groups was also sufficient to elicit an in-group bias in empathy for pain. This was the first clear demonstration that the processes involved in the formation of a shared agent mentioned as central in Veissière et al.'s target article influenced the psychological processes of empathy.

What about the motor processes involved in socioaffective responses? We have been able to address this question by measuring the postural correlates of empathy for pain. The interrelation between the motor and affective components of behavior has been studied for a long time. For some theoretical models, emotion shapes behavior so that pleasant events should trigger approach whereas unpleasant events should trigger withdrawal. The ability to simulate another person's emotional response in a particular situation could be the basis for the development of empathic skills (Meltzoff & Decety 2003) and the instruction to adopt another person's perspective modulates pain rating according to the affective link between the observer and the individual experiencing the outcome (Singer et al. 2006). In a first study (Lelard et al. 2013), we used posturography to record differential postural responses when participants were instructed to imagine themselves in a painful or non-painful situation within the functional context of empathy for pain. This study demonstrated for the first time a stiffening response to pain visual stimulation, showing that postural responses were dependent of the perceived pain during the induced simulation process. These results laid the basis for further studies the basis for further studies concerning the role of perspective-taking in motivational dimension of motor control and social interaction. However, a main limitation of this study was that the effects of mental simulation were not tested, being unable to determine whether the reported effects were because of embodiment of the situation or to the valence of the visual scene.

A second study (Lelard et al. 2017) was designed to record the differential postural correlates of empathy for pain according to whether or not participants were instructed to imagine themselves in a painful or non-painful situation. Both painful visual scenes

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

(as in the preceding study) and instructions to embody the displayed situation were hypothesized to induce postural postural and physiological changes. The results demonstrated a posterior displacement of the body in the mental simulation condition compared to the passive observation condition, supporting the hypothesis that instruction to imagine ourselves in a painful situation activates internal models that lead to an embodiment of the situation (Zahavi 2008). This was the first study to describe adjustments of postural control in response to mental simulation of affective/motor pictures.

By summarizing these results in this commentary, we wanted to support the hypothesis of the target paper of the main importance of sharing agency in socioaffective and motivational processes. To us, these studies show that this process (i.e., the social categorization of the other as an in-group or out-group member) is quite arbitrary (as demonstrated by the easy experimental manipulation of social link) and influence not only the central processes involved in empathy (broadly socioaffective processes), but also the motor correlates of these responses.

The future of TTOM

Søren Overgaard

Center for Subjectivity Research, University of Copenhagen, DK-2300 Copenhagen S, Denmark.
s.overgaard@hum.ku.dk https://cfs.ku.dk/staff/?pure=en/persons/259148

doi:10.1017/S0140525X19002668, e111

Abstract

"Thinking through other minds," or TTOM, is defined in two different ways. On the one hand, it refers to something people *do* – for example, inferences they make about others' expectations. On the other hand, it refers to a particular theoretical *model* of those things that people do. If the concept of TTOM is to have any future, this ambiguity must be redressed.

What is the future of theory of mind, or TOM? That depends on what you mean by "TOM." If by "TOM" you mean something people *do* – say, attribute mental states to others – the future looks bright. There is little chance people are going to stop doing those things anytime soon. But, if you mean a particular theoretical *model* of those things people do – say that they use something akin to a *theory* – then it is less clear what the answer is. Simulation theorists and others have challenged the (theory-) theory model for decades. Because the *concept* of TOM is ambiguous in this way (something people do, or a particular theoretical model thereof), however, you might think that concept is best avoided in future research on social cognition (see, e.g., Apperly 2011, p. 3; Goldman 2006, p. 10; Nichols & Stich 2003, p. 2).

When Veissière et al. raise the question of what the future of "thinking through other minds," or TTOM, is, there is a similar ambiguity. Again, it all depends on what they mean by "TTOM." Running through their paper is a fundamental conceptual muddle, as "TTOM" is defined – and the notion is employed – in two entirely different ways.

According to one definition, it is "a model of implicit cultural learning that we call 'thinking through other minds' (TTOM)" (sect. 1.3, para. 2). TTOM in this sense integrates other approaches, argues for compromises between externalist and internalist accounts, does (or does not) have certain ontological commitments, supports this or that theoretical view, provides mathematical formalizations, and so on. And of course, it is TTOM in this sense that is "a generic active inference (also known as FEP or variational) account of the acquisition of culture and mind-reading abilities" (sect. 5.1, para. 4).

According to another definition, however, TTOM refers to something people *do* – in particular, the forming of expectations about other people and their expectations. Thus, for example, "We call this intersubjective process of engaging others' expectations and inferences "*thinking through other minds*" (sect. 2.4, para. 8). TTOM in this sense consists of inferring others' expectations, or learning to infer such things; individuals or groups of people may vary in terms of their capacities to "leverage" TTOM in this sense, and so on.

TTOM in this latter sense has a bright future, of course. Humans form habits and attitudes according to how they expect certain others to think and act. And they construct physical environments in such a way as to encourage certain behaviours (and discourage others), or make certain behaviourally relevant information salient, etc. A well-trodden path in the woods is a simple example; traffic lights, designed and positioned in such a way as to attract road users' attention, a more complex one. None of this is likely to change, and Veissière et al. nicely bring out how important TTOM in *this* sense might be to solving the mystery of implicit cultural learning.

TTOM understood as a theoretical model might also fare well, although naturally the situation here is less clear. That model's proposed compromise between "internalist" and "externalist" approaches to social cognition is likely to meet resistance from hardliners on both sides. And even those who are sympathetic to the idea of such a compromise may quibble about the details.

The *concept* of TTOM, however, faces very uncertain prospects. A theoretical concept as fundamentally ambiguous as this is only fit to breed confusion. Thus, if Veissière et al. want to hold on to the concept of TTOM, it is crucial that they clear up the conceptual muddle sooner rather than later.

Choosing a Markov blanket

Thomas Parr

Wellcome Centre for Human Neuroimaging, Queen Square Institute of Neurology, University College London, London WC1N 3AR, UK.
thomas.parr.12@ucl.ac.uk https://tejparr.github.io/

doi:10.1017/S0140525X19002632, e112

Abstract

This commentary focuses upon the relationship between two themes in the target article: the ways in which a Markov blanket may be defined and the role of precision and salience in mediating the interactions between what is internal and external to a system. These each rest upon the different perspectives we might take while "choosing" a Markov blanket.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

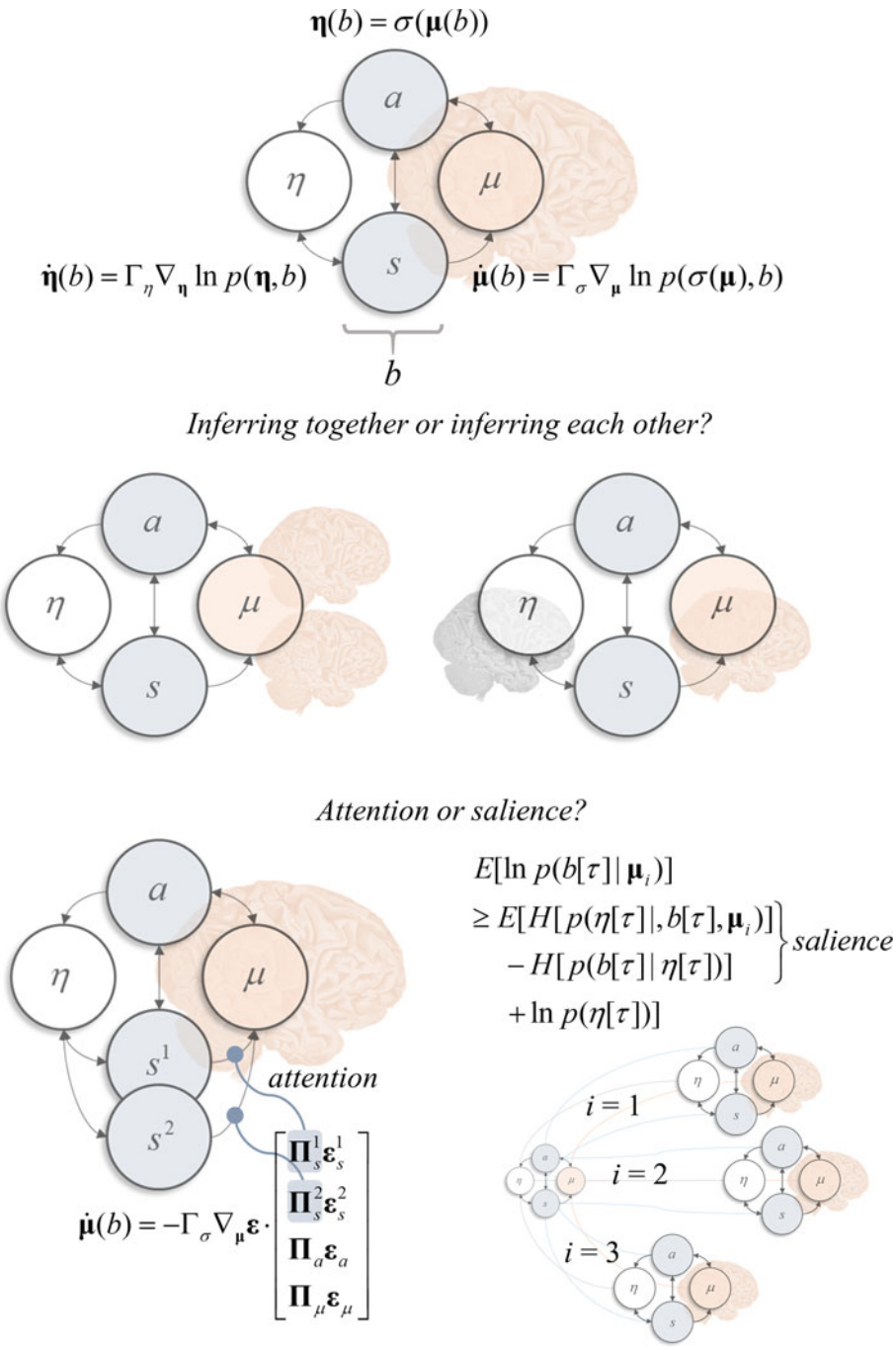


Figure 1 (Parr) Blankets, inference, and attention. This figure sets out the various ways in which we can "choose" a Markov blanket. The top image sets out the conditional dependencies between internal (μ), external (η), and blanket (b) states, where arrows show the direction of causation. Blanket states comprise active (a) and sensory (s) states. Given a mapping (σ) between the most likely internal and external states ($\boldsymbol{\mu}$ and $\boldsymbol{\eta}$, respectively) as a function of b, both internal and external states can be viewed as performing a gradient ascent on the same log-probability density. This is the non-equilibrium steady state density, or generative model. Technically, these dynamics maximize the evidence for the generative model, and are sometimes described as "self-evidencing" (Hohwy 2016). The middle schematics show two alternative delineations of a Markov blanket in a social context. Either two brains sit within the same Markov blanket and can be thought of as jointly inferring their environment (η), but not each other, or they could be thought of as being on either side of a blanket. In this setting, each individual draws inferences about the other as the other is part of the environment. The bottom images set out the distinction between attention and salience from the perspective of a Markov blanket. The image on the left shows multiple sensory states (superscripted) and shows the form of the internal state dynamics (under Gaussian assumptions). The magnitude of the influence of each s on $\boldsymbol{\mu}$ depends on the precision $\boldsymbol{\Pi}_s$ with which that sensory state depends on external states. The image on the right shows a different sort of selection, cartooning three alternative paths (indexed by i) that can be scored in terms of the expected log probability of the blanket states following that trajectory (where $x[\tau]$ means the path x follows over time). Note that the expression for the bound on this probability includes a relative entropy (the difference between the two H terms) that quantifies the salience or information gain expected along that trajectory. Interestingly, the entropy of blanket states conditioned on external states is inversely related to $\boldsymbol{\Pi}_s$, highlighting the point of connection between attention and salience that often underwrites their conflation.

Veissière et al. provide a compelling account of the use of variational principles (specifically, active inference) to provide a formal basis for understanding culture and cognition. This affords an opportunity to exploit the tools that come along with active inference in the context of social sciences. As an example of this, Veissière et al. highlight the importance of shared "regimes of attention," and the way in which the elusive concept of attention may be pinned down in formal terms. This relies upon the concept of a Markov blanket (Pearl 1998) that statistically insulates the inside of a system from the outside. In this commentary, we discuss three perspectives on *choosing a Markov blanket* (Fig. 1). The first is the choice that we as scientists make when deciding upon our object of study. It is this that underwrites the application of Bayesian mechanics across interdisciplinary boundaries. The second is the implicit choice made by internal states of a Markov blanket as to which blanket states most influence their dynamics. This is an attentional process modulating influences from the outside in. The third is a choice between hypothetical blankets in a dynamical setting. It is this that determines how inside influences outside and gives rise to the concepts of salience and novelty (Clark 2017a).

Selecting an object of study involves explicitly or implicitly segregating that thing from other things. This selection defines a Markov blanket that mediates interactions between that object and everything else. If we select the nervous system, the blanket comprises sensory receptors and muscles, whereas internal states are those neurons that respond to the former and drive changes in the latter. Because the system as a whole persists over time, it must

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

be at (non-equilibrium) steady state. This implies dynamics that correct deviations from a steady state density, ensuring the system continues to occupy regions of high probability. Because the internal states of the system are coupled to external states, but only via the blanket states, it appears that internal states vicariously infer what is happening in the outside world (Friston 2019). This Bayesian mechanical perspective says something intuitively sensible from the perspective of a nervous system: the brain draws inferences about the world based upon its sensory input.

We can frame the dynamics of any blanketed system in the same way. For example, if we take a more fine-grained approach we could treat an individual neuron as our system of interest (Palacios et al. 2019), with its blanket comprising pre- and post-synaptic membrane potentials. The interesting thing about this is that the other neurons in the brain, previously internal states, have become external states. This means they have gone from *performing* inference to *being* inferred. Nothing has changed in the dynamics of the system itself, but by changing our perspective, we change the inference problem (i.e., generative or internal model) that is being solved.

This has two interesting consequences. The first is that it endorses the use of inferential formalisms at a range of scales (Kirchhoff et al. 2018), whether cellular, cognitive, or cultural. The second is that the choice we make as to where the Markov blanket is drawn has consequences for how we think about the interactions between different parts of a system. Bringing this back to the question of cognition and culture, we could think of many individuals as the internal states of a system jointly inferring their shared environment or we could think of an individual drawing inferences about other individuals. In either setting, the challenge going forward is to set out the generative model from which inferential dynamics at a cultural scale emerge.

We now take the perspective from inside a blanket, and ask what it means to choose between alternative blanket states. This choice has two parts to it. The first is deciding which blanket states should influence internal state dynamics. The second is deciding between hypothetical trajectories the blanket states could follow. The distinction between these is formally identical to that between attention and salience attribution; two important features that emerge from solving a generative model.

Starting with attention, imagine we have multiple sensory states in a blanket. The degree to which each of these may be used to draw inferences depends upon the precision (inverse variance) with which they are predicted by external states, under the non-equilibrium steady state density. This manifests as a form of gain control, where those sensory states that are precisely predicted by external states are amplified relative to others in setting internal dynamics, exactly as in attentional gain control (Desimone 1996; Hillyard et al. 1998; Shipp 2016).

Attentional gain must be distinguished from the process of salience attribution (Parr & Friston 2019). The latter involves overtly (Rizzolatti et al. 1987) acting upon the world to obtain more information (Mirza et al. 2016). This requires the capacity to score alternative trajectories (e.g., eye movements to different locations) in terms of their anticipated information gain (Lindley 1956). The relative probability for each trajectory is bounded by an expected free-energy functional. This functional favours those trajectories for which the salience is greatest (Parr et al. 2020). As such, the process of salience attribution may be formalized as the process of choosing between alternative blanket trajectories.

Once a Markov blanket has been drawn around a system of interest, this licences an inferential interpretation of its dynamics. The choice of blanket tells us what is being inferred (external states) and what is doing the inferring (internal states). The advantage of appealing to a formalism of this sort is that it provides an opportunity to precisely define and simulate cognitive (and cultural) processes. We highlight the examples of attention and salience. These may be understood through the metaphor of a scientist who decides upon the quality of her data (i.e., attention) before drawing inferences, and then decides upon the next experiment to perform (i.e., salience) to optimize the quality of future data.

Acknowledgments. TP is supported by the Rosetrees Trust (Award Number 173346).

Social epistemic actions

Giovanni Pezzulo[a], Laura Barca[a], Domenico Maisto[b] and Francesco Donnarumma[a]

[a]Institute of Cognitive Sciences and Technologies, National Research Council, Rome 00185, Italy and [b]Institute for High Performance Computing and Networking, National Research Council, Naples 80131, Italy.
giovanni.pezzulo@istc.cnr.it
https://sites.google.com/site/giovannipezzulo/home
domenico.maisto@icar.cnr.it https://www.icar.cnr.it/persone/maisto/
laura.barca@istc.cnr.it
https://sites.google.com/site/laurabarcahomepage/home
francesco.donnarumma@istc.cnr.it
https://www.istc.cnr.it/it/people/francesco-donnarumma

doi:10.1017/S0140525X19002802, e113

Abstract

We consider the ways humans engage in *social epistemic actions*, to guide each other's attention, prediction, and learning processes towards salient information, at the timescale of online social interaction and joint action. This parallels the active guidance of other's attention, prediction, and learning processes at the longer timescale of niche construction and cultural practices, as discussed in the target article.

Veissière et al. convincingly argue that we collectively build niches and cultural practices, which guide our attention towards salient information, facilitating cultural learning, and the acquisition of shared expectations about norms and conventions. This allows us to "acquire culture by being immersed in specific, culturally patterned practices that modulate salience."

Here, we consider that not only we guide each other's attention, prediction, and learning processes towards salient information at a long time-scale (e.g., niche construction); but also at a faster time-scale (e.g., during teaching and joint action).

In active inference, the salience of stimuli depends on both their quality and on the agent's belief about the world (Parr & Friston 2017a). Salient stimuli are those that are expected to change the agent's belief, such as those about which the agent is uncertain, but (once gathered) would clearly disambiguate the agent's alternative hypotheses. Conversely, stimuli that were predicted, are of poor quality or ambiguous (and if gathered, would not disambiguate the agent's hypotheses) have little salience, as they would not change the agent's belief significantly.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

By directing our information-gathering actions (e.g., saccades) to high-salience locations or stimuli, we gain the most (in information terms) from our engagement with the environment. Information-gathering actions are sometimes called *epistemic actions*, as they aim at changing one's contextual beliefs (e.g., when exiting from an unknown underground parking, looking around to resolve uncertainty about one's current location); and are distinguished from *pragmatic actions*, which aim to achieve goals (e.g., drive to an intended destination, after having resolved the above contextual uncertainty) (Friston et al. 2015; 2017; Pezzulo et al. 2015; 2018).

Crucially, during social interactions, we can perform epistemic actions for the sake of others – or *social epistemic actions*. These actions aim at gathering salient (belief-changing) information for others and hence are quintessentially communicative. However, they don't need to be verbal, but can exploit sensorimotor channels, that is, sensorimotor communication.

An extensive body of research shows that during social interactions, we modify our behavior in communicative ways, to render it more "legible," easier to understand and to predict by others (Pezzulo et al. 2018). A well-known example is that of a mother who amplifies her speech (e.g., the vowels) and exaggerates her bodily movements, when interacting with her child, that is, *motherese* and *motionese*. Such sensorimotor communication may simultaneously help capture the child's attention and simplify her learning task (e.g., by stressing what is salient).

Sensorimotor communication is ubiquitous during social interaction. It does not necessarily require "specialized" behaviors, such as pointing or gazing at some object. Rather, virtually any action can be used (or modified) to convey sophisticated communicative messages. Even simple actions, such as passing a glass to somebody can express (intentionally or unintentionally) love, hate, or deference. These ("hidden") emotional states can be conveyed by – and inferred from – subtle kinematic cues, for example, slightly faster or slower arm movements (Becchio et al. 2012; Pezzulo et al. 2013).

Sensorimotor communication is especially effective during joint actions. For example, during a joint action as simple as moving a table together, we can push the table to signal in which direction we want to go or where we want to place the table. During more complex social interactions, such as during a soccer match, we can exaggerate our movements to help our teammates inferring our intention (e.g., where we want to pass the ball), or hide it from opponents (by feinting). Sensorimotor communication goes also beyond the body. For example, when driving a car, we can decelerate or move to the side of a road, to signal that we want to leave room to other drivers (Chater et al. 2018; McEllin et al. 2018; Pezzulo & Dindo 2011; Vesper et al. 2011).

These examples can be conceptualized within active inference, as *social epistemic actions* that unveil salient information, for the sake of somebody else. Consider the case of a person, who is helping his roommate to move a table, but does not know whether she wants to place it to the left or the right of a chair. He can engage in *mindreading*, to infer the roommate's intention ("hidden state") based on her movements ("observables"). Simultaneously, the roommate can select a plan that is maximally informative about her intentions (e.g., push to the left earlier and harder) – an example of *social epistemic action*, to optimize interactive success. Interestingly, the roommate helps her coactor inferring her intention, by "surprising" him, because her "exaggerated" force deviates from the most likely (predicted) action plan. It is this unpredicted (and informative) deviation that has high salience and signals the teammate's communicative intention to place the table to the left (Pezzulo 2011; Pezzulo et al. 2013). Importantly, not all unpredictable behaviors are salient, but (like standard epistemic actions) only those that resolve the coactor's uncertainty. Hence, selection of effective social epistemic actions requires tailoring them to the current interactive situation. Recent research showed that coactors consider elements such as others' uncertainty, and what they can or cannot see, to modulate their social epistemic actions. Furthermore, social epistemic actions can be bidirectional, with coactors continuously helping each other inferring their intentions and predicting their actions (Leibfried et al. 2015; Pezzulo & Dindo 2011; Pezzulo et al. 2017; Vesper & Richardson 2014).

In sum, we flexibly use social epistemic actions to drive others' attention on patterns and regularities that we want them to infer (or learn). Social epistemic actions may thus work synergistically with "culturally-patterned practices" to afford social and cultural learning. Of note, although social epistemic actions can be realized verbally, here we focused on sensorimotor realizations, which (in certain conditions) may signal more directly what and where salient information is. Furthermore, humans may be exquisitely sensible to recognizing under which conditions certain actions or demonstrations are executed for pedagogical purposes, and hence learn faster and more efficiently under these conditions (Csibra & Gergely 2011).

Thinking through others' emotions: Incorporating the role of emotional state inference in thinking through other minds

Ryan Smith[a] and Richard D. Lane[b]

[a]Laureate Institute for Brain Research, Tulsa, OK 74136 and [b]Department of Psychiatry, University of Arizona, Tucson, AZ 85724.
rsmith@laureateinstitute.org lane@psychiatry.arizona.edu

doi:10.1017/S0140525X19002644, e114

Abstract

The active inference framework offers an attractive starting point for understanding cultural cognition. Here, we argue that affective dynamics are essential to include when constructing this type of theory. We highlight ways in which interactions between emotional responses and *the perception of* those responses, both within and between individuals, can play central roles in both motivating and constraining sociocultural practices.

In their thoughtful review, Veissière and colleagues motivate active inference as an integrative framework for understanding and modeling the dynamics of cultural cognition. Conspicuously absent was discussion of the role of emotions in "thinking through other minds." However, emotions plausibly underwrite the motivation, maintenance, and enforcement of many sociocultural norms and practices (e.g., feeling an automatic aversion to breaking a social norm, or anticipating that others would react with anger/disappointment if one failed to engage in a cultural practice). Here, we argue that there is an important opportunity to expand their model by explicitly incorporating emotion. Motivated by the active inference framework, we will focus on dynamic bidirectional

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

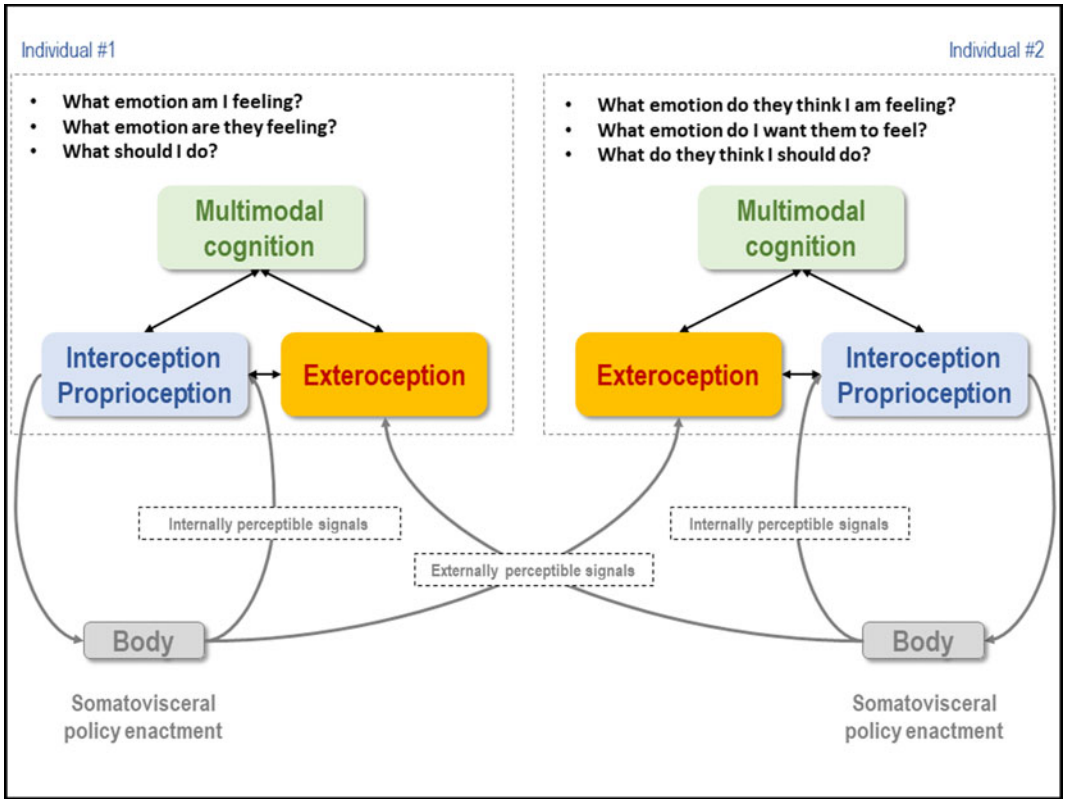


Figure 1. (Smith & Lane) Depiction of "thinking through others' emotions" as an extension of the "thinking through other minds" framework. Based on the (implicitly or explicitly) perceived thoughts and feelings of others, quick/involuntary somatovisceral (e.g., valenced changes in facial expression, body posture, and autonomic state) and cognitive (e.g., selective attention) policies are enacted and perceived by both self and others. Subsequent inferences about one's own emotional state and the emotional states of others (both implicit and explicit) then further inform sociocultural decision-making (e.g., conforming to social norms and engaging in cultural practices).

interactions between two broad emotion-related processes: affective response generation (cast as policy selection) and affective response perception (cast as Bayesian state inference).

As elsewhere (Smith et al. 2018c), we use the broad term "affective response generation" to denote the flexible engagement of multiple quick/involuntary changes across visceromotor, skeletomotor, and attentional states in response to (current, remembered, or imagined) interoceptive or exteroceptive stimuli of perceived significance to an organism's needs, goals, and values. For example, consider the simultaneous elicitation of unpleasant changes in posture, facial expression, autonomic arousal, threat-biased attention, and avoidance motivation that can quickly ensue in response to a simple social gesture. Although not always, and not always within awareness, these types of multimodal internal responses often occur in response to the perceived thoughts and feelings of others, and they can play an essential role in the sociocultural phenomena that Veissière and colleagues discuss. Within the active inference framework, recent work (Allen et al. 2019; Smith et al. 2019a; 2019b) has shown how these responses can be cast in terms of multimodal policy selection. That is, based on a set of (e.g., interoceptive or social) signals, the states of the world, the body, and of the minds of self and others can be (implicitly or explicitly) inferred, which can then engage predictions about how these states will change over time if different policies (i.e., different sets of visceromotor, skeletomotor, and attentional changes) were selected. The "affective response" that is generated then corresponds to the enactment of the policy predicted to lead to the states most consistent with an organism's preferences (e.g., perceiving social approval) – often involving automatic attention to salient sources of information (e.g., the locations of friends and enemies) and visceromotor adjustments based on the predicted metabolic demands associated with navigating the environment so as to attain preferred states (e.g., staying close to friends and away from enemies).

There are many cases in which this type of quick/involuntary policy selection process can play adaptive (and often unintentional) social roles. For example, although typically not intentional, the automatic production of tears (crying) can elicit helpful social support from others. Unintended changes in posture and muscle tension in response to social norm violations can also convey implicit signals to others about the probability that aggressive action will ensue if such violations continue.

However, the presence of an affective response does not entail that it will be perceived/interpreted correctly by oneself or others. The ability to infer the interoceptive and emotional states of self and others, based on the internally/externally observable sensory consequences that follow from the enactment of affective policies (e.g., perceived heart rate or facial expression changes), has also been formulated as Bayesian inference within the active inference framework (e.g., inferring the probability that an individual is sad based on information about their body state and the context [Barrett 2017; Smith et al. 2018b]). When interacting with affective response generation, affective response perception can produce iterative bidirectional interactions between the minds and bodies of

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

multiple individuals. For example, an individual could feel an aversion to participate in particular social practices, but then also react with frustration because they don't want to feel that aversion. Or an individual might be happy because they believe they are meeting others' sociocultural expectations, but then react with disappointment when they perceive that others are displeased.

The relevant inferential dynamics also appear to play out at multiple hierarchical levels. For example, there is evidence that individuals automatically simulate the body states they perceive in others in order to infer what emotions they are feeling (Niedenthal 2007). There are also important cases where correctly inferring the emotions of others requires the deployment of additional higher-level knowledge (e.g., "even though I like this restaurant, she will be sad if we go there"). These different levels also plausibly facilitate different social dynamics. For example, simulating the discomfort of another individual's tense/shaky posture could promote automatic empathic responding (i.e., making the other individual feel better would make you feel better [Lamm et al. 2011]). In contrast, the explicit semantic inference that what another person is feeling corresponds to the concept FEAR can facilitate important social inferences about both the past and the future by drawing on emotion knowledge, such as the likely causes and consequences of that state (e.g., "it was likely caused by a perceived threat" and "the person is now likely to try to avoid that threat") – which can then inform subsequent decision-making (Fig. 1).

Successfully navigating the hypersocial human niche requires iterative and nested use of these processes – in which one's goals can only be accomplished by predicting the emotions of others in the future under different courses of action (e.g., "he's feeling angry because I'm driving below the speed limit – if I speed up then he will calm down"). This in turn leads to complex and continuous feedback loops that can facilitate both adaptive (Smith et al. 2019c) and maladaptive (Smith et al. 2018a) social dynamics. The resulting dynamics lead to a type of "thinking through others' emotions" that depends jointly on interoceptive and exteroceptive inference and on bidirectional brain-body interactions across individuals throughout a society. We put forward these additional affective dynamics as essential to completing the authors' account of thinking through other minds.

A deeper and distributed search for culture

Paul S. Strand

Department of Psychology, Washington State University Tri-Cities, Richland, WA 99354
pstrand@wsu.edu https://psychology.wsu.edu/people/faculty/paul-s-strand/

doi:10.1017/S0140525X19002693, e115

Abstract

The target article does not address the neural mediation of complex social behavior. I review evidence that such mediation may be compatible with proposed Bayesian information-processing principles. Notably, however, such mediation occurs subcortically as well as cortically, concerns reward uncertainty and information uncertainty, and impacts culture via group-level payoff structures that define individualism and collectivism.

The formal-mathematical constructs that are the focus of the target article are not grounded in nervous system functioning. Conciliatory efforts would call into question informational uncertainty as a singular phenomenon mediated by higher-level neural processes. That is, because a form of informational uncertainty – *reward uncertainty* – is subcortically mediated (Anselme & Güntürkün 2019; Hart et al. 2015). Furthermore, evidence suggests that reward uncertainty gives rise to complex behavioral phenomena that can form the basis for cultural constraints on behavior (Strand et al. 2019). These ideas are expanded upon in what follows, in the hopes that formal-mathematical models such as those presented in the target article will, in the future, enumerate the place of reward uncertainty within the domain of informational uncertainty.

To understand reward uncertainty and its importance for social behavior, let's review a recent study concerned with the formation of the attachment patterns (Beckes et al. 2017). *Attachment patterns* describe social–behavioral propensities of individuals arising in response to relationships with caretakers or supportive others. Therefore, attachment patterns are relevant for considering how information affects social behavior. As we discuss below, they also align with the dominant cultural institutions, collectivism, and individualism.

Beckes et al. (2017) experimentally investigated how supportive social responsiveness influences attachment-related behavior using a shock threat support-seeking paradigm. Participants were adults who ostensibly sought help from another participant ("supporter") seated in another room. They pressed a button to indicate their need for help every time a shock threat signal appeared. The supporter could stop the imminent shock in response to this support-seeking behavior. In a randomized design, one group of participants experienced a supporter who acted consistently and predictably to prevent the shock. Another group experienced a supporter who acted inconsistently and unpredictably to prevent the shock. Therefore, in response to support-seeking behavior, the former group experienced reward certainty (continuous reinforcement) and the latter group reward uncertainty (intermittent reinforcement).

Study results revealed that exposure to reward uncertainty led to behavior toward the supporter that was characterized by attitudinal ambivalence and high levels of approach motivation – a behavioral constellation consistent with an insecure-anxious attachment pattern (Ainsworth et al. 1978). By contrast, exposure to reward certainty led to unambiguous positive attitudes and moderate approach motivation toward the supporter – the behavioral equivalent of a secure attachment pattern. Not studied by Beckes et al. (2017), insecure-avoidant is a third primary attachment pattern. It is thought to be induced by an extinction schedule for security-seeking behavior and characterized by attitudinal neutrality and low approach motivation to caretakers.

The Beckes et al. (2017) findings reveal the existence of complex socio-behavioral response patterns induced by simple forms of environmental stimulation (i.e., intermittent and continuous schedules of reinforcement). Moreover, neuroscientific evidence reveals that these behavioral responses are mediated by dopamine activity within the subcortical nucleus accumbens (Hart et al. 2015). Therefore, a confluence of neural and behavioral evidence reveals the existence of forms of complex social behavior that are *schedule-induced* and subcortically mediated (Anselme & Güntürkün 2019). As such, they do not reflect the "first principles" of the target article, which are cortically mediated and include perspective-taking and mind reading. I return to the

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

issue of neural mediation after considering the importance for culture of the attachment patterns.

The impact of the attachment patterns on culture is suggested by the results of a world-wide review of their distribution across collectivist and individualist cultures (Mesman et al. 2016). Findings reveal that collectivist cultures have a higher relative percentage of insecure-anxious individuals. Such individuals tend to prioritize interactions within strong-tie social networks (reflecting high approach motivation), which is a defining characteristic of collectivist cultures (Yamagishi & Hashimoto 2016). Individualist cultures, on the other hand, have a higher relative percentage of insecure-avoidant individuals. Such individuals tend to prioritize interactions in weak-tie social networks (reflecting low approach motivation), a defining characteristic of individualist cultures. Therefore, participation in strong-tie relationships is relatively advantageous in collectivist settings, and participation in weak-tie relationships is relatively advantageous in individualist settings. In this way, culture is defined not by co-constructed minds but, rather, foundationally, by group-level payoff structures that constrain the reward maximization behavior of individuals. Those constraints reflect the aggregate of the behavioral propensities of the individuals who comprise the group (Strand et al. 2019).

Consistent with the target article, work cited above supports an informational account of reward uncertainty. Behaviorally, a richer reward schedule (continuous reinforcement) led to less extreme approach motivation than did a leaner schedule (intermittent reinforcement; Beckes et al. 2017). Neurally, dopamine release was highest at maximum reward uncertainty (probability of reward = 0.5), and lowest at both extremes of certainty (probability of reward = 0.0 and 1.0; Hart et al. 2015). Therefore, the behavioral response is determined by the informational value of the schedule, not the reward amount. The question remains: Does the formal-mathematical model of the target article align with what is known about reward uncertainty and the neural mediation of complex behavior?

The dark side of thinking through other minds

Sander Van de Cruys[a] and Francis Heylighen[b]

[a]Laboratory of Experimental Psychology, KU Leuven, BE-3000 Leuven, Belgium and [b]Center Leo Apostel, Vrije Universiteit Brussel, Brussels, Belgium.
sander.vandecruys@kuleuven.be www.sandervandecruys.be
fheyligh@vub.ac.be http://pespmc1.vub.ac.be/HEYL.html

doi:10.1017/S0140525X19002796, e116

Abstract

We show that TTOM has a lot to offer for the study of the evolution of cultures, but that this also brings to the fore the dark implications of TTOM, unexposed in Veissière et al. Those implications lead us to move beyond meme-centered or an organism-centered concept of fitness based on free-energy minimization, toward a social system-centered view.

TTOM and the underlying FEP framework allow us to revise and refine, not only theory of mind accounts, but also theories of the evolution of culture, most notably memetics. At least at three levels, TTOM/FEP provides important correctives to reinvigorate this field. In doing this, Veissière et al.'s account explains how the "imagined communities" we live in (Anderson 1983/2006), come to be. In this commentary, we specifically consider the darker implications of this, as these remain unexplored in the target article.

First, the analysis of Veissière et al. shows that the concept of memes as independent units of cultural information is rather deceptive. An FEP-based account of culture highlights that ideas or hidden causes are not insular units but parts of hierarchically and laterally-structured belief networks, or narratives. Such networks also include so-called auxiliary hypotheses that can take the blame when other, high precision beliefs are under threat of being disproven. The networks even include socially-expected ways of sampling evidence (expected precision or epistemic value of different sources). These observations suggest that cultural beliefs are socially constructed and *self-sustaining*. Thus, fabricated beliefs such as conspiracy theories can spread easily and take root (Gershman 2019).

Second, Veissière et al. rightly call attention to the importance of embodied cultural practices (rather than just "ideas," as in memetics) in the evolution of culture. Indeed, practices have primacy in steering not just behavior but thought. This "practice before ideology" principle can be seen in enculturation through religious rituals. Heylighen et al. (2018) observe that: "the undeniable act of praying to God can only be safeguarded from cognitive dissonance by denying any doubts you may have about the existence of God." In FEP terms, the irrefutable perceptual evidence created by the active practice can only be explained away by adopting the ideological "hidden causes." In rituals (as in many cultural practices), actions are triggered by cultural markers in the environment – I do it because others like me do/did it – circumventing explicit thought but at times also the actual interests of the participant. Here, practices become a tool for control of individual action by the social system (conformity pressure).

Third, TTOM/FEP may provide a unified selection criterion ("fitness") for the evolution of cultures. Culture constrains the behavioral paths for its individual members, generally because its practices and narratives have shown to be efficient free-energy reduction vehicles, for "agents like you." Ideas and practices that reduce free energy more efficiently tend to spread and become dominant in a culture. These ideas could concern hidden causes of the environment and the challenges it provides (e.g., a god causing thunderstorms) but also hidden causes of the behavior of other individuals in your community. This gives cultural ideas a circular, self-reinforcing character. For example, the cultural expectation that sinning requires guilt and atonement reduces the free energy of the harmed party, but guilt also becomes a hidden cause efficiently explaining away someone's behavior in the eyes of others belonging to the culture. However, note that the success of these expectations depends on the conservative perpetuation of the culture, and the exclusion of "dissident" behavior.

Similarly, ideologies such as religion or nationalism, as interconnected sets of hidden causes and shared expectations and practices, are an efficient means of free-energy minimization. As Atran and Ginges (2012) remark, most religions have at their core a limited set of principles (expectations) that they consider "sacred." In essence, to be sacred implies unconditionality. Indeed, expectations that are independent of contextual parameters provide a simple, dependable (high precision) foundation for how to act in and explain the (social) world. It makes

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

these principles into very powerful, socially fulfilled hidden causes. The same free-energy minimization logic explains why strictly patterned (hence predictable) religious rituals are especially successful at important transition points in life (such as the transition to adulthood), characterized by higher uncertainty about how one should act. In the same vein, Hogg (2014) reports evidence that individuals that experience high personal uncertainty (e.g., adolescents going through identity problems) tend to strongly identify with a group and (radical) ideology to easily resolve their self-uncertainty. Examples can even be found of cultures systematically plunging their members into uncertainty to increase allegiance. Thus, cultures and their "sacred" rules often actually harm their members, hence outright increasing their free energy. Think, for example, of rules inducing genital mutilation, suicide terrorism, honor killings, or more mundanely, chronic stress because of a ruthless, sacred rule of productivity.

The above examples show that a meme-centric concept of fitness will not do (Ramsey & De Block 2015), but, more interestingly, they also suggest that a purely organism-centric concept of fitness (organism-centered free-energy minimization) is unsatisfactory to explain the power of cultures on their members. Indeed, internalized and environmentally anchored cultural expectations (behavioral "rules") often take on a life of their own, not necessarily benefitting the individual that follows them, but rather maintaining the very system of social ideas and practices they are part of. Luhmann (1986) has argued that social systems should be seen as autopoietic, organism-like agents that, via their human constituents, actively counteract any deviation from their organization, so as to ensure the continuation and self-regeneration of the system (Heylighen et al. 2018). Hence, these social systems seem to also reduce their free energy, consistent with a multiscale formulation of the FEP (Ramstead et al. 2018). On the one hand, the relationship between individual and social system is one of symbiosis or mutual benefit, with social systems providing means for reducing free energy to the individual through coordination of action and prevention of conflicts. On the other hand, social systems, via TTOM mechanisms, can also veer into dogmatism, radicalism, and mind control that suppresses individual expression, creativity, and well-being (Heylighen et al. 2018). We believe that the account of Veissière et al. should also provide insight into this dark side of cultural phenomena.

Participating in a musician's stream of consciousness

Björn Vickhoff

Sahlgrenska Academy, University of Gothenburg, 413 90 Göteborg, Sweden.
bjorn.vickhoff@aniv.gu.se

doi:10.1017/S0140525X19002759, e117

Abstract

Do we acquire culture through other minds, or do we get access to other minds through culture? Music culture is a practice as well as the people involved. Sounding music works as a script guiding action, as do, to varying degrees, many rituals and customs. Collective co-performance of the script enables intersubjectivity, which arguably contributes to the formation of subcultures. Shared-emotional experiences give material to the narrative of who we are.

Do we acquire culture through other minds, or do we get access to other minds through culture?

"Culture" usually refers to the arts or to a people sharing customs and codes. In the case of music, these understandings are intertwined in two ways:

1. Music is governed by customs defining the style. It almost invariably follows some system creating expectancies, implicitly known and applied by musicians and listeners.
2. Musical styles typically shape *neo-tribes* (Bennett 1999) of followers who share a spectrum of social codes and customs.

Music culture, thus, is a practice as well as neo-tribes of followers. Both aspects are governed by patterned behaviour, generating anticipation.

The Austrian phenomenologist Alfred Schütz, contemplating music listening, once wrote:

> Although separated by hundreds of years, [the listener] participates with quasi-simultaneity in [the composer's] stream of consciousness, by performing with him step by step the ongoing articulation of his musical thought. The beholder, thus, is united with the composer by a time dimension common to both (Schütz 1951).

Schütz puts the listener in the head of the musician. He is not proposing that listeners *follow* or *imitate* the composer, but *perform* the music *with* the composer.

Co-performance with a musician implies shared brain activity as well as a shared time dimension. Except for the parallel activation of the auditory systems, several mutual pre-motor areas are activated. The shared time dimension is induced through entrainment (synchronization to musical beats). Synchronization demands prediction. Periodical sounds produce bursts of neural oscillations on the beats. These bursts continue when the stimulus is omitted (Tal et al. 2017). They are, thus, predictions. "Entrainment," says anthropologist Judith Becker, "is the strongest form of interaction … It actualizes a supra-individual state" (Becker 2001).

Synchronization also depends on familiarity with the piece and/or style. But, even when the prediction is wrong, synchronized activities occur, namely the processing of *prediction errors*. Deviances from musical customs elicit oscillatory reactions (the *event-related potential* ERAN), not only in the listener, but even in the musician (who planted the surprising event) (Koelsch et al. 2019). Surprises attract attention. In a sequence, changes between the expected and the unexpected are felt. This exemplifies listening as a participation in the musician's stream of consciousness.

Overt synchronization engenders pro-social behaviour (Repp & Su 2013). Further, the ability to synchronize is improved by oxytocin – a peptide associated with social integration (Gebauer et al. 2016). These findings may be associated with bird courtship, where the male and the female of some species engage in synchronized rituals, and with human courtship, where the dance floor provides a scene for synchronization. The synchronized other, thus, can be considered an affordance – an epistemic resource, reducing uncertainty concerning the other's suitability as a partner. It confirms: same species, same fitness, and same interest.

Subjectivity involves emotions, and inter-subjectivity thus entails shared emotions. With an emotional completion, the

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

TTOM theory would catch how we, not just *think*, but *live* through other minds.

Could emotions be computed along the lines of *the free-energy principle* (FEP)? *Interoceptive inference* entails that neural activity caused by visceral reactions is compared to predictions reflecting homoeostasis (Seth & Friston 2016). The only ascending information is the degree of deviance. Is it possible to access the rich palette of emotions this way? It has been argued from an enactive perspective that "emotional experience simply cannot be reduced to a frame of reference" (Roesch et al. 2012).

In the circumplex model of affect, emotions are plotted in a two-dimensional space defined by valence and arousal coordinates (Russell & Pratt 1980). Valence, according to FEP, could be defined as "the negative rate of change of free energy over time" (Joffily & Coricelli 2013). This means that reduction of uncertainty is rewarding. This definition offers a solution to the "dark-room problem" (Friston et al. 2012): If there is no uncertainty to reduce, the organism cannot be happy. Valence over certainty must follow an inverted U-curve, starting in boredom, ending in chaos, and peaking at an optimal level of *active inference*, as defined by Seth and Friston (2016). Neither the uneventful, nor chaos can provide the epistemic affordance needed for active inference. Arousal could be caused by the degree and precision of the prediction error. This is in accordance with the pattern of the *piloerection* reaction to music, which is often caused by music expressing longing and triggered by an unexpected change in the music (e.g., a change of tonality). Here, physiologically assessed arousal, which may reflect uncertainty, is followed by relaxation, which may reflect certainty (Vickhoff et al. 2012). The experience is delightful.

Musical emotion is an excellent subject to test interoceptive inference. Music is computable, it plays with expectancies and it makes us happy without obvious rewards.

A short answer to the initial question: Music exemplifies that we do not just acquire culture through other minds. It is a reciprocal dependency. Music, written, or sounding, is a script guiding action. So are, to varying degrees, many rituals and customs. Collective co-performance of the script enables inter-subjectivity, which arguably contributes to the formation of neo-tribes. Shared-emotional experiences give material to the narrative of who we are.

A unified account of culture should accommodate animal cultures

Andrew Whiten

Centre for Social Learning and Cognitive Evolution, School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK.
a.whiten@st-andrews.ac.uk

doi:10.1017/S0140525X1900270X, e118

Abstract

Discoveries about social learning and culture in non-human animals have burgeoned this century, yet despite aspiring to offer a unified account of culture, the target article neglects these discoveries almost totally. I offer an overview of principal findings in this field including phylogenetic reach, intraspecies pervasiveness, stability, fidelity, and attentional funnelling in social learning. Can the authors' approach accommodate these?

Despite promising a "unified account" of culture, the authors neglect to define what they take culture to encompass. Because they cite Henrich (2015) with approval, perhaps we can borrow his definition of culture as "the large body of practices, techniques, heuristics, tools, motivations, values and beliefs that we all acquire while growing up, mostly by learning from other people" (p. 3). Substituting "individuals" for "people," culture thus defined has been discovered to be widespread among non-human animals (henceforth "animals") (Whiten 2017a). Yet, it receives virtually no mention in the target article. This lacuna contrasts starkly with another recent peer-commentary articles offering broad theories of culture (Tamariz 2019; Osiurak & Reynaud 2020) that do incorporate core animal literature (I note there is virtually no overlap in the sources cited in these theoretical offerings!). Here, I outline dimensions of animal culture that I suggest any unifying account of culture must recognize.

First, phylogenetic reach. Traditions meeting the above specifications have been identified in hundreds of studies of mammals (particularly primates [Whiten & van de Waal 2018] and cetaceans [Whitehead & Rendell 2015]) as well as birds (Aplin 2019), and fish (Laland et al. 2011). Transmission chain experiments have demonstrated inheritance across multiple "cultural generations" (i.e., from individual or group A, to B to C, etc.); the content of these ranges from foraging techniques in chimpanzees and children (Horner et al. 2006) to those of bumble bees (Alem et al. 2016) and mate choice preferences in fruit flies (Danchin et al. 2018; Whiten 2018a).

Second is intra-species pervasiveness. Whitehead and Rendell (2015, p. 17) conclude from their book-length survey that "culture, we believe, is a major part of what the whales are." Whiten and van de Waal (2018) distinguish three main phases of cultural acquisition in primates: first, juveniles learn much about foraging and other activities from their mother, whom they may accompany closely for years; in a second phase learning expands to a wider social circle, when young males may apprentice themselves to adult males who exploit a somewhat different cultural foraging niche than their mothers; and finally, in a third phase dispersing individuals learn from individuals in their new group who are already informed about local resources and group dynamics. Schuppli and van Schaik (2019) and Whiten (2019a; 2019b; 2019c) infer from records of juvenile orangutans' close peering at adult activities that the number of behaviours so learned may exceed 190. "It seems that immatures learn virtually all of their skills socially," they conclude (p. 5). What animals learn from the accumulated knowledge of conspecifics spans dietary profiles, foraging techniques, predator avoidance, mate choice, courtship behaviour, vocal communication, migration routes, tool use, social customs, circadian rhythms, and locomotion styles.

These varied domains of social learning raise the prospect that animal cultures may be constituted of multiple traditions, paralleling the way we refer to a whole array of traditions (technologies, diet, communications, etc.) when we contrast any two human cultures (Whiten & van Schaik 2007). Such patterning, including regional cultural variations, has so far appeared most extensive in great apes (Schuppli & van Schaik 2019; Whiten 2017b) but has also been documented in cetaceans (Whitehead & Rendell 2015)

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

and likely applies to many birds where traditions span such diverse contexts as song dialects, migration, and foraging (Aplin 2019). Veissière et al. state that "it is precisely because of the existence of inter group behavioural and cognitive variations that arise through social learning … that we can speak of culture" – but surely not: certainly Henrich's definition does not require this. A universal behaviour would still be cultural, if routinely acquired by social learning from others (Schuppli & van Schaik 2019).

A third important discovery about animal cultures concerns stability and fidelity over time. Archaeological excavations revealed chimpanzee nut-cracking technologies based on natural stone hammers at a level corresponding to 4,300 years ago, at sites where the skill still occurs today (Mercader et al. 2007). How many contemporary human traditions have shown such fidelity over this period? Cultural changes can also be quite rapid, an important characteristic of this "second inheritance system" in nature, contrasting with the primary inheritance system of genetics, that is less nimble (Whiten 2017a). Humpback whale songs, for example, rise in objectively measurable complexity over time, then every few years the currently common song is superseded by a new "revolutionary" but simpler song, soon adopted by the whole population (Allen et al. 2018). Such new songs are known to be passed across the Pacific Ocean populations over a period of years, to be followed by further waves of new songs as these emerge.

A fourth major discovery has concerned adaptive preferences guiding animals' social learning. Veissière et al. focus much on adaptive attentional funnelling in the context of human culture, but this also occurs in animals, using a variety of cues that allow social learning to deviate from randomly copying others and achieve adaptive selectivity (Kendal et al. 2018; Price et al. 2017). These cues are diverse but span some that Veissière et al. discuss in the human case. They include conformity, expressed as a motivation to do what a majority of one's companions do (Aplin et al. 2015; Watson et al. 2018; Whiten 2019c for avian and primate examples, respectively). Other biases include copying high ranking and more successful individuals (Bono et al. 2018; Kendal et al. 2015). The latter study illustrates how complex combinations of preferences may arise: vervet monkeys' selectivity depends not only on the relative foraging payoffs they observe accruing to others, but also on the sex of the observer and the sex of the forager observed. Preferential learning from high rankers may sometimes be more akin to prestige effects in humans than often asserted; lemurs who were more successful at a novel foraging opportunity have been shown to receive more affiliative responses from companions later (Kulahci et al. 2018; Whiten 2018b).

The above can be no more than a selective sketch of recent discoveries about animal cultures. But, these phenomena surely enrich our understanding of the nature and evolution of culture: A unified theory of culture must surely accommodate them?

Integrating models of cognition and culture will require a bit more math

Matthew R. Zefferman[a] and Paul E. Smaldino[b]

[a]Department of Defense Analysis, Naval Postgraduate School, Monterey, CA 93943 and [b]Department of Cognitive and Information Sciences, University of California, Merced, Merced, CA 95343.

mrzeffer@nps.edu https://www.zefferman.com
psmaldino@ucmerced.edu http://smaldino.com/

doi:10.1017/S0140525X1900267X, e119

Abstract

We support the goal to integrate models of culture and cognition. However, we are not convinced that the free energy principle and Thinking Through Other Minds will be useful in achieving it. There are long traditions of modeling both cultural evolution and cognition. Demonstrating that FEP or TTOM can integrate these models will require a bit more math.

There is a decade-long tradition in which mathematical and computational models of social learning and cultural evolution demonstrate why social learning evolves; when it is useful; why it can be biased toward learning from the majority or from prestigious, successful, or similar individuals; and how learning biases help create the population-level dynamics of cultural change (Boyd & Richerson 1985; Henrich & McElreath 2003; McElreath & Henrich 2007). In turn, an understanding of these population-level dynamics has helped us understand and explain a wide range of phenomena including the origins of human cooperation (Richerson et al. 2016), civilization (Bowles & Gintis 2013), social identity (Smaldino 2019), hipsters (Smaldino & Epstein 2015), warfare (Zefferman & Mathew 2015), sex-biased tool use in dolphins (Zefferman 2016), and environmental sustainability (Waring et al. 2017). This theoretical framework already answers a number of questions Veissière et al. care about.

A criticism of this body of literature is that cultural evolutionary theory does not adequately consider cognition (Heyes 2018b). There are indeed questions about cultural evolution and social learning that a better understanding of cognition might help answer. In particular, cultural evolution likely shapes the sociocultural environments in which the cognitive machinery that facilitates social transmission develops. Learning more about the dynamical interaction between cognitive development and cultural evolution is an important and, until very recently, underappreciated research area (Contreras Kallens et al. 2018; Smaldino & Spivey 2019).

We find much to agree with in Veissière et al.'s qualitative description of how cognition and culture interact. However, much of this is well-worn territory. The main advance proposed in the target article is that free energy principle (FEP) and thinking through other minds (TTOM) integrate an impressively large number of theories and hypotheses of cognition, social learning, and cultural evolution. However, given the authors' extremely underspecified mathematical model, we remain unconvinced by this claim.

The mathematical model presented in the target article is too imprecise to be useful in its current form. The authors do not describe what the terms of the model represent in a cultural or cognitive system or show how the model does any scientific work. Indeed, the FEP is not actually a model of *anything*, but rather a paradigm for describing the behavior of systems that must themselves be modeled. It is fine to start theorizing with a general model (see, e.g., Frank [1995] on the Price equation), but useful models must eventually map onto relevant aspects of the world. In the target article's entire mathematical appendix, no parameter is described as representing any aspects of culture, brains, agents, ideas, cognitive processes, or any tangible or

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

intangible object related to the purported subject matter. Without this basic modeler's due diligence, we find the discussion of the potential contributions of FEP and TTOM to contribute little.

Take, for example, the "dark room" problem posed in the target article: "if agents aim to avoid unexpected encounters with their environment, we should expect minimally changing sensory environments like dark rooms and correspondingly monotonous sensations to be the most frequently (re)visited states of an organism." Of course, even a moment's consideration resolves this quandary, as strategic action can often dominate over passive non-action, and natural selection will favor it when it does. What does the FEP add to this non-dilemma? The authors state that:

> the FEP deals with the issue of novelty seeking behavior by formalizing action as being in the game of maximizing the epistemic value of action (or epistemic affordance) … [F]ree energy minimizing agents seek to sample the world in the most efficient way possible. Since the information gain (i.e., salience) is the amount of uncertainty resolved, it makes good sense for the agent to selectively sample regions of environment with high uncertainty, which will yield the most informative observations… In effect, agents will act to optimize the epistemic value or affordance of an action *before* acting on its pragmatic value, which is essentially its expected utility.

This is qualitatively appealing, but adds little value to the current state of scientific understanding. There is already a long history of modeling the trade-off between exploring an uncertain environment and "acting on its pragmatic value" in the social and biological sciences (e.g., Hills et al. 2015; Rendell et al. 2010; 2011). To demonstrate how a model *can* add value, contrast the musings above with a mathematical model studied by Perreault et al. (2012). In their model, a population of agents uses Bayesian learning to integrate environmental and social cues and eventually make decisions in an uncertain environment. They show that individuals optimally weigh social cues (relative to environmental cues) more heavily when the environment is more stable and when environmental cues are more uninformative. They also show that conformist-biased social learning (weighting common behaviors above chance) readily outperforms unbiased social learning across a broad range of conditions, especially when environments are novel or cultural transmission is error-prone. Most importantly, they provide an explicit model where agents make decisions and perform actions with measurable outcomes that potentially provide insight into the world. They have gone from the "first principles" of a Bayesian-learning process and evolutionary selection to the consequent changes in individual agents' cognition in response to socio-environmental forces.

What added value does the FEP give above and beyond the type of modeling done by Perreault et al.? Would an agent using the FEP or TTOM outperform their Bayesian-learning agents? Or is the Bayesian-learning agent a special case of an agent using FEP or TTOM? Should the overall model dynamics be analyzed using principles of the FEP, and if so, why?

Veissière et al. assert that agents will maximize the epistemic value of an action before its pragmatic value. However, a long history of models in many disciplines suggest fundamental trade-offs between these goals. As researchers interested in marrying cognition to cultural transmission, the determination of whether the FEP or TTOM are useful cannot be assessed unless their proponents can show how they do useful work in explaining the world.

Authors' Response

TTOM in action: Refining the variational approach to cognition and culture

Samuel P. L. Veissière[a,b,c], Axel Constant[b,d,e], Maxwell J. D. Ramstead[a,b,e], Karl J. Friston[e] and Laurence J. Kirmayer[a,b,c]

[a]Division of Social and Transcultural Psychiatry, Department of Psychiatry, McGill University, Montreal, Quebec, Canada H3A 1A1; [b]Culture, Mind, and Brain Program, McGill University, Montreal, Quebec, Canada H3A 1A1; [c]Department of Anthropology, McGill University, Montreal, Quebec, Canada H3A 2T7; [d]Charles Perkins Centre, The University of Sydney, Sydney, New South Wales, Australia 2006 and [e]Wellcome Centre for Human Neuroimaging, University College London, London, WC1N 3AR, UK.
samuel.veissiere@mcgill.ca axel.constant.pruvost@gmail.com
maxwell.ramstead@mcgill.ca k.friston@ucl.ac.uk laurence.kirmayer@mcgill.ca

doi:10.1017/S0140525X20000011, e120

Abstract

The target article "Thinking Through Other Minds" (TTOM) offered an account of the distinctively human capacity to acquire cultural knowledge, norms, and practices. To this end, we leveraged recent ideas from theoretical neurobiology to understand the human mind in social and cultural contexts. Our aim was both *synthetic* – building an integrative model adequate to account for key features of cultural learning and adaptation; and *prescriptive* – showing how the tools developed to explain brain dynamics can be applied to the emergence of social and cultural ecologies of mind. In this reply to commentators, we address key issues, including: (1) refining the concept of culture to show how TTOM and the free-energy principle (FEP) can capture essential elements of human adaptation and functioning; (2) addressing cognition as an embodied, enactive, affective process involving cultural affordances; (3) clarifying the significance of the FEP formalism related to entropy minimization, Bayesian inference, Markov blankets, and enactivist views; (4) developing empirical tests and applications of the TTOM model; (5) incorporating cultural diversity and context at the level of intracultural variation, individual differences, and the transition to digital niches; and (6) considering some implications for psychiatry. The commentators' critiques and suggestions point to useful refinements and applications of the model. In ongoing collaborations, we are exploring how to augment the theory with affective valence, take into account individual differences and historicity, and apply the model to specific domains including epistemic bias.

We are grateful to the commentators for their critiques, challenges, and elaborations of our model of human cognition, action, and cultural learning called *Thinking Through Other Minds* (TTOM). Several commentators provided arguments and examples that address points raised by others, which suggests – to our great satisfaction – that the TTOM model is useful, not only as a step toward an integrative theory of enculturation, but

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

also as a framework for interdisciplinary collaboration and knowledge exchange.

The target article offered an account of the distinctively human capacity to acquire – and think through – cultural knowledge, norms, and practices. To this end, we leveraged recent ideas from theoretical neurobiology to understand the human mind in social and cultural contexts. Our aim was both *synthetic* – building an integrative model adequate to account for key features of cultural learning and adaptation; and *prescriptive* – showing how the tools developed to explain brain dynamics can be applied to the emergence of social and cultural ecologies of mind.

Our commentators raised important issues regarding definitions, concepts, and methods, and called for further development of the model. Given space constraints, we cannot respond to every point in each commentary. To address the key issues, we have organized this response thematically into six sections. In what follows, we discuss: (1) refining the concept of culture to show how TTOM and the free-energy principle (FEP) can capture essential elements of human adaptation and functioning; (2) addressing cognition as an embodied, enactive, affective process involving cultural affordances; (3) clarifying the significance of the formalisms related to the FEP and active inference, including (3.1) entropy minimization, (3.2) Bayesian inference, (3.3) Markov blankets, and (3.4) enactivist views of cognition; (4) developing empirical tests and applications of the model; (5) incorporating cultural diversity and context at the levels of (5.1) intra-cultural differences, (5.2) individual differences, and (5.3) the transition to a digital niche; and (6) potential applications to psychiatry.

Our aim is to clarify a picture of human cognition – not simply in terms of a generic "first principle" paradigm – but one designed to account for specific kinds of adaptation that are reflected in the patterns of social and cultural organization that depend on the body, emotions, interpersonal perception, and epistemic biases.

R1. The domain of culture

Several commentators noted the need to clarify the notion of culture that underwrites the TTOM model. **Whiten**'s insightful commentary suggests that accounts of culture should explain the cultural behaviors of non-human animals, which in recent decades have been documented extensively. In particular, Whiten notes features of animal culture that a unified theory of culture will have to account for, including: *phylogenetic reach*, that is, the ubiquity of transgenerational transmission of learning behaviors in many species; *intraspecies pervasiveness*, that is, the inclusion of many forms of behavior within culturally learned repertoires, which may exhibit diversity within a specific or community; *stability* and *fidelity* over time; and finally, *adaptive preferences* for certain modes or types of social learning.

Whiten suggests that, from the perspective of animal cultures, some of the features of culture we discuss are optional, for example, inter-group differentiation. However, it is precisely cultural diversity, borrowing, exchange, hybridization, and competition (i.e., differences that arise primarily through cultural learning) that allow the processes of cultural change and elaboration distinct from the rudimentary replication of limited behavioral repertoires. The 4300-year-long "fidelity" in nut-cracking technologies mentioned by Whiten, along with the minor (and similarly stable) regional variations in ape-foraging strategies, reflects local affordances rediscovered in each generation, apparently without intergenerational cultural transmission (Moore 2013). On this view, such "fidelity" may actually reflect a rigid dependence on what we have called natural affordances (Ramstead et al. 2016), without the cumulative intergenerational cultural elaboration that Tomasello has called the "ratchet effect," which appears to be unique to humans (Tennie et al. 2009).

The potential phylogenetic reach of basic forms of culture raised by **Whiten** may yield important insights into forms of learning we share with other animals. Our account, however, focuses on the phylogeny of the enhanced theory of mind modalities and cognition–culture iterative loops that appear unique to humans. Central to our account is the cognitive package that we call *Thinking Through Other Minds* (TTOM). Under the FEP, we construe this as package as comprising a set of abilities and a domain of statistical regularities that are coupled: abilities enabling us to tune to the minds of others and to navigate environmental uncertainty. The TTOM model explains this coupling by appealing to evolved and learned priors about conspecifics and their mental states.

Among hominids, current models for the evolution of TTOM emphasize cooperative breeding; a strategy that likely evolved in the Homo Erectus lineage circa 2 million years ago (MYA) (Hrdy 2011) that selects for individuals who are skilled at understanding others' needs, giving care, and eliciting care. Across species, the evolution of cooperative breeding likely follows different timescales and different pathways to similar traits. New World monkeys such as marmosets, for example, who share a last common ancestor (LCA) with humans 35 MYA, are cooperative breeders with better mind-reading and prosocial abilities than the non-cooperative breeding great apes (LCA with humans 5 MYA), and have as such been recognized as better models than chimpanzees for understanding human evolution (Miller et al. 2016). Humans, in turn, have refined skills for detecting the epistemic cues and their precision allowing for more complex forms of cultural transmission.

We use the computational construct of *precision of epistemic cues* to account for the relative stability, fidelity, *and scaffolded elaboration* of cultural forms of life over time. To account for *preference guiding behavior* based on the formal construct of culture acquisition proposed by TTOM, one might appeal to gene-culture coevolutionary explanations. This is what we suggested with the example of prestige biases and regimes of attention in humans. Prestige biases are an example of external component of the regime of attention – passed on as high-precision epistemic cues (scaffolded on, then divorced from physical dominance hierarchies) – whose effect is enabled by genetically inherited predispositions to social learning functioning as an internal component of regimes of attention (see Figure 4 of the target article).

In contrast to **Whiten**'s appeal to animal culture, **Vickhoff** starts from the example of the arts, forms of human culture that involve elaborate systems of shared customs, codes, and scripts. Vickhoff considers music as an instance of culture, because it is governed by a system of shared expectations, and conforms to a "style" – a type of convention with esthetic value. To this, we would add a third salient characteristic of music, the arts and, indeed, of culture more generally: improvisation and creative invention (Torrance & Schumann 2019). All of these aspects are related to TTOM. Vickhoff suggests that, as an enculturated agent, the listener can access the composer's mind as expressed through culturally scripted features of music. Hence, in Vickhoff's words, we not only access culture through other minds but also engage with other minds through culture. Although the TTOM model accounts for the idea of "set of

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

expectations," it may be less intuitive to apply the framework to thinking about cultural genres or styles. This is so because such uses speak to notions of esthetic value and of creativity and novelty-seeking, which do not have obvious interpretations under TTOM and the FEP.

Creativity is a kind of exploration of new possibilities for perception and action. Similar to other exploratory activities, it is driven by the human propensity for novelty-seeking or curiosity. Several authors have provided accounts of culture and curiosity in terms of the epistemic value of exploring a niche or a larger adaptive landscape to identify current or future possibilities (Friston et al. 2017a; 2017b; Moulin & Souchay 2015; Schmidhuber 2006; Schwartenbeck et al. 2013). On this view, the FEP applies to a second-order process of optimization on longer time scales and, in some instances, across many alternate niches (Bengio 2014). A similar argument can be used for the value of creativity to generate an expanded adaptive repertoire of actions and responses. Improvisation, invention, and innovation are basic to human cultures, not only in the domains of esthetic experimentation that have come to be designated as "the arts" in contemporary societies, but equally in the most quotidian activities.

The relationship between culture and creativity in the framework of TTOM needs further development, to be sure, and **Vickhoff** gives some clues about how to address this. He mentions recent work on synchrony in human action, which is beginning to reveal the mechanisms of micro-coordination of action in real time and its consequences for feelings of emotional attunement and social affiliation (Kiverstein et al. 2019; Parkinson 2019; Tschacher & Haken 2007; Van de Cruys & Wagemans 2011; Van de Cruys et al. 2017). Vickhoff notes, as well, that the reduction of uncertainty can be pleasurable (Vuust et al. 2018). The esthetic pleasure of music may arise in part from a balance between the setting up expectations (through rhythm and repetition of melodic and harmonic structures) and violating expectations with novelty, and then re-establishing order by further repetition or thematic resolution (Huron 2006; Vuust et al. 2018). This play with the tensions of confirmation of expectation and surprise then constitutes a microcosm of the larger adaptive tasks of dealing with an often-unpredictable world (Friston & Friston 2013).

A similar challenge to the view that all human action interactions are motivated by the minimization of uncertainty is spelled out in **Mirski, Bickhard, Eck, & Gut**'s (**Mirski et al.**) commentary. For these authors, a free-energy account of culture fails to explain the emergence of new or competing normativities, and the varied ways in which many human actions appear to possess no fitness-enhancing or uncertainty-reducing function. They mention "people deciding to die or suffer for some highly abstract cause" as one such candidate exception to our model. First, we should note that the pursuit of "suffering" is typically patterned and prescribed socially, and seen in many cultural domains from rites of initiation, to religion and athletics (Atkinson 2008; Coakley & Shelemay 2007; Gaines & Farmer 1986). We see this as paradigmatic of, rather than an exception to the search for high-precision cultural modeling under the FEP. The most widely accepted account of "righteous violence" of dying for an abstract cause, in turn, uses Terror Management Theory to explain how perceived threats to one's group and belief systems coming from external agents can motivate altruistic death as a group fitness-enhancing strategy (Pyszczynski et al. 2009). In the framework of active inference, altruistic self-sacrifice is motivated by an effort to reduce a perceived increase in the uncertainty of the social world. Cognitive anthropologist Scott Atran's work on suicide terrorism has yielded further clues on how some people come to die for "abstract ideas." For Atran (2003), strong motives to defend an imagined community are not simply abstract: they are extrapolated from small-scale group bonding, where strong ties of solidarity – installed by military training – produce the motivation or willingness to die for one's friends, seen as an extension of a broader community. These examples can help us understand how new and competing normativities are also competitions for generative models of the world. Conflict over meaning and how the world ought to be invariably occurs within and between groups. We understand social change as occurring under these dynamics of optimizing world models through intra-group and inter-group competitions. We think that the FEP offers a way to account for both social stabilization and social change, and indeed for human historicity itself.

R2. Culture as embodied affective engagement with affordances

In describing human behavior as culturally patterned via regimes of attention that modulate salience and guide action, we aimed to ground our account in so-called 4EA approaches that emphasize the embodied, embedded, extended, enactive, and affective nature of cognition. As pointed out by several commentators, our attention-and-expectation-centered, information-heavy account left out important details on the affective, sensorimotor, physiological, and phenomenological dimensions of this puzzle.

Baggs & Chemero express the concern that our choice of terminology leads us to pursue an approach that is disembodied and inferential rather than embodied and enactive. They note that Vygotsky long ago proposed an approach to cultural learning as developmental engagements between social actors (initially child and parent) that is enabled by language (Vygotsky 1980). This developmental trajectory has been well documented (Tomasello et al. 1993). The mother's discursive framing of the child's activity becomes the child's self-talk, which in turn becomes inner speech. The narrative self is born in this move from internalized dialog to monolog. Beyond this developmental picture, Baggs & Chemero insist that "language-involving cognition operates according to a different set of norms, and is not merely a more elaborated form of adaptive fitness."

We agree that the ability to use language is a game changer for cultural learning. It allows humans to explore imaginary fitness landscapes, install higher-order priors without the need for individual lived experience, expand action repertoires by following recipes or instructions, and explicitly name, frame, debate, and deliberate about social norms. Most powerfully, language allows recursion and self-reflexivity. In theorizing how most human priors become saturated with other minds, and in this sense divorced from *direct* interaction with the world, we agree with **Baggs & Chemero** that this process can be cashed out under a Vygotskyian model of scaffolding via social learning and the education of attention. We have made this argument elsewhere (Ramstead et al. 2016).

We strongly disagree, however, with **Baggs & Chemero**'s claim that the affordances construct cannot be used to explain socially and culturally appropriate behavior. We see no reason to think the engagement of humans with their social and cultural worlds cannot be cast in terms of affordances. Indeed, Gibson himself used the concept of affordance to describe human social and cultural behavior (1986, p. 119ff). The difficulty in accepting that

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

dynamics of enculturation can be affordance-like, as we have argued elsewhere (Ramstead et al. 2016), stems from the conservative definition of "affordance" adopted by many cognitive scientists, which applies the concept only to possibilities for action that can be "picked up" without instruction and do not require social learning to recognize and use (e.g., Moore 2013).

Drawing in part from Chemero's (2011) work, we have defined a cultural affordance as a relationship between the embodied skills of an encultured agent and relevant aspects of the cultural environment (Ramstead et al. 2016). In our model, the relevance of an affordance is cashed out in terms of culturally shared priors about salience; and the ability to engage with cultural affordances corresponds to policy selection in active inference. In the target article, we provide an embodied dynamical account of how such engagement with a cultural niche is possible. That account explains what goes on "under the hood" (and indeed, "around the hood" in the environment) in scaffolded learning processes that stretch from imitative learning to explicit rule following and deliberation. It is with this view in mind that we have described culture as involving a set of *deontic* affordances. From this perspective, the cultural world not only *solicits* certain modes of attention and action, but also entails the *obligation* to respond to the world and other people in constrained ways. These obligations are felt in the body, expressed in habitual stances and actions, as well as (sometimes) in elaborated moral language and deliberation.

Clement & Dukes offer the important specification that culturally patterned practices do not simply direct attention in a value-free fashion, but entail the structuring of *valence*, as well. For these authors, "it is the valence and the intensity of others' emotional expressions in particular that can be used to detect what is expected from each member" (para. 4). We could not agree more with this description of social learning as fundamentally *affective*, and culture as deeply *axiological* – less about "what is," as Clement and Dukes put it, and more about "what matters and what is meaningful." We understand this process as operating beyond explicitly normative situations, and extending to the affective qualities of the cultural world. For example, inferring whether someone's style of dress signals high or low social status, or assessing the "intrinsic" desirability of a prospective purchase, involves picking up *on what relevant others expect the world to afford* – simply learning fixed associations to social cues without this component of TTOM might lead to frequent errors as we move across contexts or niches. In other words, we are able to pick up the multidimensional *valences* assigned to specific features of the world by local regimes of attention. We can learn these valences as embodied dispositions to respond to cultural affordances, without explicit awareness of the relevant norms.

In their commentary, **Smith & Lane** show how the TTOM model could be expanded readily to make an explicit place for emotional processing and interoceptive embodied inference. In their approach, emotional processes are cast as emotional policy selection and emotional state inference (Smith et al. 2019a). This extension of the formalism rests on an additional layer of parametric depth that biases action toward inference by the agent of its own affective states. We would add that interoceptive inferences are also modeled socially against the affective states of others – indeed, they are sometimes taught more or less explicitly in the way caregivers react to, and thereby teach us to react to and assign predictive meaning to our own internal states and social positions (Gendron & Barrett 2018; Hoemann et al. 2019). This hierarchical and interpersonal structure of an active inference agent fits well with TTOM, which extends all the way from explicit high-level inferences involving emotional concepts to low-level automatic emotional response generation (Seth & Friston 2016).

This model of emotion, when coupled with the concept of *deontic policy selection* (Constant et al. 2019b) discussed in the concluding section of the target article, can get us closer to what **Buskell** expects of TTOM: a view of characteristically human phenomena (such as radical organizational change), which requires an understanding of affect and normative behavior. We have begun some work in this direction (Constant et al. 2019b) showing how concept of *deontic policy selection* can help account for phenomena such as social conformity (Asch 1955; Toelch & Dolan 2015). In this model, we provide a formal description of how the acquisition of regimes of attention allows enculturated agents to zero in on what appears to be the most socially relevant response to a situation. This policy selection may have its own affective consequences which can influence subsequent action. This looping process follows directly from the TTOM model augmented with the concept of deontic action selection and the implementation of affect under the FEP by Smith et al. (2019a).

Van de Cruys & Heylighen provide a compelling example of affective social phenomena that is readily implementable with a model of affect coupled to TTOM. They propose that TTOM can help account for social phenomena that evoke, maintain, and are amplified by negative affect (e.g., dogmatism or radicalism). For instance, "guilt [*can*] also become a hidden cause efficiently explaining away someone's behavior in the eyes of others belonging to the culture." Learning the mapping between one's actions, others' suffering and feelings of guilt attaches an affective valence to policies that can guide behavior. Moral emotions such as guilt, then, contribute to social coordination in particular cultural contexts and indeed, as others have suggested, to the unique forms of pro-sociality found in human forms of life (Tomasello 2009).

The commentary by **Allen, Legrand, Correa, & Fardo** suggests that our account is overly focused on cognitive and brain-bound prior beliefs, and does not sufficiently address embodied forms of inference that are ongoing at other scales. Indeed, work by Allen's group demonstrates that visceromotor processes and interoception can be usefully cast as forms of active inference (Allen et al. 2019). We agree that social and cultural processes such as those described in the TTOM model may be built on a basis of embodied, interoceptive priors some of which have a long phylogenetic history. We would simply add that cultural processes also allow priors to be installed from the top down, through linguistic coding and mimetic processes, so that in any actual instance, culturally learned behavior is likely to be an outcome of both interoceptive and exteroceptive influences and both embodied and linguistic practices (Di Paolo et al. 2018). This puts cultural history and learning on an equal footing with evolutionary and co-evolutionary influences.

The ongoing social dynamics of affective influence are foregrounded in several commentaries. As **Mouras** reminds us in his commentary, shared affective, motivational, and sensorimotor processes can be observed and experimentally manipulated directly in social interactions. Mouras's discussion of experiments on postural correlates of in-group bias in empathy for pain offers many useful pointers on how to operationalize models of shared agency in TTOM. Mouras and colleagues have made elegant use of the Minimal Group Paradigm, in which in-groups and out-

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

groups are arbitrarily constituted in a lab environment (Tajfel et al. 1971), to demonstrate the quick, flexible, yet precise ways in which embodied and affective joint models can emerge as long as an assumption of shared goals is present. The experiments described by Mouras show that empathy for others' pain can occur quickly once in-group conditions obtain, and that merely imagining fellow group members in pain affects sensorimotor processes with effects such as automatic (nonconscious) stiffening of posture. These findings echo recent studies on physiological correlates of group bonding that evince shared responses in both participants and spectators during high-arousal conditions of group synchrony, such as marching in a stadium (Jackson et al. 2018) and ritual firewalking (Konvalinka et al. 2011). Taken together, these studies point to an emerging socio-physiology that reveals some of the embodied interactions that undergird TTOM.

Further points raised by commentators help us describe social interactions as shaped by relational dynamics that include developmental histories and processes of attachment and affiliation. **Strand**'s commentary, on the importance of attachment patterns with supportive others as potential modulators of cultural learning, raises important issues, including the claims that: (1) reward learning mechanisms are central to the patterning of culture via Bayesian inference; (2) neural pathways subserving these mechanisms are likely evolutionarily old, and (3) attachment styles are an important locus of cultural differences. In support of this view, Strand cites recent studies showing that collectivistic cultures exhibit a higher prevalence of insecure-anxious individuals, compared to higher rates of insecure-avoidants in individualistic cultures (Yamagishi & Hashimoto 2016). We agree that attachment and reward mechanisms are central in the phylogeny and ontogeny of culture. On this view, cultural evolution has favored modes of interaction that fulfill and leverage our evolved need for rewarding attachments, enabling the construction of adaptive social ties in each generation (Boyer & Liénard 2006; 2008). We can think of the patterning of culture in terms of kinds of affiliative relationships, communal activities, rituals, and ceremonies that serve a general function of social uncertainty minimization, and contribute to specific, adaptive cooperative action, whereas offering a protective buffer against the risks of loneliness and rejection.

Strand's comments are also helpful to understand another strange loop in cultural evolution: attachments are central to the patterning of culture, yet are themselves culturally patterned. In Yamagishi and Hashimoto's (2016) niche construction model of cross-cultural differences in attachment, an avoidant style is understood as adaptive in individualistic cultures (in which help-seeking is socially costly), whereas anxious attachments confer social adaptiveness in cultures that require high levels of cooperativity and social deference (where individualism is socially costly). Of note here, both "individualism" and "collectivism" are collectively patterned via social norms and deontic cultural affordances that demand different levels of autonomy in specific contexts.

Pezzulo, Barca, Maisto, & Donnarumma's (**Pezzulo et al.**) insightful commentary on social epistemic action further assists us in specifying that, in addition to being rewarding for its own sake, human joint action often entails conveying cues for the sake of others. Pezzulo et al. explain that unlike epistemic actions (aimed at updating one's contextual beliefs about relevant features of the world) and pragmatic actions (acting on the world after resolving contextual uncertainty), social epistemic actions serve the function of directing others' attention to patterns and regularities that we want them to infer. This is precisely the phenomenon we have referred to as *regimes of attention* (Ramstead et al. 2016). The authors point out that although such actions can be verbal, humans routinely employ a wealth of kinetic and sensorimotor cues (what they cleverly call "motionese") to convey the internal states and intentions they want others to infer, or to "signal more directly what and where salient information is." Pezzulo et al. rightly emphasize the importance of social epistemic actions on "faster timescales" (e.g., during teaching, learning, and real-time communication) but we also find the notion useful for an account of the evolution of human sociality on longer co-evolutionary and developmental timescales. The existence of species-wide ostensive or indexical cues, such as pointing (to direct others' attention), or holding out a hand (to signal helping behavior) offers strong evidence for evolutionarily old strategies for communicating universally legible intentions and emotions to others. The presence of laughter and vocal crying in apes, humans, and pre-verbal human infants offers further evidence for phylogenetically old and involuntary (or automatic) drives for social epistemic actions. Indeed, both actions occur under weak or absent voluntary control, and are best described as forms of honest signaling that reveal true internal states (Provine 2017). For example, an experiment using short audio clips found that hearing joint laughter was sufficient for participants from 24 diverse language groups to differentiate pairs of friends from strangers (Bryant et al. 2016).

This view of cultural transmission as entailing multiple forms of learning falling on a spectrum from automatic "picking up" of information to explicit teaching and deliberation assists us in taking up **Gweon**'s helpful pointers for "a more complete picture of social learning." Gweon points out, building on Cisbra and Gergely's (2009) Natural Pedagogy paradigm, that children are intrinsically motivated to identify high-utility, high-quality information from reliable sources, and are subsequently motivated to teach that information to their peers. We agree with Gweon that this view of children as active agents of both Bayesian inference and attention-directing communication is key to understanding cultural transmission. In this line, we welcome **Michael & de Bruin**'s invitation to consider the importance of *mind-shaping* mechanisms and to clarify how the TTOM model of enculturation accounts for a full implicit-to-explicit inference spectrum. We find the postulated existence of evolutionarily old, developmentally early behavior-influencing mind shaping mechanisms (De Bruin and Strijbos 2020; Zawidzki 2013) fully compatible with the account we have outlined above. Our account of social cognition recognizes multiple levels and instances of inference from the "quick and embodied" to the effortful and deliberative. In doing so, we join previous "multi-system" efforts by Christensen and Michael (2016) intent on resolving tensions between theory–theory, simulation, and embodied accounts of mind-shaping and mindreading.

R3. Formal clarifications: What the FEP adds to our understanding of cultural learning

The TTOM model gains precision and explanatory power by leveraging the active inference framework. Unfortunately, some of the key features of active inference have been misunderstood in the recent literature, as well as in some of the commentaries. Clarifying these misunderstandings is crucial to appreciate how TTOM can be applied to model cultural learning.

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

R3.1. Optimization and "entropy minimization"

One frequent confusion concerns the construct of entropy and its relationship with variational free energy in the active inference framework. The commentary by **Fortier-Davy**, for example, characterizes our view in the paper as the claim that "humans tend to minimize entropy" or, still more strongly, that "humans *seek* to minimize entropy" [emphasis added]. We note that nowhere in the text do we actually make either of these claims. However, statements such as Fortier-Davy's, although inaccurate, are often presented in the literature as a criticism of active inference. These critiques treat active inference as if it were equivalent to entropy minimization. The accompanying critique, then, is that, because there is a large body of evidence from psychology and social science that human behaviors do not always (or usually) act to minimize entropy, active inference must not apply globally to human cognition.

Certainly, entropy reduction and FEP minimization are related, but the link is more nuanced than simple equivalence. Variational free energy is an upper bound on surprisal; and assuming that the system we are considering is measurable (i.e., is equipped with a nonequilibrium steady state), the time average of surprise will converge toward entropy. This means that it is true that a system that is able to track and minimize free energy from moment to moment will also place an upper bound on entropy on average and over time (equivalently, place a lower bound on model evidence or marginal likelihood). At this first level of analysis, which considers only moment-to-moment dynamics, **Fortier-Davy**'s claim about the relationship between the FEP and entropy seems to hold. However, there is more to this story.

The relationship between entropy and variational free energy is enriched when considering the enactive aspect of self-evidencing, that is, by the consideration of planning sequences of actions. Contemporary formulations of active inference work not merely with variational free energy but with *expected free energy*, which is the (average) expectation value of free energy for a given policy (Friston et al. 2017a; 2017b). If we consider a probability distribution over the space of available policies, minimizing expected free energy actually means that we will end up *maximizing* the entropy of that distribution: in effect, a maximally entropic or flat distribution over the space of policies means that the agent is "keeping its options open." Thus, it follows from our formal account that humans (and all creatures) pursue actions that increase the entropy or spread of their available action repertoire. This idea is reflected in recent work on the epistemic value of actions (Parr & Friston 2017a; Pezzulo et al. 2016).

Further, the idea that living creatures minimize the entropy of their states is not contradictory to the idea that humans (and other animals) will seek out novel, high entropy stimuli. Indeed, this is just what active inference agents do when they select the policy that minimizes expected free energy. This, again, follows from the formal construction of expected free energy (Parr & Friston 2017a; Pezzulo et al. 2016). One way to express the expected free energy of a policy is as risk minus novelty (the same formula used in economics); and another is as pragmatic value plus epistemic value. Both of these formulations account for the novelty-seeking behavior that is typical of agents such as humans.

An agent that acts to minimize expected free energy will first explore the world and resolve uncertainty by *seeking out* high-entropy (and therefore, information rich) stimuli, before acting on the pragmatic value of a policy. Indeed, information is defined as the amount of uncertainty resolved by making an observation. As such, the most informative observation is the one that resolves the most amount of uncertainty, which means that agents may seek out the *most entropic* stimuli in order to derive the greatest amount of information. Finally, recent active inference formulations parameterize expected free energy (Hesp et al. 2019); in this setting, agents have beliefs or preferences about how surprised they should typically be. These qualifications about the relationship between the entropy and active inference naturally account for the findings discussed by **Fortier-Davy** that humans tend to prefer medium-entropy rather than low-entropy stimuli.

A broader concern about the utility of TTOM is raised by **Overgaard**, who finds the model ambiguous insofar as it seems to conflate a simple description of the ethnographic reality of what people do (namely, the obvious truth that thought and action occur cooperatively with other people) with a more specific computational model of specific kinds of learning (about which he is dubious or noncommittal). We think there is no real ambiguity here, but a simple progression from description to explanation. TTOM begins with the descriptive facts that Overgaard acknowledges but moves on to account for what is going on in these interactions – and the phenomena of cultural learning. To do this, TTOM incorporates a model of implicit learning that involves our brains and bodies interacting with one another and the world (e.g., as our neural networks are tuned to predict our social and designed environment through rolling cycles of sensorimotor action and perception). The subtlety is that TTOM is a theoretical model of cultural learning that describes cultural learning and cognition as following particular constraints (formalized mathematically in the target article). Yet at the same time, TTOM is something that people "do" insofar as those equations capture the dynamics of neuronal and social ecologies that constitute the sort of enculturated beings that we are. TTOM is at once a name *for* and a map *of* what we do.

R3.2. Building on Bayes: Cultural learning, optimization, and adaptation

Colombo gives us the opportunity to clarify another, often ill understood fact about the FEP and active inference. Colombo poses the following problem: if "utility (or adaptive value) of an outcome is equivalent to its probability [then] complying with social norms [which entails maximising the probability of certain outcomes] always has adaptive value." Obviously, we agree with Colombo that this is not the case. Sociocultural norms and dynamics do not need to promote individual fitness or cultural progress. Indeed, norms and culturally prescribed goals can also lead to death or societal collapse (Diamond 2010). If cultural progress occurs, it is most likely because cultural norms promoted changes that yielded population-level advantages in response to environmental pressures over long stretches of time. Conformity is useful to social groups and, indeed, the tendency to conform varies both among individuals within a group and, on average, between groups. But, the norms to which individuals learn to conform need not benefit everyone individually, and could even involve harm for some individuals or subgroups, yet persist because they enhance population-level fitness. The discrepancies, conflicts, and trade-offs between these utilities at different levels and timescales can account for the persistence of norms that

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

are not beneficial to the individual or to some segment of a social group.

Colombo's main concern, however, is with another kind of conflation between descriptive expectations, normative expectations, and preferences for conformity to social norms. The FEP is a normative framework in the sense that it states what *all organisms must have done, given that they have existed for some duration* (i.e., minimized their free energy), in the same spirit that natural selection will tend to maximize (expected) fitness at the population level. The FEP is normative in the sense that it tells us about the conditions that must have been met by environment-sensitive systems, if they have maintained themselves at the non-equilibrium steady state. The FEP is not normative in the sense that it tells us what all organisms *ought to do*, or what will *necessarily happen to them*. Thus, it leaves plenty of space for thinking about things such as branching processes, spandrels, maladaptive traits, constraints, trade-offs, and mismatches – all the stuff we need to understand the place of adaptive value in the contingent history and development (ontogenetic or evolutionary) of the system of interest. The observation that particular suboptimal or maladaptive norms obtain is not a problem but an observation that forces us to explore the impact of initial and boundary conditions and historical trajectories on the constitution of humans and their cultural niches.

A related objection is raised by **Zefferman & Smaldino**, who point out that our model does not sufficiently account for the learning and cognitive biases that drive cultural transmission. According to Zefferman and Smaldino, the FEP offers little explanatory power when compared to these learning rules. We agree that cultural evolution likely favored multiple rules reflecting diverse mechanisms, optimized over time, or recruited for specific kinds of cultural learning and adaptive tasks. In the target article, we made a case for prestige, expertise, and in-group attentional biases as obvious candidates for high-precision learning rules driving cultural evolution. Rules that may be maladaptive (in that they are mismatched with the environment, or confer fitness on certain traits or groups at the expense of other rules that are more adaptive) or that may not operate under active inference do pose a challenge to our model. For example, Perreault, Moya, and Boyd argue that "many adaptive problems are difficult because the environment does not provide clear cues to the best behavior. What is the best design for a bow? What causes malaria? It is not clear what decision rule will be favored by selection when the environmental cue does not allow accurate inference" (Perreault et al. 2012). Our work so far has spelled a general model of the *architecture* of cultural learning systems as whole (see Figure 4 of the target article) at the expense of an exhaustive description of the specific rules that emerge in the system. We welcome the challenge to enhance our model with a better account of invariant rules, which are likely to be domain specific.

The commentary by **Mirski et al.** also addresses the need to specify the mechanisms of cultural learning but misses another key point about the general TTOM framework – by misunderstanding how new local "rules" can emerge in the patterning of culture. They suggest that approaches to cognition based on active inference are incorrect because they cannot explain how non-phylogenetic priors, such as those learned through experience, could come to guide the behavior of agents. This ignores the fact that our model accommodates the learning of (empirical) priors at various timescales, and most notably through immersion in coordinated action and experience – a view of learning in context often called empirical Bayes (Friston et al. 2016; Kass & Steffey 1989; MacKay 1992). Mirski et al. merely stipulate that the FEP cannot account for normative phenomena and do not engage the framework and formalism that we provide precisely for online learning of such patterns.

R3.3. The Markov blankets of culture

In conjunction with the FEP, TTOM relies on the notion of Markov blankets as ways of demarcating cognitive systems and their environments. Thomas **Parr** adds helpful refinements to the cultural affordance model we deploy in our paper by focusing on the ways in which humans (and those studying them) draw Markov blankets. Parr brings to the fore three perspectives on the ways in which we, as scientists, choose to draw Markov blankets around systems of interest to build a computational model. First, there is the selection of the blanketed system itself (the set of internal states and their blanket), which licenses an inferential interpretation of the system dynamics. This interpretation chooses a set of systemic or internal states that infer external, non-systemic states through their vicarious coupling via the blanket states. Second, once such a blanketed system is carved out, one needs to select which blanket states will drive the activity of sensory states; this is the question of attention. Third, the system must also select which parts of the environment are most relevant to its policy selection; this is the issue of salience, novelty, and the more general issue of selective sampling.

An important empirical question to understand the behavior of any organism, on **Parr**'s view, is to identify the relevant "influences from the outside in" that modulate "implicit choices made by internal states of a Markov blanket as to which blanket states most influence their dynamics." A central argument in our account is that for *Homo sapiens*, "the outside" is culturally encoded with group-level, action-generating *expectations about the world* that predict differentiated (hence patterned) experiences of, and action in the world. Through developmental scaffolding via shared modulations of attention ("enculturation"), the most relevant buffer of statistical regularities for humans gradually becomes *other people's expectations about the world and how to function in it optimally given a set of social norms and environmental constraints* (Veissière 2018). It is this buffering dynamics of a shared Markov blanket (Poirier et al. 2019) between the human groups and the world that we have called TTOM.

R3.4. Addressing enactivist objections

Several objections to the active inference approach are raised from enactivist perspectives echoing those of **Baggs & Chemero**. In the case of the commentaries by **Hutto**, and **Kiverstein & Rietveld**, these objections are based on a very narrow reading of computation, information, and representation. Active inference is a theory of belief-guided action and related information flows. Information, in this context, is not a cognitivist or Cartesian construct that would render our theory incompatible with models focused on dynamic coping and attunement. Rather, variational free energy is an information theoretic measure that *emerges from* dynamic interactions between a Markov-blanketed system (with its own set of beliefs or generative model) and its embedding environment – which as we have argued, crucially includes culture via ecologies of other minds. We have known since the pioneering work of Pattee (1977) and Kelso (1994) that informational and dynamical descriptions are complementary, not contradictory. Variational free energy quantifies precisely the extent

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

to which a system becomes a model of its environment through their reciprocal dynamic interactions. Second, and most importantly, where we decide to draw a Markov blanket has consequences for our inferential interpretation of its dynamics. For human systems, this means distinguishing between two cases: (1) those in which two agents (as blanketed systems) are inferring each other; and (2) those in which a higher-order system, which has two agents as its internal states or component parts, infers aspects of the external world. We believe TTOM applies to both cases; in the first case, it provides a new take on Theory of Mind abilities of agents (i.e., their capacity to infer each others' mental states); in the second, it provides a new account of cultural dynamics as a form of group inference (Clark 2017b; Kirchhoff et al. 2018). The flexibility of this formalism licenses our use of it as a model of human dynamic interactions across spatial and temporal scales.

Hutto's commentary suggests that the TTOM model is making a "spectatorial assumption," in that it suggests that humans do not have direct access to another's mind and that they must employ inferential abilities to access the mental states of others. We accept this characterization of our view. It seems obvious to us. Even if one were to focus on the phenomenology of interpersonal engagement, it also seems like a plain fact about experience that human agents never have direct, unmediated access to the minds of other agents. Humans have a rich phenomenological experience, on the basis of which we must infer the mental states of others (i.e., facets of their own experience). This follows quite simply from the basic setup of the problem that all living things face: they only have access to the environment through their sensory states (including interoception), and must reconstruct the causes of their sensations – other persons included – in an inferential matter (Hohwy 2016). This, however, does not mean that human agents are cut-off from their social world. Nor does it mean that all knowledge of others involves explicit inference. Indeed, as we have argued elsewhere (Ramstead et al. 2018), the multiscale active inference formulation entails that – to the extent that they form a higher-order social ensemble – systems that are segregated at one scale are integrated into a higher-order dynamics. The *regulated coupling* of internal and external states via the dynamics of blanket states implies a *semantics* and indeed semantic content in the strongest sense (Constant et al. 2019a). This inherently representational semantics of active inference is frequently missed by enactivist interpretations.

R4. Applications and empirical tests of the model

We welcome comments from **Brown, Brusse, Huebner, & Pain** (**Brown et al.**), who speak from the vantage point of cognitive archeology. They agree that unifying models can help "dissolve disputes by bringing rival positions under a single theoretical framework," but remain unconvinced that the FEP can offer such a framework. In particular, Brown et al. express their doubt that our model can generate testable predictions. This concern is shared by **Dołęga, Schlicht, & Dennett** (**Dołęga et al.**), who argue that our account does not sufficiently distinguish between the levels of explanation. They suggest that our model is best understood as applying to Marr's computational level of explanation (i.e., characterizing the information processing task). Of course, this specification has consequences for theory building. For example, by showing that one can cast both mind-reading and niche construction as forms of active inference, we show that there is no need to choose between competing alternatives. Although our account shows how internalist and externalist accounts can both be justified in some contexts, Dolega et al. claim that it does not provide us with a way of deciding between the two options for any given phenomenon. We are thus, in their view, over-inclusive and gain explanatory scope at the cost of explanatory power and precision. These charges are well put and pose an important challenge to the usefulness of the TTOM model. However, we think the apparent empirical limitations of the model reflect its current formulation in generic terms. Detailed examples in particular domains are needed to produce specific hypotheses and test the fit of models with experimental data. Moreover, plenty of evidence (both neural and behavioral) is already available to demonstrate the viability of the principles that ground our model as applied to individual cognition (e.g., Friston 2010; Keller & Mrsic-Flogel 2018).

The methods we employ to formulate TTOM are borrowed from theoretical neurobiology. These basic principles have been applied a wide range of problems and tested against empirical data, and we know they can explain and reliably predict many features of individual human behavior (Cullen et al. 2018; Mirza et al. 2016). The application of these methods to cognitive systems beyond individual humans and their brains, however, is more recent. The challenge is to figure out what should count as empirical evidence for an ecologically extended, and spatio-temporally scalable model like TTOM. We think the paleoanthropological record holds some clues that may help us answer these questions, whereas addressing a second point raised by **Brown et al.** about what they perceive as our model's failure to address debates between the internalist and externalist approaches to cognition.

The TTOM model, to clarify, proposes an "interactionist" view of cognition as at once internal to individual brains and bodies, and extended, embedded, and enhanced through external features of the social world. Work in paleo-anthropology – that emphasizes the co-evolution of human cognition with their anthropogenic environment modifications – provides explicit examples of this process. For example, Dietrich Stout's analysis of cumulative cultural evolution in the lower paleolithic suggests that capacity for Theory of Mind was significantly enhanced through selective pressure to teach and learn the making of stone tools (Stout 2011; Stout et al. 2011). In other words, tool production played a crucial role in selecting for better perspective-taking abilities. Using ethnographic evidence from the Ju/'hoansi Bushmen, Wiessner (2014) argued that the invention of fire, and the ritual practice of firelight talks at nighttime played an important role (in both phylogeny and ontogeny) in "evoking higher orders of theory of mind via the imagination, conveying attributes of people in broad networks (virtual communities), and transmitting the 'big picture' of cultural institutions that generate regularity of behavior, cooperation, and trust at the regional level." On this interactionist view, then, fire also played a role in the evolution of cognition. On more recent accounts of the "Broad Spectrum Revolution," a further leap in human cognitive, technological, and cooperative foraging skills occurred around 40,000 years ago after anthropogenic depletion of megafauna provided new selective pressures for diversification of hunting and gathering strategies (Sterelny 2007; 2011; Zeder 2012).

These diverse examples point to multiple ways in which humans have learned to resolve uncertainty through various TTOM-enhanced, environmentally mediated pathways. We have outlined a unifying account of human cultural co-evolution, which can be operationalized under the FEP. However, detailed

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

predictions and testable hypothesis will have to take on board historical contingency and boundary conditions.

R5. Modeling diversity and difference in TTOM

TTOM models the acquisition of culture by a generic agent in a stable, relatively homogeneous social niche. However, the reality in most contemporary societies is one of the high levels of diversity. In our responses to **Mirski et al.**, we have discussed how social change, emerging normativities, and social conflict can be described as processes of group selection and optimization. Many levels of difference and variation remain to be examined in our account.

Intra-cultural and inter-individual diversity, as well as nonconformity and intra-individual change (when individuals change their mind over time or in response to changes in context) raise thorny questions for TTOM. Whether our ongoing transition to a digital world with unpresented access to billions of minds is a game-changer for TTOM is another pertinent question. The commentaries by **Christopoulos & Hong**, **Bouizegarene**, and **Clark** are especially helpful in addressing these essential refinements to the model.

R5.1. Cultural diversity and contextual variation

In their commentary on the multicultural mind, **Christopoulos & Hong** remind us that humans are not simply passive recipients of the cultural models through which they interact with their world, but instead actively construct their world through a variety of learned strategies. Christopoulos & Hong discuss the case of individuals exposed to multiple cultural environments, who thereby acquire an intuitive understanding that "the very same behavior could have different causes, interpretations and consequences" depending on the cultural context in which it is manifested. We might add that such individuals may also possess reflective knowledge that the same environment affords different actions to different people. "Cultural-frame switching," thus, can be conceptualized as the ability to switch between different possibilities for thought, affect, and action by leveraging priors from different regimes of attention. This capacity is particularly evident in the case of migrants who integrate into their culture of adoption, but also in those who learn new ways of seeing and doing through travel, or who grow up with a "home" culture different that the dominant culture of the larger society around them. Christopoulos & Hong's observation that many individuals do not fit neatly within a single cultural model, we think, applies to our understanding of "culture" more generally because most people are exposed to more than one cultural framework.

The encoding of the external world with skill-bound, value-laden patterned possibilities for attention and action is only made possible via TTOM. Whether some or any affordances are entirely TTOM-free is an open question. **Fortier-Davy**, for example, disputes our claim that differences in optical illusions across culture provide an example of TTOM. On Fortier-Davy's account, the attunement of visual priors depends on external features of the environment alone. Of course, these physical environments are humanly constructed and so bear the traces of other minds but, as Fortier-Davy avers, no expectations about those minds may be necessary to learn some of the perceptual biases or expectations associated with specific cultural contexts. However, for more complex inferences the process of TTOM is essential. Thus, Fortier-Davy presents the scenario of individuals growing up amid "complex and ambiguous scenes," which he understands as structuring a "holistic perceptual style" that may induce vulnerability to "illusions requiring context-independent scrutiny." Cultural differences in automatic attention to contextual versus isolated information are well documented (Kitayama et al. 2003). Context-dependent perceptual styles are typically found in more *collectivistic* cultures in which strong attention to *social context* is primed from early childhood. Attention to complex systems of causality in the external world (such as weather patterns or the distribution of plant and animal species relevant to hunter–gatherer cultures) is also learned *socially* via immersion in particular cultural contexts and the education of attention guided by more experienced role models. Clearly, *other minds* in such settings are always instrumental in the structuring of perceptual priors. We remain convinced that for all culturally proficient humans, perceptual priors, policies, and actions are patterned in a similar fashion.

A fully worked out, testable Bayesian model of culture – operating via TTOM – would need to identify the relevant "cultural base-rates" to which people outsource their priors in different situations. Most people (even those not explicitly raised in different cultures) likely draw on a variety of cultural, subcultural, class, educational, political, and esthetic models to navigate their world from context to context and task to task. In each case, the evocation and deployment of particular cultural repertoires depends on the interaction of the individual's learning history with particular contextual affordances, norms, and expectations. How might a testable model intent on predicting such a person's behavior identify a set of relevant, *switchable* cultural base rates according to situation? The Minimal Group Paradigm experiments mentioned by **Mouras**, as we discussed above, clearly show that humans can switch group-based frames of references and allegiances in quick, effortless, and seemingly arbitrary ways (such as being randomly assorted into blue or red T-Shirt groups). The limits of such flexibility remain poorly understood. Are there locally stable, conventionally constituted base-rates (such as those associated with age, group-affiliation, gender, or class) that will tend to prevail over other markers of group identity? We are not aware of any studies testing these important questions.

R5.2. Individual variation

Bouizegarene's commentary similarly calls for refinements to our generic account of cognition and culture – this time by appealing to individual differences in conformity and receptivity to cultural affordances. Discussing individual differences in normative identity styles, Bouizegarene asks how our account could explain how "some cultured agents seek and tolerate the uncertainty of questioning their identity beyond social norms and voluntarily go about a long process of thinking autonomously about themselves, rather than using norms as an antidote to this uncertainty?" These, we agree, are pertinent questions that a fine-grained variational account of culture must address. First, we should clarify that a higher tolerance for individual uncertainty and questioning, in relation to social norms, does not entail a complete divorce from social norms. Indeed, beyond the intrinsically social (and hence patterned) dimensions of language, meaning, and collective goals in which questioning takes place, such questioning of norms is made possible only *in relation to social norms*. A fuller niche construction account of sociocultural evolution, thus, could cast the co-evolution of a spectrum of personality traits with a population from "conservative" to "novelty-seeking" (e.g.,

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

conscientiousness, anxious attachments, and agreeableness on the one hand, openness to experience, thrill-seeking, and oppositionality on the other) as a group-level solution to ensure optimal adaptability to environmental change, and optimal conservation of cumulative culture.

R5.3 Transition to the digital niche

Clark's questions about human–machine interactions and the "thermodynamics" of digitally mediated human life are important for a fuller articulation of our model that can address the changing techno-cultural landscape. For example, the emerging problem of "smartphone" or "screen" addiction may reflect new dynamics of information foraging and TTOM. We have argued elsewhere (Stendel et al., in press; Veissière & Stendel 2018) that a hyper-abundance of informational uncertainty online solicits the hyper-activation of evolutionarily old attentional biases for social and group-fitness-enhancing information. In turn, this affords an addictive relationship with screens through a constant search for social rewards, social comparison, and high-precision cultural information. We have termed these dynamics the "hypernatural monitoring hypothesis" (Veissiére & Stendel 2018). The human minds' limitations on processing vast amounts of information online have recently been described as "bottlenecking" mechanisms that favor belief-consistent, negatively valenced, predictive information of a social nature (Hill 2019). These recent mechanisms of digital niche construction have been proposed as candidate explanations for the rise of new social challenges, such as increasing extremism, political polarization, and the proliferation of misinformation (Hill 2019). The TTOM model may provide a way to formulate some of the online dynamics that contribute to these social problems.

The socially leveraged processes of epistemic foraging that underpin TTOM can be described with a few simple algorithms. The bottlenecking mechanisms of informational uncertainty minimization observed on the Internet show us just how much information about threats and group affiliations matter to human minds. Still, much work remains to be done to typologize the *attentional biases* ("choices," or "policies") that underpin TTOM. Future work will need to distinguish between biases directly geared toward other minds (such as attentional preferences for eyes, faces, group affiliation, and propositional attitudes), those that harness, enrich, and "anthropomorphize" evolutionarily older, developmentally earlier biases, and those that are not about other minds at all. Thus, we can think of automatic mechanisms that track prestige, social status, and reputation (Henrich & Gil-White 2001) as "recycling" general epistemic-foraging mechanisms (found in all living organisms) scaffolded into dominance hierarchies (found in all social mammals). These high-precision cues offer relevant information about which fitness-enhancing model to track and learn from. For humans, these dynamics operate both automatically and self-reflectively through symbolic and status cues grounded in *normatively configured hierarchies* governing optimal moral standing and social functioning. Similarly, we can think of evolutionarily old threat-detection modalities such as the negativity bias (Vaish et al. 2008) and pollution avoidance mechanisms (Stevenson et al. 2019) as becoming re-encoded via TTOM through such symbolically enriched processes as superstition, xenophobia, bullying non-conformists, hypochondria, paranoia, magical thinking, conspiracy theories, and the myriad other metaphors and narrative models that postulate the existence of "dark," "malefic" forces and agents as the "cause" of negative internal states and social problems (Boyer 2018).

R6. Applications of TTOM to psychiatry

In addition to a better understanding of social problems, we think our model can help advance the study of psychopathology by highlighting interactions between the social and neural dimensions of various disorders.

In their wide-ranging commentary, **Dumas, Gozé, & Micoulaud-Franchi** (**Dumas et al.**) provide a very useful extension of our discussion of the shared affective and phenomenological processes that underwrite human experience. To advance an "interactive turn in psychiatric semiology," Dumas et al. call for a unified computational framework that could account both for sociocultural variations in psychiatric symptomatology (including how patients interpret and enact different explanatory models of illness, and how clinicians leverage different explanatory models in their interactions with patients) and current neuroscientific findings on physiological disturbances in the brain. Indeed, they go beyond our original argument, to suggest that the TTOM model can contribute to the development of methods that examine how the whole extended cultural phenotype of humans interacts with the whole genotype in what they call *social physiology*. We welcome this ambitious project, and applaud Dumas et al.'s broader discussion of multi-scale approaches to psychiatric disorders grounded in active inference – in particular, their comments on neurodevelopmental conditions such as autism and schizophrenia that have been described as minimal self-impairments, and that may be operating at more basic levels than the cognitive-behavioral loops that are the focus of the current TTOM model.

The potential relevance of TTOM to psychiatry is also brought out by **Lifshitz & Luhrmann** in their commentary on the ways in which culture can structure the affective valence and content of hallucinations. In addition to pointing out, as Gold and Gold (2015) have done elsewhere, that all types of hallucinations appear to involve *social scenarios* and "the relevant others that humans think through" to guide their existence, Lifshitz and Luhrmann provide compelling examples of sensorially rich "hallucinations" (interacting with imagined agents who are not actually present) in contexts such as religious practice that do not involve any pathology or underlying brain dysfunction. This points to the far-reaching ways in which culture can affect perception. Their review of recent findings on the cross-cultural patterning of experience in schizophrenia also shows how basic TTOM impairments, which likely do entail dysfunctions of an organic nature (and, as such, might lead to behavior recognized as dysfunctional in any cultural context), predict widely different degrees of distress depending on the specific cultural information conveyed in hallucinations, and on local cultural assumptions about the nature of affliction. Lifshitz and Luhrmann thus raise important questions on the reach and limits of cultural influences on phenomenological plasticity of mental disorders in general.

Finally, we welcome the suggestions for extending TTOM offered by **Bolis & Schilbach**. We find their dialectical account of the interactions and relationships between the agents compelling and appreciate their call to extend our model to take into account the diversity of human experience, including neurodiversity. The idea that difficulties in interaction between the neurotypical and neurodiverse individuals result from the misalignment *between* individuals, rather than simply from some underlying deficit in the neurodiverse population, has far-reaching

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

implications for our understanding of and attitudes toward individual differences and difficulties in adaptation. Their suggestion to focus more closely on the dynamics of real-time interactions among humans points to fruitful avenues for empirically testing the TTOM model.

R7. Conclusion: The future of TTOM

We opened this discussion by presenting the "puzzle" of culture for a species characterized by immense diversity in skill sets, along with ways of thinking, feeling, and perceiving the world. We sought to clear the conceptual muddle and dispel any just-so story to describe the patterning of culture around evolved capacities for shared attention. Our model can thus be read as extending earlier multidisciplinary efforts to investigate the ways in which human language, systems of meaning, kinship, social institutions, norms and organization reflect, extend, and are constrained by basic evolved structures of the human mind. In addition to outlining how the brain and socially constructed niches are dynamically coupled in cultural learning, we have argued that these dynamics can be usefully operationalized under the FEP.

By highlighting the patterned dynamics through which the world comes to afford different things to different groups (and that lead different groups to be treated in different ways by other groups), our model offers a naturalistic account that may help operationalize some views typically understood as "socioconstructivist." This should not be read as a radical endorsement of culture as an anything-goes process, entirely divorced from natural and biological constraints. The evolutionary and developmental acquisition of cultural affordances, as we have argued, builds on a set of attentional biases for coalitional intention-tracking, threat-avoidance, and prestige-cued, and social fitness-enhancing information – where the latter maximizes an individual's access to relevant skills, explanatory models, values, moral status, social recognition, and social support. All cultures and cultural subgroups operate with these dynamics.

Although we are convinced that many human affordances are collectively modulated via TTOM, our model does not deny the existence and importance of other external and natural affordances, many of which may be configured by cultural activities that construct our material world and local niches. Once a body is equipped with culture-bound skills, and once a world is layered with culture-bound meaning, patterned dynamics of cooperative action and improvisation become possible.

After engaging with our commentators' provocative critiques and suggestions for refinement we are left with the sense that the future of TTOM looks bright. In ongoing collaborations, we are exploring how to augment the theory with affective valence, take into account individual differences and historicity, and begin to model specific domains such as epistemic bias. Once again, we are deeply grateful for this creative colloquy and look forward to continued TTOM.

References

[The letters "a" and "r" before author's initials stand for target article and response references, respectively]

Adams, R. A., Huys. Q. J. M. & Roiser, J. P. (2016) Computational psychiatry: Towards a mathematically informed understanding of mental illness. *Journal of Neurology, Neurosurgery, and Psychiatry* **87**(1):53–63. [aSPLV]

Ainley, V., Apps, M. A. J., Fotopoulou, A. & Tsakiris, M. (2016) "Bodily precision": A predictive coding account of individual differences in interoceptive accuracy. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371** (1708):20160003. https://doi.org/10.1098/rstb.2016.0003. [MA]

Ainsworth, M. D. S., Blehar, M. C., Waters, E. & Wall, S. (1978) *Patterns of attachment: A psychological study of the strange situation*. Erlbaum. [PSS]

Alem, S., Perry, C. J., Zhu, X., Loukola, O. J., Ingraham, T., Søvik, E. & Chittka, L. (2016) Associative mechanisms allow for social learning and cultural transmission of string pulling in an insect. *PLoS Biology* **14**:e1002564. [AW]

Allen, J., Garland, E. C., Dunlop, R. A. & Noad, M. J. (2018) Cultural revolutions reduce complexity in the songs of humpback whales. *Proceedings of the Royal Society B: Biological Sciences* **285**:20182088. [AW]

Allen, M. & Friston, K. J. (2018) From cognitivism to autopoiesis: Towards a computational framework for the embodied mind. *Synthese* **195**(6):2459–82. https://doi.org/10.1007/s11229-016-1288-5. [MA]

Allen, M., Levy, A., Parr, T. & Friston, K. J. (2019) In the body's eye: The computational anatomy of interoceptive inference. *BioRxiv* 603928. https://doi.org/10.1101/603928. [MA, RS, rSPLV]

Allen, M. & Tsakiris, M. (2019) The body as first prior: Interoceptive predictive processing and the primacy. In *The interoceptive mind: From homeostasis to awareness*, 1st Edition, pp. 27–45. Oxford University Press. [MA]

Anderson, B. (1983/2006) *Imagined communities: Reflections on the origin and spread of nationalism*. Verso. (Original work published in 1983). [SVdC]

Anderson, M. L. & Chemero, T. (2013) The problem with brain GUTs: Conflation of different senses of "prediction" threatens metaphysical disaster. *Behavioral and Brain Sciences* **36**(3):204–05. [EB]

Andrews, P. W., Gangestad, S. W. & Matthews, D. (2002) Adaptationism–how to carry out an exaptationist program. *Behavioral and Brain Sciences* **25**(4): 489–504, discussion 504–53. [aSPLV]

Anselme, P. & Güntürkün, O. (2019) Incentive hope: A default psychological response to multiple forms of uncertainty. *Behavioral and Brain Sciences* **42**:E58. doi:10.1017/S0140525X18002194. [PSS]

Antin, M. (1912) *The promised land*. Penguin. [GIC]

Aplin, L. M. (2019) Culture and cultural evolution in birds: A review of the evidence. *Animal Behaviour* **147**:179–87. [AW]

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A. & Sheldon, B. C. (2015) Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature* **518**:538–41. [AW]

Apperly, I. (2011) *Mindreaders: The cognitive basis of "theory of mind."* Psychology Press. [SO]

Apperly, I. A. & Butterfill, S. A. (2009) Do humans have two systems to track beliefs and belief-like states? *Psychological Review* **116**(4):953–70. [aSPLV]

Apps, M. A. & Tsakiris, M. (2014) The free-energy self: A predictive coding account of self-recognition. *Neuroscience & Biobehavioral Reviews* **41**:85–97. [MA]

Arbib, M. A. & Fellous, J. M. (2004) Emotions: From brain to robot. *Trends in Cognitive Science* **8**(12):554–61. [KBC]

Asada, M (2015) Development of artificial empathy. *Neuroscience Research* **90**:41–50. [KBC]

Asch, S. E. (1955) Opinions and social pressure. *Scientific American* **193**(5):31–35. [rSPLV]

Asch, S. E. (1956) Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied* **70**(9):1–70. [aSPLV]

Astuti, R. & Bloch, M. (2015) The causal cognition of wrong doing: Incest, intentionality, and morality. *Frontiers in Psychology* **6**:136. [aSPLV]

Atkinson, M. (2008) Triathlon, suffering and exciting significance. *Leisure Studies* **27** (2):165–80. [rSPLV]

Atran, S. (2003) Genesis of suicide terrorism. *Science* **299**(5612):1534–1539. [rSPLV]

Atran, S. & Ginges, J. (2012) Religious and sacred imperatives in human conflict. *Science* **336**(6083):855–57. https://doi.org/10.1126/science.1216902. [SVdC]

Badcock, P. B. (2012) Evolutionary systems theory: A unifying meta-theory of psychological science. *Review of General Psychology* **16**(1):10–23. [aSPLV]

Badcock, P. B., Davey, C. G., Whittle, S., Allen, N. B. & Friston, K. J. (2017) The depressed brain: An evolutionary systems theory. *Trends in Cognitive Sciences* **21**(3):182–94. [aSPLV]

Badcock, P. B., Friston, K. J. & Ramstead, M. J. (2019) The hierarchically mechanistic mind: A free-energy formulation of the human psyche. *Physics of Life Reviews* **31**: 104–21. [aSPLV]

Baldwin, M. W. (1992) Relational schemas and the processing of social information. *Psychological Bulletin* **112**(3):461–84. [aSPLV]

Barrett, L. (2017) *How emotions are made: The secret life of the brain*. Houghton Mifflin Harcourt. [RS]

Barrett, L. F. & Simmons, W. K. (2015) Interoceptive predictions in the brain. *Nature Reviews Neuroscience* **16**(7):419–29. https://doi.org/10.1038/nrn3950. [MA]

Becchio, C., Manera, V., Sartori, L., Cavallo, A. & Castiello, U. (2012) Grasping intentions: From thought experiments to empirical evidence. *Frontiers in Human Neuroscience* **6**:117. https://doi.org/10.3389/fnhum.2012.00117. [GP]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Becker, J. (2001) Anthropological perspectives on music and emotion. In: *Music and emotion: Theory and research*, eds. P. N. Juslin & J. A. Sloboda, pp. 135–60. Oxford University Press. [BV]

Beckes, L., Simons, K., Lewis, D., Le, A. & Edwards, W. (2017) Desperately seeking support: Negative reinforcement schedules in the formation of adult attachment associations. *Social Psychological and Personality Science* **8**(2):229–38. doi:10.1177/1948550616671402. [PSS]

Benet-Martínez, V., Leu, J., Lee, F. & Morris, M. W. (2002) Negotiating biculturalism: Cultural frame switching in biculturals with oppositional versus compatible cultural identities. *Journal of Cross-Cultural Psychology* **33**(5):492–516. [GIC]

Bengio, Y. (2014) Evolving culture versus local minima. In: *Growing adaptive machines: Combining development and learning in artificial neural networks*, eds. T. Kowaliw, N. Bredeche & R. Doursat, pp. 109–138. Studies in Computational Intelligence. Springer. [arSPLV]

Bennett, A. (1999) Subcultures or neo-tribes? Rethinking the relationship between youth, style and musical taste. *Sociology* **33**(3):599–617. [BV]

Berlyne, D. E. (1966) Curiosity and exploration. *Science* **153**(3731):25–33. https://doi.org/10.1126/science.153.3731.25. [MF-D]

Berrios, G. E. (1996) *The history of mental symptoms: Descriptive psychopathology since the nineteenth century*. Cambridge University Press. [GD]

Bertolotti, T. & Magnani, L. (2017) Theoretical considerations on cognitive niche construction. *Synthese* **194**(12):4757–79. [aSPLV]

Berwick, R. C., Chomsky, N. & Piattelli-Palmarini, M. (2013) Poverty of the stimulus stands: Why recent challenges fail. In: *Rich Languages from Poor Inputs*, eds. M. Piattelli-Palmarini & R. C. Berwick, pp. 19–42. Oxford University Press. [aSPLV]

Berzonsky, M. D. (1989) Identity style conceptualization and measurement. *Journal of Adolescent Research* **4**(3):268–82. [NB]

Berzonsky, M. D. (2011) A social-cognitive perspective on identity construction. In *Handbook of identity theory and research*, eds. S. J. Schwartz, K. Luyckx & V. L. Vignoles, pp. 55–76. Springer. [NB]

Bicchieri, C. (2006) *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press. [MC]

Bicchieri, C. (2016) *Norms in the wild: How to diagnose, measure, and change social norms*. Oxford University Press. [MC]

Bickhard, M. H. (1992) How does the environment affect the person? In *Children's development within social context*, eds. L. T. Winegar & J. Valsiner, pp. 63–92. L. Erlbaum. [RM]

Bickhard, M. H. (2007) Language as an interaction system. *New Ideas in Psychology* **25**(2):171–87. https://doi.org/10.1016/j.newideapsych.2007.02.006. [RM]

Bickhard, M. H. (2008) Social ontology as convention. *Topoi* **27**(1–2):139–49. https://doi.org/10.1007/s11245-008-9036-1. [RM]

Bickhard, M. H. (2009) The interactivist model. *Synthese* **166**(3):547–91. https://doi.org/10.1007/s11229-008-9375-x. [RM]

Bickhard, M. H. (2015) Toward a model of functional brain processes II: Central nervous system functional macro-architecture. *Axiomathes* **25**(4):377–407. https://doi.org/10.1007/s10516-015-9276-9. [RM]

Bickhard, M. H. (2016) The anticipatory brain: Two approaches. In *Fundamental issues of artificial intelligence*, ed. V. C. Müller, pp. 259–81. Springer International Publishing. [RM]

Bijleveld, E., Scheepers, D. & Ellemers, N. (2012) The cortisol response to anticipated intergroup interactions predicts self-reported prejudice. *PLoS ONE* **7**(3):e33681. [aSPLV]

Bilu, Y. (2013) We want to see our king: Apparitions in Messianic Habad. *Ethos* **41**:98–126. [ML]

Binmore, K. (1994) *Game theory and the social contract, vol. I. Playing fair*. MIT Press. [MC]

Birch, J. (m.s.) Toolmaking and the origin of normative cognition. [AB]

Bloom, P. (2005) *Descartes' baby: How the science of child development explains what makes us human*. Random House. [aSPLV]

Bloom, P. (2017) *Against empathy: The case for rational compassion*. Random House. [aSPLV]

Bolis, D., Balsters, J., Wenderoth, N., Becchio, C. & Schilbach, L. (2017) Beyond autism: Introducing the dialectical misattunement hypothesis and a Bayesian account of intersubjectivity. *Psychopathology* **50**(6):355–72. https://doi.org/10.1159/000484353. [aSPLV, DB, GD]

Bolis, D. & Schilbach, L. (2017) Beyond one Bayesian brain: Modeling intra- and interpersonal processes during social interaction: Commentary on "Mentalizing homeostasis: The social origins of interoceptive inference" by Fotopoulou & Tsakiris. *Neuropsychoanalysis* **19**(1):35–38. [DB]

Bolis, D. & Schilbach, L. (2018a) Observing and participating in social interactions: Action perception and action control across the autistic spectrum. *Developmental Cognitive Neuroscience* **29**:168–75. https://doi.org/10.1016/j.dcn.2017.01.009. [aSPLV, DB]

Bolis, D. & Schilbach, L. (2018b) "I Interact Therefore I Am": The self as a historical product of dialectical attunement. *Topoi* 1–14. https://doi.org/10.1007/s11245-018-9574-0. [DB, GD]

Bonawitz, E., Shafto, P., Gweon, H., Goodman, N. D., Spelke, E. & Schulz, L. (2011) The double-edged sword of pedagogy: Instruction limits spontaneous exploration and discovery. *Cognition* **120**(3):322–30. http://doi.org/10.1016/j.cognition.2010.10.001. [HG]

Bono, A. E. J., Whiten, A., van Schaik, C., Krutzen, M., Eichenberger, F., Schnider, A. & van de Waal, E. (2018) Payoff- and sex-biased social learning interact in a wild primate population. *Current Biology* **28**:2800–805. [AW]

Borsboom, D., Cramer, A. & Kalis, A. (2018) Brain disorders? Not really… Why network structures block reductionism in psychopathology research. *Behavioral and Brain Sciences* **42**:1–54. https://doi.org/10.1017/S0140525X17002266. [GD]

Bostrom, N. (2014) *Superintelligence: Paths, dangers, strategies*. Oxford: Oxford University Press. ISBN 978-0199678112. [KBC]

Bourdieu, P. (1977) *Equisse D'une Théorie de La Pratique*. Cambridge University Press. [aSPLV, JK]

Bourdieu, P. (1984) *Distinction: A social critique of the judgement of taste*. Harvard University Press. [aSPLV, JK]

Bowles, S. & Gintis, H. (2013) *A cooperative species: Human reciprocity and its evolution*. Princeton University Press. [MRZ]

Boyd, R. & Richerson, P. (1985) *Culture and the evolutionary process*. University of Chicago Press. [MRZ]

Boyd, R. & Richerson, P. (2001) Norms and bounded rationality. In *Bounded rationality: The adaptive toolbox*, eds. G. Gigerenzer & R. Selten, pp. 281–96. MIT Press. [MC]

Boyd, R. & Richerson P. J. (2005) *The origin and evolution of cultures*. Oxford University Press. [aSPLV]

Boyd, R., Richerson, P. J. & Henrich, J. (2011) The cultural niche: Why social learning is essential for human adaptation. *Proceedings of the National Academy of Sciences* **108**(Suppl_2):10918–25. http://doi.org/10.1073/pnas.1100290108. [HG]

Boyer, P. (2018) *Minds make societies: How cognition explains the world humans create*. Yale University Press. [arSPLV]

Boyer, P. & Liénard, P. (2006) Why ritualized behavior? Precaution systems and action parsing in developmental, pathological and cultural rituals. *The Behavioral and Brain Sciences* **29**(6):595–613, discussion 613–50. [rSPLV]

Boyer, P. & Liénard, P. (2008) Ritual behavior in obsessive and normal individuals: Moderating anxiety and reorganizing the flow of action. *Current Directions in Psychological Science* **17**(4):291–94. [rSPLV]

Brandi, M. L., Kaifel, D., Bolis, D. & Schilbach, L. (2019) The interactive self – A review on simulating social interactions to understand the mechanisms of social agency. *i-com* **18**(1):17–31. [DB]

Brewer, M. (1979) In-group bias in the minimal intergroup situation. A cognitive-motivational analysis. *Psychological Bulletin* **86**(2):307. [MH]

Bridgers, S., Jara-Ettinger, J. & Gweon, H. (2019) Young children consider the expected utility of others' learning to decide what to teach. *Nature Human Behaviour* **4**(2):144–52. [HG]

Briegel, H. J. (2012) On creative machines and the physical origins of freedom. *Scientific Reports* **2**:522. [KBC]

Briley, D. A., Morris, M. W. & Simonson, I. (2005) Cultural chameleons: Biculturals, conformity motives, and decision making. *Journal of Consumer Psychology* **15**(4):351–62. [GIC]

Brown, D. E. (2004) Human universals, human nature & human culture. *Daedalus* **133**(4):47–54. https://doi.org/10.1162/0011526042365645. [aSPLV]

Bruineberg, J., Kiverstein, J. & Rietveld, E. (2018a) The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese* **195**(6):2417–44. https://doi.org/10.1007/s11229-016-1239-1. [MA]

Bruineberg, J. & Rietveld, E. (2014) Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience* **8**:599. [aSPLV]

Bruineberg, J., Rietveld, E., Parr, T., van Maanen, L. & Friston, K. J. (2018b) Free-energy minimization in joint agent-environment systems: A niche construction perspective. *Journal of Theoretical Biology* **455**:161–78. https://doi.org/10.1016/j.jtbi.2018.07.002. [aSPLV, JK, MA]

Bryant, G. A., Fessler, D. M., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., Petersen, M. B., Bickham, S. T., Bolyanatz, A., Chavez, B., De Smet, D., Díaz, C., Fančovičová, J., Fux, M., Giraldo-Perez, P., Hu, A., Kamble, S. V., Kameda, T., Li, N. P., Luberti, F. R., Prokop, P., Quintelier, K., Scelza, B. A., Shin, H. J., Soler, M., Stieger, S., Toyokawa, W., van den Hende, E. A., Viciana-Asensio, H., Yildizhan, S. E., Yong, J. C., Yuditha, T. & Zhou, Y. (2016) Detecting affiliation in colaughter across 24 societies, *Proceedings of the National Academy of Sciences* **113**(17):4682–87. [rSPLV]

Burkart, J. M., Hrdy, S. B. & Van Schaik, C. P. (2009) Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology: Issues, News, and Reviews* **18**(5):175–86. [aSPLV]

Campbell, R. J. (2015) *The metaphysics of emergence*. Palgrave Macmillan. [RM]

Cardon, A. (2006) Artificial consciousness, artificial emotions, and autonomous robots. *Cognitive Processes* **7**(4):245–67. [KBC]

Carruthers, P. (2015) Perceiving mental states. *Consciousness and Cognition* **36**:498–507. [KD]

Carruthers, P. & Smith, P. K. (1996) *Theories of theories of mind*. Cambridge University Press. [aSPLV]

Cermolacce, M., Sass, L. & Parnas, J. (2010) What is bizarre in bizarre delusions? A critical review. *Schizophrenia Bulletin* **36**(4):667–79. https://doi.org/10.1093/schbul/sbq001. [GD]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Chanes, L. & Barrett, L. F. (2016) Redefining the role of limbic areas in cortical processing. *Trends in Cognitive Sciences* **20**(2):96–106. https://doi.org/10.1016/j.tics.2015.11.005. [MA]

Charafeddine, R., Mercier, H., Clément, F., Kaufmann, L., Berchtold, A., Reboul A. & Van der Henst, J.-B. (2015) How preschoolers use cues of dominance to make sense of their social environment. *Journal of Cognition and Development* **16**(4):587–607. [FC]

Chater, N., Misyak, J., Watson, D., Griffiths, N. & Mouzakitis, A. (2018) Negotiating the traffic: Can cognitive science help make autonomous vehicles a reality? *Trends in Cognitive Sciences* **22**(2):93–95. [GP]

Chemero, A. (2009) *Radical embodied cognitive science*. MIT Press. [AB, aSPLV]

Chemero, A. (2011) *Radical embodied cognitive science*. MIT press. [rSPLV]

Cheng, C. Y., Lee, F. & Benet-Martínez, V. (2006) Assimilation and contrast effects in cultural frame switching: Bicultural identity integration and valence of cultural cues. *Journal of Cross-Cultural Psychology* **37**(6):742–60. [GIC]

Cheng, J. T., Tracy, J. L., Foulsham, T., Kingstone, A. & Henrich, J. (2013) Two ways to the top: Evidence that dominance and prestige are distinct yet viable avenues to social rank and influence. *Journal of Personality and Social Psychology* **104**(1):103–25. [aSPLV]

Cheon, B. K. & Hong, Y. (2019) Aversive responses towards culture fusion is moderated by the source of foreign cultural inflows. *PLoS One* (under review). [GIC]

Chetverikov, A. & Kristjánsson, Á. (2016) On the joys of perceiving: Affect as feedback for perceptual predictions. *Acta Psychologica* **169**:1–10. https://doi.org/10.1016/j.actpsy.2016.05.005. [MF-D]

Chew, Y. H., Wenden, B., Flis, A., Mengin, V., Taylor, J., Davey, C. L., Tindal, C., Thomas, H., Ougham, H. J., de Reffye, P., Stitt, M., Williams, M., Muetzelfeldt, R., Halliday, K. J. & Millar A. J. (2014) Multiscale digital Arabidopsis predicts individual organ and whole-organism growth. *Proceedings of the National Academy of Science of the USA* **111**(39):E4127–36. [KBC]

Chmiel, A. & Schubert, E. (2017) Back to the inverted-U for music preference: A review of the literature. *Psychology of Music* **45**(6):886–909. https://doi.org/10.1177/0305735617697507. [MF-D]

Chomsky, N. (1996) *Studies on semantics in generative grammar*. Walter de Gruyter. [aSPLV]

Christensen, W. & Michael, J. (2016) From two systems to a multi-systems architecture for mindreading. *New Ideas in Psychology* **40**:48–64. [arSPLV]

Christopoulos, G. I. & Tobler, P. N. (2016) Culture as a response to uncertainty: Foundations of computational cultural. In *The Oxford handbook of cultural neuroscience*, eds. J. Y. Chiao, S.-C. Li, R. Seligman & R. Turner, pp. 81–104. Oxford University Press. [GIC]

Chudek, M., Heller, S., Birch, S. & Henrich, J. (2012) Prestige-biased cultural learning: Bystander's differential attention to potential models influences children's learning. *Evolution and Human Behavior* **33**(1):46–56. [FC]

Chudek, M., McNamara, R., Burch, S., Bloom, P. & Henrich, J. (2013) Developmental and cross-cultural evidence for intuitive dualism. Unpublished manuscript, University of British Columbia. [aSPLV]

Cialdini, R. B. & Goldstein, N. J. (2004) Social influence: Compliance and conformity. *Annual Review of Psychology* **55**:591–621. [aSPLV]

Clark, A. (2006) Language, embodiment, and the cognitive niche. *Trends in Cognitive Sciences* **10**(8):370–74. [aSPLV]

Clark, A. (2008) *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University Press. [aSPLV]

Clark, A. (2013a) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *The Behavioral and Brain Sciences* **36**(3):181–204. https://doi.org/10.1017/S0140525X12000477. [aSPLV, RM]

Clark, A. (2013b) Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* **36**(3):181–253. [MF-D]

Clark, A. (2017a) A nice surprise? Predictive processing and the active pursuit of novelty. *Phenomenology and the Cognitive Sciences* **17**(3):521–34. doi:10.1007/s11097-017-9525-z. [TP]

Clark, A. (2017b) How to knit your own Markov blanket. In *Philosophy and predictive processing*, eds. T. K. Metzinger & W. Wiese. Frankfurt am Main: MIND Group. [rSPLV]

Clark, A. & Chalmers, D. (1998) The extended mind. *Analysis* **58**(1):7–19. [aSPLV]

Clark, K. B. (1988) *Prejudice and your child*. Wesleyan University Press. [aSPLV]

Clark, K. B. (2012) A statistical mechanics definition of insight. In *Computational intelligence*, ed. A. G. Floares, pp. 139–62. Nova Science. ISBN 9781620819012. [KBC]

Clark, K. B. (2015) Insight and analysis problem solving in microbes to machines. *Progress in Biophysics and Molecular Biology* **119**:183–93. [KBC]

Clark, K. B. (2017) The humanness of artificial nonnormative personalities. *Behavioral and Brain Sciences* **40**:e259. [KBC]

Clark, K. B. (2019) Unpredictable homeodynamic and ambient constraints on irrational decision making of aneural and neural foragers. *Behavioral and Brain Sciences* **42**:e40. [KBC]

Clark, K. B. (in press a) Humanizing social digibots for personalized mobile health applications. In *Artificial intelligence in precision health*, ed. D. Barh. Elsevier. [KBC]

Clark, K. B. & Clark, M. K. (1939) The development of consciousness of self and the emergence of racial identification in Negro preschool children. *Journal of Social Psychology* **10**(4):591–99. [aSPLV]

Clément, F. & Dukes, D. (2017) Social appraisal and social referencing: Two components of affective social learning. *Emotion Review* **9**(3):253–61. [FC]

Clément, F. & Dukes, D. (in press) Affective social learning: A lens for developing a fuller picture of socialization processes. In *The Oxford handbook of emotional development*, eds. D. Dukes, A. Samson & E. Walle. Oxford University Press. Forthcoming in 2021. [FC]

Coakley, S. & Shelemay, K. K. (2007) *Pain and its transformations: The interface of biology and culture*. Harvard University Press. [rSPLV]

Cole, J. (2016) Accessing hominin cognition: Language and social signaling in the lower to middle Palaeolithic. In *Cognitive models in Palaeolithic archaeology*, eds. T. G. Wynn & F. L. Coolidge, pp. 157–95. Oxford University Press. [RLB]

Cole, M. (1996) *Cultural psychology: A once and future discipline*. Harvard University Press. [EB]

Colombo, M. (2014) Two neurocomputational building blocks of social norm compliance. *Biology & Philosophy* **29**(1):71–88. [MC]

Colombo, M. (2017) Social motivation in computational neuroscience. Or if brains are prediction machines, then the Humean theory of motivation is false. In *Routledge handbook of philosophy of the social mind*, ed. J. Kiverstein, pp. 320–40. Routledge. [MC]

Colombo, M. & Wright, C. (2018) First principles in the life sciences: The free-energy principle, organicism, and mechanism. *Synthese* 1–26. https://doi.org/10.1007/s11229-018-01932-w. [MC]

Constant, A., Bervoets, J., Hens, K. & Van de Cruys, S. (2018a) Precise worlds for certain minds: An ecological perspective on the relational self in autism. *Topoi* 1–12. https://doi.org/10.1007/s11245-018-9546-4. [aSPLV]

Constant, A., Clark, A. & Friston, J. (2019a) Representation wars: Enacting an armistice through active inference. *PhilSci Archive* preprint. http://philsci-archive.pitt.edu/16641/. [rSPLV]

Constant, A., Ramstead, M. J. D., Veissière, S. P. L., Campbell, J. O. & Friston, K. (2018b) A variational approach to niche construction. *Journal of the Royal Society Interface* **15**(141):20170685. [aSPLV]

Constant, A., Ramstead, M. J. D., Veissière, S. P. L. & Friston, K. J. (2019b) Regimes of expectations: An active inference model of social conformity and decision making. *Frontiers in Psychology* **10**:679. https://doi.org/10.3389/fpsyg.2019.00679. [arSPLV, DB]

Contreras Kallens PA, Dale, R. & Smaldino, P. E. (2018) Cultural evolution of categorization. *Cognitive Systems Research* **52**:765–774. [MRZ]

Coolidge, F. L. & Wynn T. (2018) *The rise of Homo sapiens: The evolution of modern thinking*, 2nd Edition. Oxford University Press. [RLB]

Coopersmith, J. (2017) *The lazy universe: An introduction to the principle of least action*. Oxford University Press. [aSPLV]

Corlett, P. R., Horga, G., Fletcher, P. C., Alderson-Day, B., Schmack, K. & Powers, A. R. III (2018) Hallucinations and strong priors. *Trends in Cognitive Sciences* **23**(2):114–27. [ML]

Csibra, G. & Gergely, G. (2009) Natural pedagogy. *Trends in Cognitive Sciences* **13**(4):148–53. [arSPLV, HG]

Csibra, G. & Gergely, G. (2011) Natural pedagogy as evolutionary adaptation. *Philosophical Transactions of the Royal Society, Series B: Biological Sciences* **366**(1567):1149–57. https://doi.org/10.1098/rstb.2010.0319. [aSPLV, GP]

Cullen, M., Davey, B., Friston, K. J. & Moran, R. J. (2018) Active inference in OpenAI Gym: A paradigm for computational investigations into psychiatric illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **3**(9):809–18. https://doi.org/10.1016/j.bpsc.2018.06.010. [arSPLV]

Danchin, E., Nöbel, S., Pocheville, A., Dagaeff, A.-C., Demay, L., Alphand, M., Ranty-Roby, S., van Renssen, L., Monier, M., Gazagne, E., Allain, M. & Isabel, G. (2018) Cultural flies: Conformist social learning in fruit flies predicts long-lasting mate-choice traditions. *Science* **362**:1025–30. [AW]

Davies, J. (2016) Program good ethics into artificial intelligence. *Nature* **538**(7625):291. doi:10.1038/538291a. [KBC]

Dayan, P. & Daw, N. D. (2008) Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience* **8**(4):429–53. [GIC]

De Bruin, L. C. & Strijbos, D. W. (2020) Does confabulation pose a threat to first-person authority? Mindshaping, self-regulation and the importance of self-know-how. *Topoi* **39**(1):151–61. doi:10.1007/s11245-019-09631-y. [rSPLV, JM]

De Castro, E. V. (2009) *Métaphysiques Cannibales: Lignes D'anthropologie Post-Structurale*. Presses universitaires de France. [aSPLV]

Dehaene, S. & Cohen, L. (2007) Cultural recycling of cortical maps. *Neuron* **56**(2):384–98. [aSPLV]

De Jaegher, H. (2013) Embodiment and sense-making in autism. *Frontiers in Integrative Neuroscience* **7**:15. https://doi.org/10.3389/fnint.2013.00015. [RLB]

De Jaegher, H. & Di Paolo, E. (2007) Participatory sense-making. *Phenomenology and the Cognitive Sciences* **6**(4):485–507. [DB]

Delplanque, J., De Loof, E., Janssens, C. & Verguts, T. (2019) The sound of beauty: How complexity determines aesthetic preference. *Acta Psychologica* **192**:146–52. https://doi.org/10.1016/j.actpsy.2018.11.011. [MF-D]

Derex, M. & Boyd, R. (2015) The foundations of the human cultural niche. *Nature Communications* **6**:8398. [KBC]

Desimone, R. (1996) Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences of the USA* **93**(24):13494–99. [TP]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Di Paolo, E. A., Cuffari, E. C. & De Jaegher, H. (2018) *Linguistic bodies: The continuity between life and language*. MIT Press. [DB, rSPLV]

Diamond, J. (2010) Two views of collapse. *Nature* **463**(7283):880–81. [rSPLV]

Diehl, M. (1990) The minimal group paradigm: Theoretical explanations and empirical findings. *European Review of Social Psychology* **1**(1):263–92. [MH]

Donaldson, M. (1978) *Children's minds*. Fontana. [EB]

Douglas, M. (1986) *How institutions think*. Syracuse University Press. [MC]

Dukes, D. & Clément, F. eds. (2019) *Foundations of* affective social learning: *Conceptualising* the social transmission of *value*. Cambridge University Press. [FC]

Dumas, G. (2011) Towards a two-body neuroscience. *Communicative & Integrative Biology* **4**(3):349–52. https://doi.org/10.4161/cib.4.3.15110. [GD]

Dumas, G., Chavez, M., Nadel, J. & Martinerie, J. (2012) Anatomical connectivity influences both intra-and inter-brain synchronizations. *PloS one* **7**(5). [GD]

Dumas, G., Kelso, J. A. & Nadel, J. (2014) Tackling the social cognition paradox through multi-scale approaches. *Frontiers in Psychology* **5**:882. https://doi.org/10.3389/fpsyg.2014.00882. [DB, GD]

Dumas, G., Nadel, J., Soussignan, R., Martinerie, J. & Garnero, L. (2010) Inter-brain synchronization during social interaction. *PLoS ONE* **5**(8):e12166. https://doi.org/10.1371/journal.pone.0012166. [GD]

Dunbar, R. I. M. (2003) The social brain: Mind, language, and society in evolutionary perspective. *Annual Review of Anthropology* **32**(1):163–81. [aSPLV]

Dunbar, R. I. M. (2004) Gossip in evolutionary perspective. *Review of General Psychology* **8**(2):100–10. [aSPLV]

Duranti, A. (2015) *The anthropology of intentions*. Cambridge University Press. [aSPLV]

Durkheim, E. (1985/2014) *The rules of sociological method: And selected texts on sociology and its method*. Simon and Schuster. (Original work published in 1985). [aSPLV]

Dweck, C. S. (2013) *Self-theories: Their role in motivation, personality, and development*. Taylor & Francis. https://content.taylorfrancis.com/books/download?dac=C2009-0-07336-6&isbn=9781317710332&format=googlePreviewPdf. [aSPLV]

Eck, D. (2015) *The Encultured Mind: From Cognitive Science to Social Epistemology* (Ph.D. thesis). University of South Florida. [RM]

Eck, D. & Levine, A. (2017) Prioritizing otherness: The line between vacuous individuality and hollow collectivism. In *Studies in the philosophy of sociality: Volume 9. Sociality and normativity for robots: Philosophical inquiries into human-robot interactions*, eds. R. Hakli & J. Seibt, pp. 67–87. Springer. [RM]

Einarsson, A. & Ziemke, T. (2017) Exploring the multi-layered affordances of composing and performing interactive music with responsive technologies. *Frontiers in Psychology* **8**:1701. [aSPLV]

Elster, J. (1989) Social norms and economic theory. *Journal of Economic Perspectives* **3**(4):99–117. [MC]

Fabry, R. E. (2018) Betwixt and between: The enculturated predictive processing approach to cognition. *Synthese* **195**:2483–518. [aSPLV]

Fan, X. & Markram, H. (2019) A brief history of simulation neuroscience. *Frontiers in Neuroinformatics* **13**:32. [KBC]

Feinman, S. (1982) Social referencing in infancy. *Merrill-Palmer Quarterly* **28**:445–70. https://www.jstor.org/stable/pdf/23086154.pdf. [aSPLV]

Feldman, H. & Friston, K. J. (2010) Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience* **4**:215. [aSPLV]

Feldman, J. (2013) Tuning your priors to the world. *Topics in Cognitive Science* **5**(1):13–34. [aSPLV]

Fiebich, A. & Coltheart, M. (2015) Various ways to understand other minds. Towards a pluralistic approach to the explanation of social understanding. *Mind and Language* **30**(3):235–58. [KD]

Fortier, M. & Kim, S. (2017) From the impossible to the improbable: A probabilistic account of magical beliefs and practices across development and cultures. In: *The science of lay theories: How beliefs shape our cognition, behavior, and health*, eds. C. Zedelius, B. Müller & J. Schooler, pp. 265–315. Springer. [MF-D]

Fortuna, M. A., Zaman, L., Wagner, A. P. & Ofria, C. (2013) Evolving digital ecological networks. *PLoS Computational Biology* **9**(3):e1002928. [KBC]

Frank, M. C. (2013) Throwing out the Bayesian baby with the optimal bathwater: Response to Endress (2013). *Cognition* **128**(3):417–23. https://doi.org/10.1016/j.cognition.2013.04.010. [MF-D]

Frank, S. A. (1995) George Price's contributions to evolutionary genetics. *Journal of Theoretical Biology* **175**(3):373–88. [MRZ]

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C. & Hewitt, J. K. (2006) Not all executive functions are related to intelligence. *Psychological Science* **17**:172–179. [GIC]

Friston, K. (2005) A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**(1456):815–36. [aSPLV]

Friston, K. (2009) The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences* **13**(7):293–301. [KD]

Friston, K. J. (2010) The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience* **11**(2):127–38. https://doi.org/10.1038/nrn2787. [arSPLV, DB, MF-D]

Friston, K. (2011) Embodied inference: Or "I think therefore I am, if I am what I think." In: *Implications of Embodiment: Cognition and Communication*, eds. W. Tschacher & C. Bergomi, pp. 89–125. [aSPLV]

Friston, K. (2012) A free energy principle for biological systems. *Entropy* **14**:2100–121. doi:10.3390/e14112100. [KD]

Friston, K. (2013) Life as we know it. *Journal of the Royal Society Interface* **10**(86):20130475. [aSPLV]

Friston K. (2019) A free energy principle for a particular physics. *arXiv e-prints*, 1–148. Retrieved from https://ui.adsabs.harvard.edu/abs/2019arXiv190610184F. [TP]

Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J. & Pezzulo, G. (2016) Active inference and learning. *Neuroscience and Biobehavioral Reviews* **68**:862–79. [arSPLV]

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. & Pezzulo, G. (2017a) Active inference: A process theory. *Neural Computation* **29**(1):1–49. https://doi.org/10.1162/NECO_a_00912. [arSPLV, GP]

Friston, K. J., Fortier, M. & Friedman, D. A. (2018) Of woodlice and men: A Bayesian account of cognition, life and consciousness. An interview with Karl Friston. *ALIUS Bulletin* **2**:17–43. [MF-D]

Friston, K. J. & Friston, D. A. (2013) A free energy formulation of music generation and perception: Helmholtz revisited. In *Sound-perception-performance*. ed. R. Bader, pp. 43–69. Springer, Heidelberg: Springer International Publishing. [rSPLV]

Friston, K. J. & Frith, C. D. (2015a) A duet for one. *Consciousness and Cognition* **36**:390–405. https://doi.org/10.1016/j.concog.2014.12.003. [MA, GD]

Friston, K. J. & Frith, C. D. (2015b) Active inference, communication and hermeneutics. *Cortex* **68**:129–43. https://doi.org/10.1016/j.cortex.2015.03.025. [aSPLV, DB, MA]

Friston, K. J., Kilner, J. & Harrison, L. (2006) A free energy principle for the brain. *Journal of Physiology, Paris* **100**(1–3):70–87. [aSPLV]

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A. & Ondobaka, S. (2017b) Active inference, curiosity and insight. *Neural Computation* **29**(10):2633–83. [arSPLV]

Friston, K. J., Redish, A. D. & Gordon, J. A. (2017) Computational nosology and precision psychiatry. *Computational Psychiatry* **1**:2–23. https://doi.org/10.1162/CPSY_a_00001. [GD, GP]

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T. & Pezzulo, G. (2015) Active inference and epistemic value. *Cognitive Neuroscience* **6**(4):187–214. https://doi.org/10.1080/17588928.2015.1020053. [aSPLV, GP]

Friston, K. J., Samothrakis, S. & Montague, R. (2012) Active inference and agency: Optimal control without cost functions. *Biological Cybernetics* **106**(8–9):523–41. https://doi.org/10.1007/s00422-012-0512-8. [RM, BV]

Friston, K. J., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T. & Dolan, R. J. (2014a) The anatomy of choice: Dopamine and decision-making. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**(1655):20130481. https://doi.org/10.1098/rstb.2013.0481. [aSPLV]

Friston, K. J. & Stephan, K. E. (2007) Free-energy and the brain. *Synthese* **159**(3):417–58. [aSPLV]

Friston, K. J., Stephan, K. E., Montague, R. & Dolan, R. J. (2014b) Computational psychiatry: The brain as a phantastic organ. *Lancet Psychiatry* **1**(2):148–58. [aSPLV]

Friston, K., Thornton, C. & Clark, A. (2012a) Free-energy minimization and the dark-room problem. *Frontiers in Psychology* **3**:130. https://doi.org/10.3389/fpsyg.2012.00130. [aSPLV, BV]

Froese, T., Iizuka, H. & Ikegami, T. (2013) From synthetic modeling of social interaction to dynamic theories of brain-body-environment-body-brain systems. *Behavioral and Brain Sciences* **36**(4):420–21. [DB]

Fuchs, T. (2015) Pathologies of intersubjectivity in autism and schizophrenia. *Journal of Consciousness Studies* **22**(1–2):191–214. [GD]

Fulda, F. C. (2017) Natural agency: The case of bacterial cognition. *Journal of the American Philosophical Association* **3**(1):69–90. [aSPLV]

Fung, P. (2015) Robots with heart. *Scientific American* **313**(5):60–63. [KBC]

Gaines, A. D. & Farmer, P. E. (1986) Visible saints: Social cynosures and dysphoria in the Mediterranean tradition. *Culture, Medicine and Psychiatry* **10**(4):295–330. [rSPLV]

Gallagher, S. (2001) The practice of mind. Theory, simulation, or primary interaction? *Journal of Consciousness Studies* **8**(5–7):83–108. [KD]

Gallagher, S. (2008) Direct perception in the intersubjective context. *Consciousness and Cognition* **17**:535–43. [KD]

Gallagher, S. (2017) *Enactivist interventions: Rethinking the mind*. Oxford University Press. [aSPLV]

Gallagher, S. & Allen, M. (2018) Active inference, enactivism and the hermeneutics of social cognition. *Synthese* **195**(6):2627–48. https://doi.org/10.1007/s11229-016-1269-8. [aSPLV, DB, MA]

Gallese, V. (2003) The manifold nature of interpersonal relations: The quest for a common mechanism. *Philosophical Transactions of the Royal Society B* **358**:517–28. doi:10.1098/rstb.2002.1234. [KD]

Gallese, V. & Goldman, A. (1998) Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* **2**(12):493–501. [aSPLV]

Gavrilets, S. & Vose, A. (2006) The dynamics of Machiavellian intelligence. *Proceedings of the National Academy of Sciences of the United States of America* **103**(45):16823–28. [aSPLV]

Gebauer, L., *et al.* (2016) Oxytocin improves synchronisation in leader-follower interaction. Scientific Reports 6:38416. [BV]

Geertz, C. (1973) *The interpretation of culture*. Basic Books. [aSPLV, ML]

Gendron, M. & Barrett, L. F. (2018) Emotion perception as conceptual synchrony. *Emotion Review* **10**(2):101–10. [rSPLV]
Gershman, S. J. (2019) How to never be wrong. *Psychonomic Bulletin & Review* **26**(1): 13–28. http://doi.org/10.3758/s13423-018-1488-8. [SVdC]
Gibson, J. J. (1979) *The ecological approach to visual perception*. Houghton Mifflin. [aSPLV, JK]
Gibson, J. J. (1986) *The ecological approach to visual perception*. Lawrence Erlbaum. [rSPLV]
Gillings, M. R., Hibert, M. & Kemp, D. J. (2016) Information in the biosphere: Biological and digital worlds. *Trends in Ecology and Evolution* **31**(3):180–89. [KBC]
Gintis, H. (2007) A framework for the unification of the behavioral sciences. *Behavioral and Brain Sciences* **30**(1):1–16. [MC]
Gödel, K. (1931) Über formal unentscheidbare Säze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik* **38**:173–98. [KBC]
Goffman, E. (2009) *Relations in public*. Transaction. [aSPLV]
Gold, J. & Gold, I. (2015) *Suspicious minds: How culture shapes madness*. Free Press. [arSPLV]
Goldman, A. I. (2006) *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford University Press. [aSPLV, SO]
Goldstein, J., Davidoff, J. & Roberson, D. (2009) Knowing color terms enhances recognition: Further evidence from English and Himba. *Journal of Experimental Child Psychology* **102**(2):219–38. [aSPLV]
Goodman, N. D. & Frank, M. C. (2016) Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences* **20**(11):818–29. http://doi.org/10.1016/j.tics.2016.08.005. [HG]
Gopnik A. & Wellman, H. M. (2012) Reconstructing constructivism: Causal models, Bayesian learning mechanisms, and the theory theory. *Psychological Bulletin* **138** (6):1085–1108. doi:10.1037/a0028044. [aSPLV, KD, MF-D]
Gowdy, J. & Krall, L. (2016) The economic origins of ultrasociality. *Behavioral and Brain Sciences* **39**:e92. [KBC]
Gozé, T., Moskalewicz, M., Schwartz, M. A., Naudin, J., Micoulaud-Franchi, J.-A. & Cermolacce, M. (2019) Reassessing "praecox feeling" in diagnostic decision making in schizophrenia: A critical review. *Schizophrenia Bulletin* **45**(5):966–70. https://doi.org/10.1093/schbul/sby172. [GD]
Grice, H. P. (1975) Logic and conversation. In *Syntax and semantics, vol. 3, speech acts,* ed. P. Cole & J. L. Morgan, pp. 41–58. Academic Press. [HG]
Griffiths, T., Chater, N., Kemp, C., Perfors, A. & Tenenbaum, J. B. (2010) Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences* **14**(8):357–64. https://doi.org/10.1016/j.tics.2010.05.004. [MF-D]
Güçlütürk, Y., Jacobs, R. H. A. H. & van Lier, R. (2016) Liking versus complexity: Decomposing the inverted U-curve. *Frontiers in Human Neuroscience* **10**:112. https://doi.org/10.3389/fnhum.2016.00112. [MF-D]
Güçlütürk, Y. & van Lier, R. (2019) Decomposing complexity preferences for music. *Frontiers in Psychology* **10**:674. https://doi.org/10.3389/fpsyg.2019.00674. [MF-D]
Gweon, H. & Asaba, M. (2017) Order matters: Children's evaluation of underinformative teachers depends on context. *Child Development* **89**(3):e278–e292. http://doi.org/10.1111/cdev.12825. [HG]
Gweon, H., Pelton, H., Konopka, J. A. & Schulz, L. E. (2014) Sins of omission: Children selectively explore when teachers are under-informative. *Cognition* **132**(3):335–41. http://doi.org/10.1016/j.cognition.2014.04.013. [HG]
Gweon, H., Shafto, P. & Schulz, L. (2018) Development of children's sensitivity to overinformativeness in learning and teaching. *Developmental Psychology* **54**(11):2113–25. http://doi.org/10.1037/dev0000580. [HG]
Gweon, H., Tenenbaum, J. B. & Schulz, L. E. (2010) Infnts consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences of the USA* **107**(20):9066–71. http://doi.org/10.1073/pnas.1003095107. [HG]
Hacking, I. (1998) *Mad travelers: Reflections on the reality of transient mental illnesses*. University of Virginia Press. [aSPLV]
Hamilton, A. F. de C. (2008) Emulation and mimicry for social interaction: A theoretical approach to imitation in autism. *Quarterly Journal of Experimental Psychology* **61** (1):101–15. https://doi.org/10.1080/17470210701508798. [aSPLV]
Han, M. J., Lin, C. H. & Song, K. T. (2013) Robotic emotional expression generation based on mood transition and personality model. *IEEE Transactions on Cybernetics* **43**(4):1290–1303. [KBC]
Hari, R. & Kujala, M. V. (2009) Brain basis of human social interaction: From concepts to brain imaging. *Physiological Reviews* **89**(2):453–79. https://doi.org/10.1152/physrev.00041.2007. [GD]
Hart, A. S., Clark, J. J. & Phillips, P. E. M. (2015) Dynamic shaping of dopamine signals during probabilistic Pavlovian conditioning. *Neurobiology of Learning and Memory* **117**:84–92. http://dx.doi.org/10.1016/j.nlm.2014.07.010. [PSS]
Heft, H. (2017) Perceptual information of "an entirely different order": The "cultural environment" in the senses considered as perceptual systems. *Ecological Psychology* **29**(2):122–45. [EB]
Heine, S. J., Takemoto, T., Moskalenko, S., Lasaleta, J. & Henrich, J. (2008) Mirrors in the head: Cultural variation in objective self-awareness. *Personality & Social Psychology Bulletin* **34**(7):879–87. [aSPLV]
Henrich, J. (2015) *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press. [aSPLV, AW]
Henrich, J. & Gil-White, F. J. (2001) The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior* **22**(3):165–96. [arSPLV]
Henrich, J. & McElreath, R. (2003) The evolution of cultural evolution. *Evolutionary Anthropology* **12**(3):123–135. [MRZ]
Hesp, C., Smith, R., Allen, M., Friston, K. & Ramstead, M. (2019) Deeply felt affect: The emergence of valence in deep active inference. https://doi.org/10.31234/osf.io/62pfd. [rSPLV]
Hewlett, B. S. (1994) *Intimate fathers: The nature and context of Aka pygmy paternal infant care*. University of Michigan Press. [aSPLV]
Hewlett, B. S. (2017) *Hunter-gatherer childhoods: Evolutionary, developmental, and cultural perspectives*. Routledge. [aSPLV]
Hewlett, B. S., Fouts, H. N., Boyette, A. H. & Hewlett, B. L. (2011) Social learning among Congo Basin hunter-gatherers. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**(1567):1168–78. http://doi.org/10.1016/j.socscimed.2009.05.016. [HG]
Heyes, C. (2018a) *Cognitive gadgets: The cultural evolution of thinking*. Belknap Press. [HG]
Heyes, C. (2018b) *Cognitive gadgets: The cultural evolution of thinking*. Harvard University Press. [aSPLV, MRZ]
Heyes, C. M. & Frith, C. D. (2014) The cultural evolution of mind reading. *Science* **344** (6190):1243091. [aSPLV]
Heylighen, F., Kingsbury, K., Lenartowicz, M., Harmsen, T. & Beigi, S. (2018) Social systems programming: Behavioral and emotional mechanisms co-opted for social control. Manuscript submitted for publication. Retrieved from http://pespmc1.vub.ac.be/Papers/SSP2mechanisms.pdf. [SVdC]
Hills, T. T. (2019) The dark side of information proliferation. *Perspectives on Psychological Science* **14**(3):323–30. [rSPLV]
Hills, T. T., Todd, P. M., Lazer, D., Redish, A. D., Couzin, I. D. & Cognitive Search Research Group (2015) Exploration versus exploitation in space, mind, and society. *Trends in Cognitive Sciences* **19**(1):46–54. [MRZ]
Hillyard, S. A., Vogel, E. K. & Luck, S. J. (1998) Sensory gain control (amplification) as a mechanism of selective attention: Electrophysiological and neuroimaging evidence. *Philosophical Transactions of the Royal Society B: Biological Sciences* **353** (1373):1257–70. [TP]
Hoemann, K., Xu, F. & Barrett, L. F. (2019) Emotion words, emotion concepts, and emotional development in children: A constructionist hypothesis. *Developmental Psychology* **55**(9):1830–49. [rSPLV]
Hogg, M. A. (2014) From uncertainty to extremism: Social categorization and identity processes. *Current Directions in Psychological Science* **23**(5):338–42. https://doi.org/10.1177/0963721414540168. [SVdC]
Hohwy, J. (2013) *The predictive mind*. Oxford University Press. [aSPLV]
Hohwy, J. (2016) The self-evidencing brain. *Noûs* **50**(2):259–85. doi:10.1111/nous.12062. [rSPLV, TP]
Hohwy, J. & Michael, J. (2017) Why should any body have a self? In: The *subject's matter: Self-consciousness and the body*, eds. F. de Vignemont & A. J. Alsmith, pp. 363–92, MIT Press. [JM]
Hong, Y., Morris, M. W., Chiu, C. & Benet-Martinez, V. (2000) Multicultural minds: A dynamic constructivist approach to culture and cognition. *American Psychologist* **55**:709–720. [GIC]
Horner, V., Whiten, A., Flynn, E. & de Waal, F. B. M. (2006) Faithful replication of foraging techniques along cultural transmission chains by chimpanzees and children. *Proceedings of the National Academy of Sciences of the United States of America* **103**:13878–83. [AW]
Howe, C. Q. & Purves, D. (2002) Range image statistics can explain the anomalous perception of length. *Proceedings of the National Academy of Sciences* **99**(20):13184–88. https://doi.org/10.1073/pnas.162474299. [MF-D]
Howe, C. Q., Yang, Z. & Purves, D. (2005) The Poggendorff illusion explained by natural scene geometry. *Proceedings of the National Academy of Sciences of the USA* **102** (21):7707–12. https://doi.org/10.1073/pnas.0502893102. [MF-D]
Howes, D. (2011) Reply to Tim Ingold. *Social Anthropology* **19**(3):318–22. [aSPLV]
Hrdy, S. B. (2011) *Mothers and others*. Harvard University Press. [arSPLV]
Huneman, P. & Machery, E. (2015) Evolutionary psychology: Issues, results, debates. In: *Handbook of evolutionary thinking in the sciences*, eds. T. Heams, P. Huneman, G. Lecointre & M. Silberstein, pp. 647–57. Springer. [aSPLV]
Huron, D. B. (2006) *Sweet anticipation: Music and the psychology of expectation*. MIT Press. [rSPLV]
Hutto, D. (2004) The limits of spectatorial folk psychology. *Mind and Language* **19** (5):548–73. [DDH]
Hutto, D. & Myin, E. (2013) *Radical enactivism: Basic minds without content*. MIT Press. [aSPLV]
Hutto, D. & Myin, E. (2017) *Evolving enactivism: Basic minds meet content*. MIT Press. [aSPLV, KD]
Hutto, D. & Satne, G. (2015) The natural origins of content. *Philosophia* **43**(3):521–36. [aSPLV]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Hutto, D. D. (2008) *Folk psychological narratives: The sociocultural basis of understanding reasons*. MIT Press. [DDH]
Hutto, D. D. (2012) *Folk psychological narratives: The sociocultural basis of understanding reasons*. MIT Press. [aSPLV]
Hutto, D. D. & Kirchhoff, M. D. (2015) Looking beyond the brain: Social neuroscience meets narrative practice. *Cognitive Systems Research* 34–35:5–17. [DDH]
Hutto, D. D., Kirchhoff, M. D. & Myin, E. (2014) Extensive enactivism: Why keep it all in? *Frontiers in Human Neuroscience* **8**:706. [aSPLV]
Huys, Q. J. M., Maia, T. V. & Frank, M. J. (2016) Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience* **19**(3):404–13. [aSPLV]
Ignatow, G. (2009) Why the sociology of morality meeds Bourdieu's *habitus*. *Sociological Inquiry* **79**:98–114. http://onlinelibrary.wiley.com/doi/10.1111/j.1475-682X.2008.00273.x/full. [aSPLV]
Ingold, T. (2001) From the transmission of representations to the education of Attention. In: *Debated Mind: Evolutionary Psychology versus Ethnography*, ed. H. Whitehouse, pp. 113–53. Berg. [aSPLV]
Ingold, T. (2016) *Lines: A brief history*. Routledge. [aSPLV]
Jack, A. I. (2014) A scientific case for conceptual dualism: The problem of consciousness and the opposing domains hypothesis. *Oxford Studies in Experimental Philosophy* **1**:1–32. [aSPLV]
Jackson, J. C., Jong, J., Bilkey, D., Whitehouse, H., Zollmann, S., McNaughton, C. & Halberstadt, J. (2018) Synchrony and physiological arousal increase cohesion and cooperation in large naturalistic groups. *Scientific Reports* **8**(1):127. [rSPLV]
Jara-Ettinger, J., Gweon, H., Schulz, L. E. & Tenenbaum, J. B. (2016) The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in Cognitive Sciences* **20**(8):589–604. http://doi.org/10.1016/j.tics.2016.05.011. [HG]
Joffily, M. & Coricelli, G. (2013) Emotional valence and the free-energy principle. *PLoS Computational Biology* **9**(6):e1003094. https://doi.org/10.1371/journal.pcbi.1003094. [aSPLV, BV]
Johnson, S. C., Dweck, C. S. & Chen, F. S. (2007) Evidence for infants' internal working models of attachment. *Psychological Science* **18**(6):501–502. [aSPLV]
Kahneman, D. (2011) *Thinking, fast and slow*. Macmillan. [aSPLV]
Kaipa, K. N., Bongard, J. C. & Meltzoff, A. N. (2010) Self discovery enables robot social cognition: Are you my teacher? *Neural Networks* **23**(8–9):1113–24. [KBC]
Kao, C.-T. & Wang, M.-Y. (2013) The right level of complexity in a Banner Ad: Roles of construal level and fluency. In: *Human interface and the management of information. Information and interaction design*, ed. S. Yamamoto, pp. 604–13. Springer, https://doi.org/10.1007/978-3-642-39209-2_67. [MF-D]
Kaplan, R. & Friston, K. J. (2018) Planning and navigation as active inference. *Biological Cybernetics* **112**:323–43. https://doi.org/10.1007/s00422-018-0753-2. [aSPLV]
Kass, R. E. & Steffey, D. (1989) Approximate Bayesian inference in conditionally independent hierarchical models (parametric empirical Bayes models). *Journal of the American Statistical Association* **84**(407):717–26. [rSPLV]
Kaufmann, L. & Clément, F. (2014) Wired for society: Cognizing pathways to society and culture. *Topoi. An International Review of Philosophy* **33**(2):459–75. [aSPLV, FC]
Keane, W. (2015) Varieties of ethical stance. In: *Four lectures on ethics: Anthropological perspectives*, eds. M. Lambek, V. Das, D. Fassin & W. Keane. Hau. [aSPLV]
Keller, G. B. & Mrsic-Flogel, T. D. (2018) Predictive processing: A canonical cortical computation. *Neuron* **100**(2):424–35. [rSPLV]
Kelly, D., Faucher, L. & Machery, E. (2010) Getting rid of racism: Assessing three proposals in light of psychological evidence. *Journal of Social Philosophy* **41**(3):293–322. [aSPLV]
Kelso, J. A. S. (1994) The informational character of self-organized coordination dynamics. *Human Movement Science* **13**(3):393–413. [rSPLV]
Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M. & Jones, P. L. (2018) Social learning strategies: Bridge-building between fields. *Trends in Cognitive Sciences* **22**(7):651–65. http://doi.org/10.1016/j.tics.2018.04.003. [AW, HG]
Kendal, R. M., Hopper, L. M., Whiten, A., Brosnan, S. F., Lambeth, S. P., Schapiro, S. J. & Hoppitt, W. (2015) Chimpanzees copy dominant and knowledgeable individuals: Implications for cultural diversity. *Evolution and Human Behavior* **36**:65–72. [AW]
Kiebel, S. J., Daunizeau, J. & Friston, K. J. (2008) A hierarchy of time-scales and the brain. *PLoS Computational Biology* **4**(11):e1000209. https://doi.org/10.1371/journal.pcbi.1000209. [aSPLV, RM]
Kiebel, S. J. & Friston, K. J. (2011) Free energy and dendritic self-organization. *Frontiers in Systems Neuroscience* **5**:80. [aSPLV]
Kinzler, K. D., Dupoux, E. & Spelke, E. S. (2007) The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America* **104**(30):12577–80. [aSPLV]
Kinzler, K. D. & Spelke, E. S. (2011) Do infants show social preferences for people differing in race? *Cognition* **119**(1):1–9. [aSPLV]
Kirchhoff, M., Parr, T., Palacios, E., Friston, K. & Kiverstein, J. (2018) The Markov blankets of life: Autonomy, active inference and the free energy principle. *Journal of the Royal Society Interface* **15**(138):1–11. 20170792. [rSPLV, TP]
Kirmayer, L. J. (1989) Cultural variations in the response to psychiatric disorders and emotional distress. *Social Science & Medicine* **29**(3):327–39. [aSPLV]
Kirmayer, L. J. (2015) Re-visioning psychiatry: Toward an ecology of mind in health and illness. In: *Re-visioning psychiatry: Cultural phenomenology, critical neuroscience and global mental health*, eds. L. J. Kirmayer, R. Lemelson & C. A. Cummings, pp. 622–60. Cambridge University Press. [aSPLV]
Kirmayer, L. J. & Crafa, D. (2014) What kind of science for psychiatry? *Frontiers in Human Neuroscience* **8**:435. https://doi.org/10.3389/fnhum.2014.00435. [DB, GD]
Kirmayer, L. J. & Gold, I. (2011) Re-socializing psychiatry: Critical neuroscience and the limits of reductionism. In: *Critical Neuroscience*, eds. S. Choudhury and J. Slaby, pp. 305–30. Wiley-Blackwell. [aSPLV]
Kirmayer, L. J., Gomez-Carrillo, A. & Veissière, S. P. L. (2017) Culture and depression in global mental health: An ecosocial approach to the phenomenology of psychiatric disorders. *Social Science and Medicine* **183**:163–68. [aSPLV]
Kirmayer, L. J., Lemelson, R. & Cummings, C. A. (2015) *Re-visioning psychiatry: Cultural phenomenology, critical neuroscience, and global mental health*. Cambridge University Press. [aSPLV]
Kirmayer, L. J. & Ramstead, M. J. D. (2017) Embodiment and enactment in cultural psychiatry. In: *Embodiment, enaction, and culture: Investigating the constitution of the shared world*, eds. C. Durt, T. Fuchs, & C. Tewes, pp. 397–422. MIT Press. [GD, aSPLV]
Kirmayer, L. J. & Sartorius, N. (2007) Cultural models and somatic syndromes. *Psychosomatic Medicine* **69**(9):832–40. https://doi.org/10.1097/PSY.0b013e31815b002c. [GD]
Kirmayer, L. J. & Young, A. (1998) Culture and somatization: Clinical, epidemiological, and ethnographic perspectives. *Psychosomatic Medicine* **60**(4):420–30. [aSPLV]
Kitayama, S., Duffy, S., Kawamura, T. & Larsen, J. T. (2003) Perceiving an object and its context in different cultures: A cultural look at new look. *Psychological Science* **14**(3):201–6. [rSPLV]
Kiverstein, J., Miller, M. & Rietveld, E. (2019) The feeling of grip: Novelty, error dynamics and the predictive brain. *Synthese* **196**(7):2847–69. https://doi.org/10.1007/s11229-017-1583-9. [arSPLV, JK]
Koelsch, S., Vuust, P. & Friston, K. (2019) Predictive processes and the peculiar case of music. *Trends in Cognitive Sciences* **23**(1):63–77. [BV]
Konvalinka, I., Xygalatas, D., Bulbulia, J., Schjødt, U., Jegindø, E. M., Wallot, S., Van Orden, G. & Roepstorff, A. (2011) Synchronized arousal between performers and related spectators in a fire-walking ritual. *Proceedings of the National Academy of Sciences* **108**(20):8514–19. [rSPLV]
Kroeber, A. L. & Kluckhohn, C. (1952) *Culture: A critical review of concepts and definitions*. Papers. Peabody Museum of Archeology & Ethnology, Harvard University. [ML]
Kulahci, I. G., Ghazanfar, A. A. & Rubenstein, D. I. (2018) Knowledgeable lemurs become more central in social networks. *Current Biology* **28**:1306–10. [AW]
Kurzban, R. & Neuberg, S. (2005) Managing ingroup and outgroup relationships. In: *The handbook of evolutionary psychology*, ed. D. M. Buss, pp. 653–75. Wiley. [aSPLV]
Laitin, D. (2007) *Nations, states, and violence*. Oxford University Press. [AB]
Lake, B. M., Ullman, T. D., Tenenbaum, J. B. & Gershman, S. J. (2018) Building machines that learn and think like people. *Behavioral and Brain Sciences* **40**:e25. [KBC]
Lakoff, G. & Johnson, M. (1980) The metaphorical structure of the human conceptual system. *Cognitive Science* **4**(2):195–208. [aSPLV]
Laland, K. N. (2018) *Darwin's unfinished symphony: How culture made the human mind*. Princeton University Press. [aSPLV]
Laland, K. N., Atton, N. & Webster, M. M. (2011) From fish to fashion: Experimental and theoretical insights into the evolution of culture. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**:958–68. [AW]
Laland, K. N. & Galef, B. G., eds. (2009) *The question of animal culture*. Harvard University Press. ISBN 9780674031265. [KBC]
Laland, K. N., Uller, T., Feldman, M. W., Sterelny, K., Müller, G. B., Moczek, A., Jablonka, E. & Odling-Smee, J. (2015) The extended evolutionary synthesis: Its structure, assumptions and predictions. *Proceedings of the Royal Society B: Biological Sciences* **282**(1813):20151019. [aSPLV]
Lamm, C., Decety, J. & Singer, T. (2011) Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. *NeuroImage* **54**:2492–502. doi:10.1016/J.NEUROIMAGE.2010.10.014. [RS]
Langton, C. G. (1990) Computation at the edge of chaos: Phase transitions and emergent computation. *Physica D* **42**:12–37. [KBC]
Larøi, F., Luhrmann, T. M., Bell, V., Christian, W. A. Jr., Deshpande, S., Fernyhough, C., Jenkins, J. & Woods, A. (2014) Culture and hallucinations: Overview and future directions. *Schizophrenia Bulletin* **40**(Suppl_4):S213–20. [ML]
Lavelle, M., Healey, P. G. T. & McCabe, R. (2014) Nonverbal behavior during face-to-face social interaction in schizophrenia: A review. *The Journal of Nervous and Mental Disease* **202**(1):47–54. https://doi.org/10.1097/NMD.0000000000000031. [GD]
Lawson, R. P., Mathys, C. & Rees, G. (2017) Adults with autism overestimate the volatility of the sensory environment. *Nature Neuroscience* **20**(9):1293–99. https://doi.org/10.1038/nn.4615. [MA]
Lebois, L. A. M., Wilson-Mendenhall, C. D., Simmons, W. K., Feldman Barrett, L. & Barsalou, L. W. (in press) Learning situated emotions. *Neuropsychologia*. [aSPLV]
Leibfried, F., Grau-Moya, J. & Braun, D. A. (2015) Signaling equilibria in sensorimotor interactions. *Cognition* **141**:73–86. https://doi.org/10.1016/j.cognition.2015.03.008. [GP]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Lelard, T., Godefroy, O., Ahmaidi, S., Krystkowiak, P. & Mouras, H. (2017) Mental simulation of painful situations has an impact on posture and psychophysiological parameters. *Frontiers in Psychology* **8**:2012. [MH]

Lelard, T., Montalan, B., Morel, M. F., Krystkowiak, P., Ahmaidi, S., Godefroy, O. & Mouras, H. (2013) Postural correlates with painful situations. *Frontiers in Human Neuroscience* **7**:4. [MH]

Lenski, R. E., Ofria, C., Collier, T. C. & Adami, C. (1999) Genome complexity, robustness and genetic interactions in digital organisms. *Nature* **400**(6745):661–64. [KBC]

Leont'ev, A. N. (1974) The problem of activity in psychology. *Soviet Psychology* **13**(2):4–33. [EB]

Levins, R. (1966) The strategy of model building in population biology. *American Scientist* **54**(4):421–31. [RLB]

Levins, R., Lewontin, R. C. (1985) *The dialectical biologist*. Harvard University Press. [DB]

Levy, R. I. (1975) *Tahitians: Mind and experience in the society islands*. University of Chicago Press. [aSPLV]

Levy, R. I. (1984) Emotion, knowing and culture. In: *Culture theory: Essays on mind, self, and emotion*, eds. R. Shweder & R. LeVine, pp. 214–37. Cambridge University Press. [aSPLV]

Libin, A. & Libin, E. (2005) Cyber-anthropology: A new study on human and technological co-evolution. *Studies in Health and Technology and Informatics* **118**:146–55. [KBC]

Lifshitz, M., van Elk, M. & Luhrmann, T. M. (2019) Absorption and spiritual experience: A review of evidence and potential mechanisms. *Consciousness and Cognition* **73**:102760. [ML]

Limanowski, J. & Blankenburg, F. (2013) Minimal self-models and the free energy principle. *Frontiers in Human Neuroscience* **7**:547. https://doi.org/10.3389/fnhum.2013.00547. [MA]

Lindley, D. V. (1956) On a measure of the information provided by an experiment. *Annals of Mathematical Statistics* **27**(4):986–1005. doi:10.1214/aoms/1177728069. [TP]

Litwin, P. & Miłkowski M. (submitted) Unification by fiat: arrested development of predictive processing. *Cognitive Science*. [RM]

Luhmann, N. (1986) The autopoiesis of social systems. In: *Sociocybernetic paradoxes, vol. 6*, eds. F. Geyer & J. van der Zouwen, pp. 172–92. Sage. Retrieved from http://cepa.info/2717. [SVdC]

Luhrmann, T. (2011) Toward an anthropological theory of mind. *Suomen Antropologi: Journal of the Finnish Anthropological Society* **36**(4):5–69. [aSPLV]

Luhrmann, T. M. (2012) *When God talks back: Understanding the American evangelical relationship with God*. Knopf. [ML]

Luhrmann, T. M., Alderson-Day, B., Bell, V., Bless, J. J., Corlett, P., Hugdahl, K., Jones, N., Larøi, F., Moseley, P., Padmavati, R. & Peters, E. (2019) Beyond trauma: A multiple pathways approach to auditory hallucinations in clinical and nonclinical populations. *Schizophrenia Bulletin* **45**(Suppl_1):S24–31. [ML]

Luhrmann, T. M. & Morgain, R. (2012) Prayer as inner sense cultivation: An attentional learning theory of spiritual experience. *Ethos* **40**(4):359–89. [ML]

Luhrmann, T. M., Padmavati, R., Tharoor, H. & Osei, A. (2015) Differences in voice-hearing experiences of people with psychosis in the USA, India and Ghana: Interview-based study. *The British Journal of Psychiatry* **206**(1):41–44. [ML]

Lumaca, M. & Baggio, G. (2017) Cultural transmission and evolution of melodic structure in multi-generational signaling games. *Artificial Life* **23**(3):406–23. [KBC]

Luo, S., Li, B., Ma, Y., Zhang, W., Rao, Y. & Han, S. (2015) Oxytocin receptor gene and racial ingroup bias in empathy-related brain activity. *NeuroImage* **110**:22–31. [aSPLV]

Luo, Y. & Baillargeon, R. (2005) Can a self-propelled box have a goal? Psychological reasoning in 5-month-old infants. *Psychological Science* **16**(8):601–608. [aSPLV]

Lutz, A., Brefczynski-Lewis, J., Johnstone, T. & Davidson, R. J. (2008) Regulation of the neural circuitry of emotion by compassion meditation: Effects of meditative expertise. *PLoS ONE* **3**(3):e1897. [aSPLV]

Mace, R. & Jordan, F.M. (2011) Macro-evolutionary studies of cultural diversity: A review of empirical studies of cultural transmission and cultural adaptation. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366**:402–411. [AB]

Machery, E. (2016) De-Freuding implicit attitudes. In: *Implicit bias and philosophy. Vol. 1. Metaphysics and epistemology*, eds. M Brownstein & J. Saul, pp. 104–29. Oxford University Press. [aSPLV]

Machery, E. & Faucher, L. (2017) Why do we think racially? Culture, evolution, and cognition. In: *Handbook of categorization in cognitive science*, 2nd edition, eds. H. Cohen and C. Lefebvre, pp. 1135–75. Elsevier. [aSPLV]

MacKay, D. J. (1992) Information-based objective functions for active data selection. *Neural Computation* **4**(4):590–604. [rSPLV]

Madoka, M. (2003) Haiku. In: *Far beyond the field: Haiku by Japanese women*, ed. M. Ueda, p. 232. Columbia University Press. [aSPLV]

Mahajan, N. & Woodward, A. (2009) Seven-month-old infants selectively reproduce the goals of animate but not inanimate agents. *Infancy* **14**(6):667–79. [aSPLV]

Malafouris, L. (2015) Metaplasticity and the primacy of material engagement. *Time and Mind* **8**(4):351–71. [aSPLV]

Malafouris, L. (2016) Material engagement and the embodied mind. In *Cognitive models in Palaeolithic archaeology*, eds. T. Wynn & F L. Coolidge, pp. 69–82. Oxford University Press, [RLB]

Mameli, M. (2001) Mindreading, mindshaping, and evolution. *Biology and Philosophy* **16**(5):595–626. [aSPLV]

Marr, D. (1982) *Vision: A computational investigation into the human representation and processing of visual information*. W. H. Freeman. [KD, GD]

Martyushev, L. M. (2018) Living systems do not minimize free energy: Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Dèsormeau Ramstead et al. *Physics of Life Reviews* **24**:40–41. https://doi.org/10.1016/j.plrev.2017.11.010. [RM]

Mathews, N., Christensen, A. L., O'Grady, R., Mondada, F. & Dorigo, M. (2017) Mergeable nervous systems for robots. *Nature Communications* **8**:439. [KBC]

Mauss, M. (1973) Techniques of the body. *Economy and Society* **2**(1):70–88. [aSPLV]

McCauley, R. N. & Henrich, J. (2006) Susceptibility to the Müller-Lyer illusion, theory-neutral observation, and the diachronic penetrability of the visual input system. *Philosophical Psychology* **19**(1):79–101. [aSPLV]

McEllin, L., Sebanz, N. & Knoblich, G. (2018) Identifying others' informative intentions from movement kinematics. *Cognition* **180**:246–58. [GP]

McElreath, R. & Henrich, J. (2007) Modeling cultural evolution. In *Oxford handbook of evolutionary psychology*, eds. R. I. M. Dunbar & L. Barrett, pp. 571–586. Oxford University Press. [MRZ]

McGeer, V. (2007) The regulative dimension of folk psychology. In: *Folk psychology re-assessed*, eds. D. D. Hutto & M. Ratcliffe, pp. 137–56. Springer. https://doi.org/10.1007/978-1-4020-5558-4_8. [aSPLV, DDH]

McLean, K. C., Lilgendahl, J. P., Fordham, C., Alpert, L., Marsden, E., & Szymanowski, K. & McAdams, D. P. (2017) Master identity development in cultural context: The role of deviating from master narratives. *Journal of Personality* **65**:1–21. [NB]

McLean, K. C. & Syed, M. (2015) Personal, master, and alternative narratives: An integrative framework for understanding identity development in context. *Human Development* **58**(6):318–49. [NB]

McShea, D. W. (2013) Machine wanting. *Studies on the History and Philosophy of Biological and Biomedical Sciences* **44**(4 pt B):679–87. [KBC]

Meltzoff, A. N. & Decety, J. (2003) What imitation tells us about social cognition: A rapprochement between developmental psychology and cognitive neuroscience. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **358**(1431):491–500. [MH]

Menary, R. (2010) The extended mind and cognitive integration. In: *The extended mind*, ed. R. Menary, pp. 267–88. MIT Press, [aSPLV]

Mercader, J., Barton, H., Gillisepie, J., Harris, S., Kuhn, S., Tyler, R., & Boesch, C. (2007) 4,300-year-old chimpanzee sites and the origins of percussive stone technology. *Proceedings of the National Academy of Sciences of the United States of America* **104**:3043–48. [AW]

Mercier, H. & Sperber, D. (2017) *The enigma of reason*. Harvard University Press. [JM, aSPLV]

Mesman, J., Van IJzendoorn, M. H. & Sagi-Schartz, A. (2016) Cross-cultural patterns of attachment: Universal and contextual dimensions. In: *Handbook of attachment: Theory, research, and clinical applications*, 3rd edition, eds. J. Cassidy & P. R. Shaver, pp. 852–77. Guilford. [PSS]

Michael, J. (2015) Cultural learning and the reliability of the intentional stance. In: *Content and consciousness 2.0: Four decades after (studies in brain and mind series*, vol. 7, ed. C. Munoz-Suárez, pp. 163–84. Springer. [JM]

Michael, J., Christensen, W. & Overgaard, S. (2014) Mindreading as social expertise. *Synthese* **191**(5):817–40. [aSPLV]

Micoulaud-Franchi, J.-A., Dumas, G., Quiles, C. & Vion-Dury, J. (2016) From clinic to the "foul and exciting field of life": A psychiatric point of view on clinical physiology. *Annales Médico-Psychologiques, Revue Psychiatrique* **175**(1):21. https://doi.org/10.1016/j.amp.2016.11.004. [GD]

Micoulaud-Franchi, J.-A., Quiles, C., Batail, J.-M., Lancon, C., Masson, M., Dumas, G. & Cermolacce, M. (2018) Making psychiatric semiology great again: A semiologic, not nosologic challenge. *L'Encephale* **44**(4):343–53. https://doi.org/10.1016/j.encep.2018.01.007. [GD]

Milgram, S. (1963) Behavioral study of obedience. *Journal of Abnormal Psychology* **67**:371–78. [aSPLV]

Millar, A. J., Urquiza, U., Freeman, P. L., Hume, A., Plotkin, G. D., Sorokina, O., Zardilis, A. & Zielinski, T. (2019) Practical steps to digital organism models, from laboratory model species to 'Crops in silico. *Journal of Experimental Botany* **70**(9):2403–418. [KBC]

Miller, C. T., Freiwald, W. A., Leopold, D. A., Mitchell, J. F., Silva, A. C. & Wang, X. (2016) Marmosets: A neuroscientific model of human social behavior. *Neuron* **90**(2):219–33. [rSPLV]

Milton, D. E. (2012) On the ontological status of autism: The "double empathy problem". *Disability & Society* **27**(6):883–87. [DB]

Miresco, M. J. & Kirmayer, L. J. (2006) The persistence of mind-brain dualism in psychiatric reasoning about clinical scenarios. *American Journal of Psychiatry* **163**(5):913–18. [aSPLV]

Mirski, R. & Gut, A. (2018) Action-based versus cognitivist perspectives on socio-cognitive development: Culture, language and social experience within the two paradigms. *Synthese* **28**(2):96. https://doi.org/10.1007/s11229-018-01976-y. [RM]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Mirza, M. B., Adams, R. A., Mathys, C. D. & Friston, K. J. (2016) Scene construction, visual foraging, and active inference. *Frontiers in Computational Neuroscience* **10**(56): 1–16. https://doi.org/10.3389/fncom.2016.00056. [arSPLV, TP]

Mithen, S. J. (1996) *The* prehistory of the *m*ind: The *c*ognitive *o*rigins of Art, *r*eligion and *s*cience. 1st paperback. Thames and Hudson. [RLB]

Miyamoto, Y., Nisbett, R. E. & Masuda, T. (2006) Culture and the physical environment: Holistic versus analytic perceptual affordances. *Psychological Science* **17**(2):113–19. https://doi.org/10.1111/j.1467-9280.2006.01673.x. [MF-D]

Montague, P. R., Berns, G. S., Cohen, J. D., McClure, S. M., Pagnoni, G., Dhamala, M., Wiest, M. C., Karpov, I., King, R. D., Apple, N. & Fisher, R. E. (2002) Hyperscanning: Simultaneous fMRI during linked social interactions. *NeuroImage* **16**(4):1159–64. https://doi.org/10.1006/nimg.2002.1150. [GD]

Montague, P. R., Dolan, R. J., Friston, K. J. & Dayan, P. (2011) Computational psychiatry. *Trends in Cognitive Sciences* 1–9. https://doi.org/10.1016/j.tics.2011.11.018. [GD]

Montague, P. R., Dolan, R. J., Friston, K. J. & Dayan, P. (2012) Computational psychiatry. *Trends in Cognitive Sciences* **16**(1):72–80. [aSPLV]

Montalan, B., Lelard, T., Godefroy, O. & Mouras, H. (2012) Behavioral investigation of the influence of social categorization on empathy for pain: A minimal group paradigm study. *Frontiers of Psychology* **3**:389. [MH]

Moore, R. (2013) Social learning and teaching in chimpanzees. *Biology and Philosophy* **28** (6):879–901. [rSPLV]

Morgan, T. J. H. & Laland, K. N. (2012) The biological bases of conformity. *Frontiers in Neuroscience* **6**:87. https://doi.org/10.3389/fnins.2012.00087. [aSPLV]

Moulin, C. & Souchay, C. (2015) An active inference and epistemic value view of metacognition. *Cognitive Neuroscience* **6**(4):221–22. [rSPLV]

Navarrete, C. D. & Fessler, D. M. T. (2005) Normative bias and adaptive challenges: A relational approach to coalitional psychology and a critique of terror management theory. *Evolutionary Psychology* **3**(1):297–325. [aSPLV]

Newen, A. (2015) Understanding others: The person-model theory. In: *Open mind: 26(T)*, eds. T. Metzinger & J. Windt, pp. 1–28. Mind Group, doi:10.15502/9783958570320. [KD]

Nichols, S. & Stich, S. (2003) Human *e*volution, *l*anguage and *m*ind: A *p*sychological and *a*rchaeological *i*nquiry. Oxford University Press. [SO]

Niedenthal, P. (2007) Embodying emotion. *Science* **316**:1002–1005. doi:10.1126/science.1136930. [RS]

Noble, W. & Davidson, I. (1996) Human *e*volution, *l*anguage and *m*ind: A *p*sychological and *a*rchaeological *i*nquiry. Cambridge University Press. [RLB]

Nordgaard, J., Sass, L. A. & Parnas, J. (2013) The psychiatric interview: Validity, structure, and subjectivity. *European Archives of Psychiatry and Clinical Neuroscience* **263** (4):353–64. https://doi.org/10.1007/s00406-012-0366-z. [GD]

Odling-Smee, J., Laland, K. N. & Feldman, M. W. (2003) *Niche construction: The neglected process in evolution*. Princeton University Press. [aSPLV]

Olson, K. R. & Spelke, E. S. (2008) Foundations of cooperation in young children. *Cognition* **108**(1):222–31. [aSPLV]

Onishi, K. H. & Baillargeon, R. (2005) Do 15-month-old infants understand false beliefs? *Science* **308**(5719):255–58. [aSPLV]

Osiurak, F. & Reynaud, E. (2020) The elephant in the room: What matters cognitively in cumulative technological culture. *Behavioral and Brain Sciences*. DOI: 10.1017/S0140525X19003236. [AW]

Oudeyer, P.-Y. & Kaplan, F. (2007) What is intrinsic motivation? A typology of computational approaches. *Frontiers in Neurorobotics* **1**:6. [aSPLV]

Overmann, K. A. (2016) Materiality and numerical cognition: A material engagement theory perspective. In: Cognitive m*odels in Palaeolithic archaeology*, eds. T. Wynn & F. L. Coolidge, pp. 42–51. Oxford University Press. [RLB]

Palacios, E. R., Isomura, T., Parr, T. & Friston, K. (2019) The emergence of synchrony in networks of mutually inferring neurons. *Scientific Reports* **9**(1):6412. doi:10.1038/s41598-019-42821-7. [TP]

Parisi, D. (1997) An artificial life approach to language. *Brain and Language* **59**(1):121–46. [KBC]

Parkinson, B. (2019) Intragroup emotion convergence: Beyond contagion and social appraisal. *Personality and Social Psychology Review*, October, 1088868319882596. [rSPLV]

Parnas, J. (2011) A disappearing heritage: The clinical core of schizophrenia. *Schizophrenia Bulletin* **37**(6):1121–30. [GD]

Parnas, J. & Zandersen, M. (2018) Self and schizophrenia: Current status and diagnostic implications. *World Psychiatry* **17**(2):220–21. https://doi.org/10.1002/wps.20528. [GD]

Parr, T., Da Costa, L. & Friston, K. (2020) Markov blankets, information geometry and stochastic thermodynamics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **378**:1–13. doi:10.1098/rsta.2019.0159. [TP]

Parr, T. & Friston, K. J. (2017a) Uncertainty, epistemics and active inference. *Journal of the Royal Society Interface* **14**(136):20170376. https://doi.org/10.1098/rsif.2017.0376. [arSPLV, GP, MA]

Parr, T. & Friston, K. J. (2017b) Working memory, attention, and salience in active inference. *Scientific Reports* **7**(1):14678. [aSPLV]

Parr, T. & Friston, K. J. (2019) Attention or salience? *Current Opinion in Psychology* **29**:1–5. https://doi.org/10.1016/j.copsyc.2018.10.006. [aSPLV, TP]

Pattee, H. H. (1977) Dynamic and linguistic modes of complex systems. *International Journal of General Systems* **3**(4):259–66. [rSPLV]

Pauker, K., Williams, A. & Steele, J. R. (2016) Children's racial categorization in context. *Child Development Perspectives* **10**(1):33–38. [aSPLV]

Pearl, J. (1998) Graphical models for probabilistic and causal reasoning. In: *Quantified representation of uncertainty and imprecision*, ed. P. Smets, pp. 367–89. Springer Netherlands. [TP]

Perreault, C., Moya, C. & Boyd, R. (2012) A Bayesian approach to the evolution of social learning. *Evolution and Human Behavior* **33**(5):449–459. [MRZ, rSPLV]

Petzschner, F. H., Weber, L. A. E., Gard, T. & Stephan, K. E. (2017) Computational psychosomatics and computational psychiatry: Toward a joint framework for differential diagnosis. *Biological Psychiatry* **82**(6):421–30. https://doi.org/10.1016/j.biopsych.2017.05.012. [MA]

Pezzulo, G. (2011) Shared representations as coordination tools for interactions. *Review of Philosophy and Psychology* **2**(2):303–33. [GP]

Pezzulo, G., Cartoni, E., Rigoli, F., Pio-Lopez, L. & Friston, K. J. (2016) Active inference, epistemic value, and vicarious trial and error. *Learning & Memory* **23**(7):322–38. [arSPLV]

Pezzulo, G. & Cisek, P. (2016) Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences* **20**(6):414–24. [aSPLV]

Pezzulo, G. & Dindo, H. (2011) What should I do next? Using shared representations to solve interaction problems. *Experimental Brain Research* **211**(3):613–30. [GP]

Pezzulo, G., Donnarumma, F. & Dindo, H. (2013) Human sensorimotor communication: A theory of signaling in online social interactions. *PLoS ONE* **8**(11):e79876. https://doi.org/10.1371/journal.pone.0079876. [GP]

Pezzulo, G., Donnarumma, F., Dindo, H., D'Ausilio, A., Konvalinka, I. & Castelfranchi, C. (2018) The body talks: Sensorimotor communication and its brain and kinematic signatures. *Physics of Life Reviews* **28**:1–21. https://doi.org/10.1016/j.plrev.2018.06.014. [GP]

Pezzulo, G., Iodice, P., Donnarumma, F., Dindo, H. & Knoblich, G. (2017) Avoiding accidents at the champagne reception: A study of joint lifting and balancing. *Psychological Science* **28**(3):338–45. [GP]

Pezzulo, G., Rigoli, F. & Friston, K. (2018) Hierarchical active inference: A theory of motivated control. *Trends in Cognitive Sciences* **22**(4):294–306. https://doi.org/10.1016/j.tics.2018.01.009. [GP]

Pezzulo, G., Rigoli, F. & Friston, K. J. (2015) Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology* **136**:17–35. [GP]

Phillips, M. L., Young, A. W., Senior, C., Brammer, M., Andrew, C., Calder, A. J., Bullmore, E. T., Perrett, D. I., Rowland, D., Williams, S. C. R., Gray, J. A. & David, A. S. (1997) A specific neural substrate for perceiving facial expressions of disgust. *Nature* **389**(6650):495–98. [aSPLV]

Pickering, M. J. & Clark, A. (2014) Getting ahead: Forward models and their place in cognitive architecture. *Trends in Cognitive Sciences* **18**(9):451–56. doi:10.1016/j.tics.2014.05.006. [KD]

Pinker, S. (1999) How the mind works. *Annals of the New York Academy of Sciences* **882**:119–27, discussion 128–34. [aSPLV]

Pinker, S. (2003) Language as an adaptation to the cognitive niche. In: *Language evolution: States of the art*, eds. S. Kirby & M. H. Christiansen, pp. 16–37. Oxford University Press. [aSPLV]

Poerio, G. L. & Smallwood, J. (2016) Daydreaming to navigate the social world: What we know, what we don't know, and why it matters. *Social and Personality Psychology Compass* **10**(11):605–18. [aSPLV]

Poirier, P., Faucher, L. & Bourdon, J.-N. (2019) Cultural blankets: Epistemological pluralism in the evolutionary epistemology of mechanisms. *Journal for General Philosophy of Science* September, 1–16. https://doi.org/10.1007/s10838-019-09472-8. [rSPLV]

Powers, A. R., Mathys, C. & Corlett, P. R. (2017) Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* **357**(6351):596–600. https://doi.org/10.1126/science.aan3458. [MA]

Price, E. E., Wood, L. A. & Whiten, A. (2017) Adaptive cultural transmission biases in children and nonhuman primates. *Infant Behavior & Development* **48**:45–53. [AW]

Provine, R. R. (2017) Laughter as an approach to vocal evolution: The bipedal theory. *Psychonomic Bulletin & Review* **24**(1):238–44. [rSPLV]

Pulcu, E. & Browning, M. (2019) The misestimation of uncertainty in affective disorders. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2019.07.007. [MA]

Pyszczynski, T., Motyl, M. & Abdollahi, A. (2009) Righteous violence: killing for God, country, freedom and justice. *Behavioral Sciences of Terrorism and Political Aggression* **1**(1):12–39. [rSPLV]

Ramsey, G. & De Block, A. (2015) Is cultural fitness hopelessly confused? *The British Journal for the Philosophy of Science* **68**(2):305–28. https://doi.org/10.1093/bjps/axv047. [SVdC]

Ramsey, W. M. (2007) *Representation reconsidered*. Cambridge University Press. [aSPLV]

Ramstead, M. J. D., Badcock, P. B. & Friston, K. J. (2018) Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews* **24**:1–16. https://doi.org/10.1016/j.plrev.2017.09.001. [arSPLV, DB, SVdC]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Ramstead, M. J. D., Constant, A., Badcock, P. B. & Friston, K. (2019a) Variational ecology and the physics of sentient systems. *Physics of Life Reviews* **31**:188–205. [aSPLV]

Ramstead, M. J. D., Kirchhoff, M. D., Constant, A. & Friston, K. J. (2019b) Multiscale integration: Beyond internalism and externalism. *Synthese* 1–30. https://doi.org/10.1007/s11229-019-02115-x. [MA]

Ramstead, M. J. D., Kirchhoff, M. D. & Friston, K. J. (2019c) A tale of two densities: Active inference is enactive inference. *Adaptive Behavior* **36**(03):181–204, 1059712319862774. https://doi.org/10.1177/1059712319862774. [MA]

Ramstead, M. J. D., Veissière, S. P. L. & Kirmayer, L. J. (2016) Cultural affordances: Scaffolding local worlds through shared intentionality and regimes of attention. *Frontiers in Psychology* **7**:1090. [arSPLV]

Ranjan, R., Logette, E., Marani, M., Herzog, M., Tâche, V., Scantamburlo, E., Buchillier, V. & Markram, H. (2019) A kinetic map of the homomeric voltage-gated potassium channel (Kv) family. *Frontiers in Cellular Neuroscience* **13**:358. [KBC]

Redcay, E. & Schilbach, L. (2019) Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience* **20**(8):495–505. https://doi.org/10.1038/s41583-019-0179-4. [DB, GD]

Rendell, L., Boyd, R., Cownden, D., Enquist, M., Eriksson, K., Feldman, M. W., Fogarty, L., Ghirlanda, S., Lillicrap, T. & Laland, K. N. (2010) Why copy others? Insights from the social learning strategies tournament. *Science* **328**(5975):208–213. [HG, MRZ]

Rendell, L., Fogarty, L., Hoppitt, W. J., Morgan, T. J., Webster, M. M. & Laland, K. N. (2011) Cognitive culture: Theoretical and empirical insights into social learning strategies. *Trends in Cognitive Sciences* **15**(2):68–76. [MRZ]

Repp, B. H. & Su, Y.-H. (2013) Sensorimotor synchronization: A review of recent research (2006–2012). *Psychonomic Bulletin & Review* **20**(3):403–52. [BV]

Richerson, P., Baldini, R., Bell, A. V., Demps, K., Frost, K., Hillis, V., Mathew, S., Newton, E. K., Naar, N., Newson, L., Ross, C., Smaldino, P. E., Waring, T. M. & Zefferman, M. (2016) Cultural group selection plays an essential role in explaining human cooperation: A sketch of the evidence. *Behavioral and Brain Sciences* **39**:e30. [MRZ]

Rietveld, E. & Brouwers, A. A. (2017) Optimal grip on affordances in architectural design practices: An ethnography. *Phenomenology and the Cognitive Sciences* **16**(3):545–64. [aSPLV]

Rietveld, E. & Kiverstein, J. (2014) A rich landscape of affordances. *Ecological Psychology* **26**(4):325–52. https://doi.org/10.1080/10407413.2014.958035. [AB, aSPLV, JK]

Rizzolatti, G., Riggio, L., Dascola, I. & Umiltá, C. (1987) Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia* **25**(1, Part 1):31–40. doi:10.1016/0028-3932(87)90041-8. [TP]

Robbins, J. (2008) On not knowing other minds: Confession, intention, and linguistic exchange in a Papua New Guinea community. *Anthropological Quarterly* **81**(2):421–29. [aSPLV]

Robbins, J., Cassaniti, J. & Luhrmann, T. M. (2011) The constitution of mind: What's in a mind? Interiority and boundedness. *Suomen Antropologi* **36**(4):15–20. [aSPLV]

Robbins, J. & Rumsey, A. (2008) Introduction: Cultural and linguistic anthropology and the opacity of other minds. *Anthropological Quarterly* **81**(2):407–20. [aSPLV]

Roepstorff, A., Niewöhner, J. & Beck, S. (2010) Enculturing brains through patterned practices. *Neural Networks* **23**(8–9):1051–59. [aSPLV]

Roesch, E. B., Nasuto, S. J. & Bishop, J. M. (2012) Emotion and anticipation in an enactive framework for cognition (response to Andy Clark). *Frontiers in Psychology* **3**:398. https://doi.org/10.3389/fpsyg.2012.00398. [BV, RM]

Rosaldo, M. Z. (1982) The things we do with words: Ilongot speech acts and speech act theory in philosophy. *Language in Society* **11**(2):203–37. [aSPLV]

Rozin, P., Haidt, J. & Fincher, K. (2009) Psychology: From oral to moral. *Science* **323**(5918):1179–80. [aSPLV]

Rumsey, A. (2013) Intersubjectivity, deception and the "opacity of other minds": Perspectives from Highland New Guinea and beyond. *Language & Communication* **33**(3):326–43. [aSPLV]

Russell, J. A. & Pratt, G. (1980) A description of the affective quality attributed to environments. *Journal of Personality and Social Psychology* **38**(2):311. [BV]

Russon, A. E., Bard, K. A. & Parker, S. T., eds. (1996) *Reaching into thought: The minds of the great apes*. Cambridge University Press. ISBN 0521644968. [KBC]

Salali, G. D., Chaudhary, N., Bouer, J., Thompson, J., Vinicius, L. & Migliano, A. B. (2019) Development of social learning and play in BaYaka hunter-gatherers of Congo. *Scientific Reports* **9**:11080. [AB]

Sarma, G. P., Lee, C. W., Portegys, T., Ghayoomie, V., Jacobs, T., Alicea, B., Cantarelli, M., Currie, M., Gerkin, R. C., Gingell, S., Gleeson, P., Gordon, R., Hasani, R. M., Idili, G., Khayrulin, S., Lung, D., Palyanov, A., Watts, M. & Larson, S. D. (2018) OpenWorm: Overview and recent advances in integrative biological simulation of *Caenorhabditis elegans*. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **373**(1758):20170382. doi:10.1098/rstb.2017.0382. [KBC]

Sass, L., Borda, J. P., Madeira, L., Pienkos, E. & Nelson, B. (2018) Varieties of self disorder: A bio-pheno-social model of schizophrenia. *Schizophrenia Bulletin* **44**(4):720–27. [GD]

Schaafsma, S. M., Pfaff, D. W., Spunt, R. P. & Adolphs, R. (2015) Deconstructing and reconstructing theory of mind. *Trends in Cognitive Sciences* **19**(2):65–72. https://doi.org/10.1016/j.tics.2014.11.007. [RLB]

Schilbach, L. (2016) Towards a second-person neuropsychiatry. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**(1686):20150081. https://doi.org/10.1098/rstb.2015.0081. [aSPLV, DB, GD]

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T. & Vogeley, K. (2013) Toward a second-person neuroscience. *Behavioral and Brain Sciences* **36**(04):393–414. https://doi.org/10.1017/S0140525X12000660. [DB, GD]

Schmidhuber, J. (2006) Developmental robotics, optimal artificial curiosity, creativity, music, and the fine arts. *Connection Science* **18**(2):173–87. [arSPLV]

Schönherr, J. & Westra, E. (2017) Beyond "interaction": How to understand social effects on social cognition. *The British Journal for the Philosophy of Science* **70**(1):27–52. [DB]

Schuppli, C. & van Schaik, C. P. (2019) Animal cultures: How we've only seen the tip of the iceberg. *Evolutionary Human Sciences* **1**:e2. [AW]

Schütz, A. (1951) Making music together: A study in social relationship. *Social Research* 76–97. [BV]

Schwartenbeck, P., Fitzgerald, T., Dolan, R. J. & Friston, K. (2013) Exploration, novelty, surprise, and free energy minimization. *Frontiers in Psychology* **4**(October):710. [rSPLV]

Schwartenbeck, P. & Friston, K. (2016) Computational phenotyping in psychiatry: A worked example. *eNeuro* **3**(4):ENEURO.0049-16.2016. [aSPLV]

Segall, M., Campbell, D. & Herskovits, M. (1966) *The influence of culture on visual perception*. Bobbs-Merrill. [MF-D]

Seligman, R. & Brown, R. A. (2009) Theory and method at the intersection of anthropology and cultural neuroscience. *Social Cognitive and Affective Neuroscience* **5**(2–3):130–37. [DB]

Seligman, R., Choudhury, S. & Kirmayer, L. J. (2016) Locating culture in the brain and in the world: From social categories to the ecology of mind. In *Handbook of cultural neuroscience*, eds. Chiao, J. Y., Li, S. C., Seligman, R., & Turner, R., pp. 3–20. Oxford University Press. [aSPLV]

Seth, A. K. (2013) Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Sciences* **17**(11):565–73. https://doi.org/10.1016/j.tics.2013.09.007. [MA]

Seth, A. K. & Friston, K. J. (2016) Active interoceptive inference and the emotional brain. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**(1708):20160007. [arSPLV, BV]

Seth, A. K. & Tsakiris, M. (2018) Being a beast machine: The somatic basis of selfhood. *Trends in Cognitive Sciences* **22**(11):969–81. https://doi.org/10.1016/j.tics.2018.08.008. [MA]

Shafto, P., Goodman, N. D. & Griffiths, T. L. (2014) A rational account of pedagogical reasoning: Teaching by, and learning from, examples. *Cognitive Psychology* **71**:55–89. http://doi.org/10.1016/j.cogpsych.2013.12.004. [HG]

Shapiro, L. (2010) *Embodied cognition*. Routledge. [aSPLV]

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C. & Frith, C. D. (2014) Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences* **18**(4):186–93. https://doi.org/10.1016/j.tics.2014.01.006. [GD]

Shelley-Tremblay, J. F. & Rosén, L. A. (1996) Attention deficit hyperactivity disorder: An evolutionary perspective. *Journal of Genetic Psychology* **157**(4):443–53. [aSPLV]

Shipp, S. (2016) Neural elements for predictive coding. *Frontiers in Psychology* **7**:1792. doi:10.3389/fpsyg.2016.01792. [MF-D, TP]

Shneidman, L., Gweon, H., Schulz, L. E. & Woodward, A. L. (2016) Learning from others and spontaneous exploration: A cross-cultural investigation. *Child Development* **87**(3):723–35. http://doi.org/10.1111/cdev.12502. [HG]

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J. & Frith, C. D. (2006) Empathic neural responses are modulated by the perceived fairness of others. *Nature* **439**(7075):466. [MH]

Smaldino, P. E. (2019) Social identity and cooperation in cultural evolution. *Behavioural Processes* **161**:108–116. [MRZ]

Smaldino, P. E. & Epstein, J. M. (2015) Social conformity despite individual preferences for distinctiveness. *Royal Society Open Science* **2**(3):140437. [MRZ]

Smaldino, P. E. & Spivey, M. J. (2019) Mills made of grist, and other interesting ideas in need of clarification [commentary on C. Heyes]. *Behavioral and Brain Sciences* **42**:e182–e182. [MRZ]

Smith, R., Alkozei, A., Killgore, W. D. S. & Lane, R. D. (2018a) Nested positive feedback loops in the maintenance of major depression: An integration and extension of previous models. *Brain, Behavior, and Immunity* **67**:374–97. doi:10.1016/j.bbi.2017.09.011. [RS]

Smith, R., Killgore, W. D. S., Alkozei, A. & Lane, R. D. (2018b) A neuro-cognitive process model of emotional intelligence. *Biological Psychology* **139**:131–51. doi:10.1016/j.biopsycho.2018.10.012. [RS]

Smith, R., Killgore, W. D. S. & Lane, R. D. (2018c) The structure of emotional experience and its relation to trait emotional awareness: A theoretical review. *Emotion* **18**(5):670–92. doi:10.1037/emo0000376. [RS]

Smith, R., Lane, R. D., Parr, T. & Friston, K. J. (2019a) Neurocomputational mechanisms underlying emotional awareness: insights afforded by deep active inference and their potential clinical relevance. *Neuroscience & Biobehavioral Reviews* **107**:473–91. [RS, rSPLV]

Smith, R., Parr, T. & Friston, K. J. (2019b) Simulating emotions: An active inference model of emotional state inference and emotion concept learning. *Frontiers in Psychology* **10**:2844. doi:10.1101/640813. [RS]

Smith, R., Weihs, K. L., Alkozei, A., Killgore, W. D. S. & Lane, R. D. (2019c) An embodied neurocomputational framework for organically integrating biopsychosocial processes: An application to the role of social support in health and disease. *Psychosomatic Medicine* **81**(2):125–45. doi:10.1097/PSY.0000000000000661. [RS]

Soenens, B., Duriez, B. & Goossens, L. (2005) Social–psychological profiles of identity styles: Attitudinal and social-cognitive correlates in late adolescence. *Journal of Adolescence* **28**(1):107–125. [NB]

Sorce, J., Emde, R., Campos, J. & Klinnert, M. (1985) Maternal emotional signaling: Its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology* **21**:195–200. [FC]

Spelke, E. S. & Kinzler, K. D. (2007) Core knowledge. *Developmental Science* **10**(1):89–96. [aSPLV]

Sperber, D. (1996) *Explaining culture: A naturalistic approach*. Wiley. [aSPLV]

Sperber, D. (1997) Intuitive and reflective beliefs. *Mind & Language* **12**(1):67–83. [aSPLV]

Stanghellini, G., Bolton, D. & Fulford, W. K. M. (2013) Person-centered psychopathology of schizophrenia: Building on Karl Jaspers' understanding of patient's attitude toward his illness. *Schizophrenia Bulletin* **39**(2):287–94. https://doi.org/10.1093/schbul/sbs154. [GD]

Stasch, R. (2009) *Society of others kinship and mourning in a West Papuan place*. University of California Press. [aSPLV]

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E. M., Breakspear, M. & Friston, K. J. (2008) Nonlinear dynamic causal models for fMRI. *NeuroImage* **42**(2):649–62. [aSPLV]

Stendel, M., Ramstead, M. J. D. & Veissiére, S. (in press) Internet sociality. In *Culture, mind, and brain: Emerging concepts, methods, and applications,* chapter 20. eds. L. J. Kirmayer, C. M. Worthman, S. Kitayama, R. Lemelson, & C. A. Cummings. Cambridge, UK: Cambridge University Press. [rSPLV]

Sterelny, K. (2007) Social intelligence, human intelligence and niche construction. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**(1480):719–30. [rSPLV]

Sterelny, K. (2011) From hominins to humans: how sapiens became behaviourally modern. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366** (1566):809–22. [rSPLV]

Sterelny, K. (2012) *The evolved apprentice*. MIT Press. [aSPLV]

Stevenson, R. J., Case, T. I., Oaten, M. J., Stafford, L. & Saluja, S. (2019) A proximal perspective on disgust. *Emotion Review* **11**(3):209–25. [rSPLV]

Stotz, K. (2017) Why developmental niche construction is not selective niche construction: And why it matters. *Interface Focus* **7**(5):20160157. [aSPLV]

Stotz, K. & Griffiths, P. E. (2017) A developmental systems account of human nature. In: *Why we disagree about human nature*, eds. T. Lewens & E. Hannon, pp. 58–75. Oxford University Press. [aSPLV]

Stout, D. (2011) Stone toolmaking and the evolution of human culture and cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences* **366** (1567):1050–59. [rSPLV]

Stout, D. & Chaminade, T. (2007) The evolutionary neuroscience of tool making. *Neuropsychologia* **45**(5):1091–100. [aSPLV]

Stout, D., Passingham, R., Frith, C., Apel, J. & Chaminade, T. (2011) Technology, expertise and social cognition in human evolution. *European Journal of Neuroscience* **33** (7):1328–38. [rSPLV]

Stout, D., Toth, N., Schick, K. & Chaminade, T. (2008) Neural correlates of early Stone Age toolmaking: Technology, language and cognition in human evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* **363** (1499):1939–49. [aSPLV]

Strand, P.S., Vossen, J.J. & Savage, E. (2019) Culture and child attachment patterns: A behavioral systems synthesis. *Perspectives on Behavior Science* **42**:835–50. doi:10.1007/s40614-019-00220-3. [PSS]

Sutton, J. (2010) Exograms and interdisciplinarity: History, the extended mind, and the civilizing process. In: *The extended mind*, ed. R. Menary, pp. 189–225. MIT Press. [aSPLV]

Swami, V., Frederick, D. A., Aavik, T., Alcalay, L., Allik, J., Anderson, D., Andrianto, S., Arora, A., Brännström, A., Cunningham, J., Danel, D., Doroszewicz, K., Forbes, G. B., Furnham, A., Greven, C. U., Halberstadt, J., Hao, S., Haubner, T., Hwang, C. S., Inman, M., Jaafar, J. L., Johansson, J., Jung, J., Keser, A., Kretzschmar, U., Lachenicht, L., Li, N. P., Locke, K., Lönnqvist, J. E., Lopez, C., Loutzenhiser, L., Maisel, N. C., McCabe, M. P., McCreary, D. R., McKibbin, W. F., Mussap, A., Neto, F., Nowell, C., Alampay, L. P., Pillai, S. K., Pokrajac-Bulian, A., Proyer, R. T., Quintelier, K., Ricciardelli, L. A., Rozmus-Wrzesinska, M., Ruch, W., Russo, T., Schütz, A., Shackelford, T. K., Shashidharan, S., Simonetti, F., Sinniah, D., Swami, M., Vandermassen, G., van Duynslaeger, M., Verkasalo, M., Voracek, M., Yee, C. K., Zhang, E. X., Zhang, X. & Zivcic-Becirevic, I. (2010) The attractive female body weight and female body dissatisfaction in 26 countries across 10 world regions: Results of the International Body Project I. *Personality & Social Psychology Bulletin* **36**(3):309–25. [aSPLV]

Swanson, J., Moyzis, R., Fossella, J., Fan, J. & Posner, M. I. (2002) Adaptationism and molecular biology: An example based on ADHD. *Behavioral and Brain Sciences* **25** (4):530–31. [aSPLV]

Tajfel, H., Billig, M. G., Bundy, R. P. & Flament, C. (1971) Social categorization and intergroup behaviour. *European Journal of Social Psychology* **1**(2):149–78. https://doi.org/10.1002/ejsp.2420010202. [MH, rSPLV]

Tal, I., *et al.* (2017) Neural entrainment to the beat: The "missing-pulse" phenomenon. Journal of Neuroscience 37(26):6331–41. [BV]

Tamariz, M. (2019) Replication and emergence in cultural transmission. *Physics of Life Reviews* **30**:47–71. https://doi.org/10.1016/j.plrev.2019.04.004. [AW]

Tauber, S., Navarro, D. J., Perfors, A. & Steyvers, M. (2017) Bayesian models of cognition revisited: Setting optimality aside and letting data drive psychological theory. *Psychological Review* **124**(4):410–41. https://doi.org/10.1037/rev0000052. [MF-D]

Taylor, C. (2016) *The language animal*. Harvard University Press. [aSPLV]

Tenenbaum, J., Kemp, C., Griffiths, T. & Goodman, N. (2011) How to grow a mind: Statistics, structure, and abstraction. *Science* **331**(6022):1279–85. https://doi.org/10.1126/science.1192788. [MF-D]

Tennie, C., Call, J. & Tomasello, M. (2009) Ratcheting up the ratchet: On the evolution of cumulative culture. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences* **364**(1528):2405–15. [rSPLV]

Thomaz, A. L. & Cakmak, M. (2013) Active social learning in humans and robots. In *Social learning theory: Phylogenetic considerations across animal, plant, and microbial taxa*, ed. K. B. Clark, pp. 113–28. Nova Science. ISBN 978-1-62618-268-4. [KBC]

Thompson, E. & Varela, F. J. (2001) Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences* **5**(10):418–25. [GD]

Timmermans, B., Schilbach, L., Pasquali, A. & Cleeremans, A. (2012) Higher order thoughts in action: Consciousness as an unconscious re-description process. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367** (1594):1412–23. https://doi.org/10.1098/rstb.2011.0421. [aSPLV]

Toelch, U. & Dolan, R. J. (2015) Informational and normative influences in conformity from a neurocomputational perspective. *Trends in Cognitive Sciences* **19**(10):579–89. [rSPLV]

Tomasello, M. (2009) *Why we cooperate*. MIT Press. [arSPLV]

Tomasello, M. (2010) *Origins of human communication*. MIT Press. [HG]

Tomasello, M. (2014) *A natural history of human thinking*. Harvard University Press. [aSPLV]

Tomasello, M. (2019) *Becoming human: A theory of ontogeny*. Belknap Press. [DB]

Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H. (2005) Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* **28**(5):675–91, discussion 691–735. [aSPLV]

Tomasello, M., Kruger, A. C. & Ratner, H. H. (1993) Cultural learning. *Behavioral and Brain Sciences* **16**(3):495–511. [NB, rSPLV]

Tomlinson, G. (2015) *A million years of music: The emergence of human modernity*. MIT Press. [RLB]

Torrance, S. & Schumann, F. (2019) The spur of the moment: What jazz improvisation tells cognitive science. *AI & Society* **34**(2):251–68. [rSPLV]

Tovo-Rodrigues, L., Rohde, L. A., Menezes, A. M. B., Polanczyk, G. V., Kieling, C., Genro, J. P., Anselmi, L. & Hutz, M. H. (2013) DRD4 rare variants in attention-deficit/hyperactivity disorder (ADHD): Further evidence from a birth cohort study. *PLoS ONE* **8** (12):e85164. [aSPLV]

Trivers, R. (2000) The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences* **907**:114–31. [aSPLV]

Tschacher, W. & Haken, H. (2007) Intentionality in non-equilibrium systems? The functional aspects of self-organized pattern formation. *New Ideas in Psychology* **25**(1):1–15. [arSPLV]

Tybur, J. M., Lieberman, D., Kurzban, R. & DeScioli, P. (2013) Disgust: Evolved function and structure. *Psychological Review* **120**(1):65–84. [aSPLV]

Ullmann-Margalit, E. (1977) *The Emergence of Norms*. Oxford University Press. [MC]

Uskul, A. K., Kitayama, S. & Nisbett, R. E. (2008) Ecocultural basis of cognition: Farmers and fishermen are more holistic than herders. *Proceedings of the National Academy of Sciences* **105**(25):8552–56. https://doi.org/10.1073/pnas.0803874105. [MF-D]

Vaish, A., Grossmann, T. & Woodward, A. (2008) Not all emotions are created equal: The negativity bias in social–emotional development. *Psychological Bulletin* **134**(3):383–403. [rSPLV]

Van de Cruys, S., Chamberlain, R. & Wagemans, J. (2017) Tuning in to art: A predictive processing account of negative emotion in art (commentary). *The Behavioral and Brain Sciences*, June **40**. http://research.gold.ac.uk/19928/. [rSPLV]

Van de Cruys, S. & Wagemans, J. (2011) Putting reward in art: A tentative prediction error account of visual art. *I-Perception* **2**(9):1035–62. [rSPLV]

Van Dijk, L. & Rietveld, E. (2017) Foregrounding sociomaterial practice in our understanding of affordances: The skilled intentionality framework. *Frontiers in Psychology* **7**:1969. doi:10.3389/fpsyg.2016.01969. [aSPLV, JK]

Veissière, S. (2016) Varieties of tulpa experiences: The hypnotic nature of human sociality, personhood, and interphenomenality. In: *Hypnosis and Meditation: Towards an Integrative Science of Conscious Planes*, eds. A. Raz & M. Lifshitz, pp. 55–76. Oxford University Press. [aSPLV]

Veissière, S. (2018) Cultural Markov blankets? Mind the other minds gap! Comment on "Answering Schrödinger's question: A free-energy formulation" by Maxwell James Désormeau Ramstead et al. *Physics of Life Reviews* **24**:47–49. [arSPLV]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press

Veissière, S. P. & Stendel, M. (2018) Hypernatural monitoring: a social rehearsal account of smartphone addiction. *Frontiers in Psychology* **9**:141. [rSPLV]

Vélez, N. & Gweon, H. (2018) Integrating incomplete information with imperfect advice. *Topics in Cognitive Science* **329**(5995):1081–17. http://doi.org/10.1111/tops.12388. [HG]

Vesper, C. & Richardson, M. J. (2014) Strategic communication and behavioral coupling in asymmetric joint action. *Experimental Brain Research* **232**(9):2945–56. https://doi.org/10.1007/s00221-014-3982-1. [GP]

Vesper, C., van der Wel, R. P., Knoblich, G. & Sebanz, N. (2011) Making oneself predictable: Reduced temporal variability facilitates joint action coordination. *Experimental Brain Research* **211**(3–4):517–30. https://doi.org/10.1007/s00221-011-2706-z. [GP]

Vickhoff, B., Åström, K., Theorell, T., von Schéele, B. & Nilsson, M. (2012) Musical piloerection. *Music and Medicine* **4**(2):82–89. [BV]

von der Lühe, T., Manera, V., Barisic, I., Becchio, C., Vogeley, K. & Schilbach, L. (2016) Interpersonal predictive coding, not action perception, is impaired in autism. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371** (1693):20150373. https://doi.org/10.1098/rstb.2015.0373. [aSPLV]

Vuust, P., Dietz, M. J., Witek, M. & Kringelbach, M. L. (2018) Now you hear it: A predictive coding model for understanding rhythmic incongruity. *Annals of the New York Academy of Sciences* **1423**(1):19–29. [rSPLV]

Vygotsky, L. S. (1930–1935/1978) *Mind in society: The development of higher psychological processes*. Harvard University Press. Original work 1930–1935. Translated in Greek, (1997) by A. Bibou and S. Vosniadou, Gutenberg. [DB]

Vygotsky, L. S. (1931/1987) The genesis of higher mental functions. In: *The history of the development of higher mental functions, vol. 4*, ed. R. Reiber, pp. 97–120. Plenum. Original work 1931. [DB]

Vygotsky, L. S. (1980) *Mind in society: The development of higher psychological processes*. Harvard University Press. [rSPLV]

Vygotsky, L. S. (1962) *Thought and language*. MIT Press. [EB]

Vygotsky, L. S. (1978) *Mind in society*. Harvard University Press. [EB]

Wagman, J. B. (2019) A guided tour of Gibson's theory of affordances. In: *Perception as information detection: Reflections on Gibson's ecological approach to visual perception*, eds. J. B. Wagman & J. J. C. Blau, pp. 130–48. Routledge. [EB]

Wallach, W., Franklin, S. & Allen C. (2010) A conceptual and computational model of moral decision making in human and artificial agents. *Topics in Cognitive Science* **2**:454–85. [KBC]

Walsh, D. (2015) *Organisms, agency, and evolution*. Cambridge University Press. [AB]

Wang, X.-J. & Krystal, J. H. (2014) Computational psychiatry. *Neuron* **84**(3):638–54. https://doi.org/10.1016/j.neuron.2014.10.018. [GD]

Wang, Y.-G., Wang, Y.-Q., Chen, S.-L., Zhu, C.-Y. & Wang, K. (2008) Theory of mind disability in major depression with or without psychotic symptoms: A componential view. *Psychiatry Research* **161**(2):153–61. [aSPLV]

Waring, T. M., Goff, S. H. & Smaldino, P. E. (2017) The coevolution of economic institutions and sustainable consumption via cultural group selection. *Ecological Economics* **131**:524–532. [MRZ]

Warnell, K. R. & Redcay, E. (2019) Minimal coherence among varied theory of mind measures in childhood and adulthood. *Cognition* **191**(October):103997. https://doi.org/10.1016/j.cognition.2019.06.009. [RLB]

Watson, S. K., Lambeth, S. P., Schapiro, S. J. & Whiten, A. (2018) Chimpanzees prioritise social information over existing behaviours in a group context but not in dyads. *Animal Cognition* **21**:407–18. [AW]

Wengrow, D. & Graeber, D. (2015) Farewell to the "childhood of man": Ritual, seasonality, and the origins of inequality. *Journal of the Royal Anthropological Institute* **21**:597–619. [AB]

Whitehead, H. & Rendell, L. (2015) *The cultural lives of whales and dolphins*. Chicago University Press. [AW]

Whiten, A. (2017a) A second inheritance system: The extension of biology through culture. *Interface Focus* **7**:20160142. [AW]

Whiten, A. (2017b) How culture extends the scope of evolutionary biology in the great apes. *Proceedings of the National Academy of Sciences of the United States of America* **114**:7790–97. [AW]

Whiten, A. (2018a) Culture and conformity shape fruit fly mating. *Science* **362**:998–99. [AW]

Whiten, A. (2018b) Social dynamics: Knowledgeable lemurs gain status. *Current Biology* **28**:R344–46. [AW]

Whiten, A. (2019a) Conformity and over-imitation: An integrative review of variant forms of hyper-reliance on social learning. *Advances in the Study of Behavior* **51**:31–75. [AW]

Whiten, A. (2019b) Cultural evolution in animals. *Annual Review of Ecology, Evolution and Systematics* **50**:27–48. [AW]

Whiten, A. (2019c) Social learning: Peering deeper into ape culture. *Current Biology* **29** (17):R845–R847. [AW]

Whiten, A. & Erdal, D. (2012) The human socio-cognitive niche and its evolutionary origins. *Philosophical Transactions of the Royal Society B: Biological Sciences* **367** (1599):2119–29. [aSPLV]

Whiten, A. & van de Waal, E. (2018) The pervasive role of social learning in primate lifetime development. *Behavioral Ecology and Sociobiology* **72**:UNSP 80. [AW]

Whiten, A. & van Schaik, C. P. (2007) The evolution of animal "cultures" and social intelligence. *Philosophical Transactions of the Royal Society B: Biological Sciences* **362**:603–20. [AW]

Wiessner, P. W. (2014) Embers of society: Firelight talk among the Ju/'hoansi Bushmen. *Proceedings of the National Academy of Sciences* **111**(39):14027–35. [rSPLV]

Wilkinson, S. & Bell, V. (2016) The representation of agents in auditory verbal hallucinations. *Mind & Language* **31**(1):104–126. [ML]

Williams, B. (2011) *Ethics and the limits of philosophy*. Taylor & Francis. [aSPLV]

Wolfram, S. (1984) Universality and complexity in cellular automata. *Physica D* **10**:1–35. [KBC]

Woodward, J. (2005) *Making things happen: A theory of causal explanation*. Oxford Studies in Philosophy of Science. Oxford University Press. [RLB]

Wright, L. T., Nancarrow, C. & Kwok, P. M. H. (2001) Food taste preferences and cultural influences on consumption. *British Food Journal* **103**(5):348–57. [aSPLV]

Xu, F. (2007) Rational statistical inference and cognitive development. In: *The innate mind. Volume 3: Foundations and the future*, eds. P. Carruthers, S. Laurence & S. Stich, pp. 199–215. Oxford University Press. [MF-D]

Yamagishi, T. & Hashimoto, H. (2016) Social niche construction. *Current Opinion in Psychology* **8**:119–24. http://dx.doi.org/10.1016/j.copsyc.2015.10.00. [PSS, rSPLV]

Zahavi, D. (2008) Simulation, projection and empathy. *Consciousness and Cognition* **17** (2):514–22. [MH]

Zahavi, D. (2014) Self and other: *Exploring subjectivity, empathy and shame*. Oxford University Press. [KD]

Zatzick, D. F. & Dimsdale J. E. (1990) Cultural variations in response to painful stimuli. *Psychosomatic Medicine* **52**(5):544–57. [aSPLV]

Zawidzki, T. W. (2008) The function of folk psychology: Mind reading or mind shaping? *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* **11**(3):193–210. [aSPLV]

Zawidzki, T. W. (2013) *Mindshaping: A new framework for understanding human social cognition*. MIT Press. [arSPLV, JM]

Zeder, M. A. (2012) The broad spectrum revolution at 40: resource diversity, intensification, and an alternative to optimal foraging explanations. *Journal of Anthropological Archaeology* **31**(3):241–64. [rSPLV]

Zefferman, M. R. (2016) Mothers teach daughters because daughters teach granddaughters: The evolution of sex-biased transmission. *Behavioral Ecology* **27**(4):1172–1181. [MRZ]

Zefferman, M. R. & Mathew, S. (2015) An evolutionary theory of large-scale human warfare: Group-structured cultural selection. *Evolutionary Anthropology* **24**(2):50–61. [MRZ]

https://doi.org/10.1017/S0140525X19001213 Published online by Cambridge University Press