Enactive-Dynamic Social Cognition and Active Inference

Abstract

The aim of this paper is twofold: it critically analyses and rejects accounts blending active inference as theory of mind and enactivism; and it advances an enactivist-dynamic account of social cognition that is compatible with active inference. While some inference models of social cognition seemingly take an enactive perspective on social cognition, they explain it as the attribution of mental states to other people, via representational machinery, in line with Theory of Mind (ToM). Holding both enactivism and ToM, we argue, entails contradiction and confusion due to two ToM assumptions rejected by enactivism: (1) that social cognition reduces to mental representation and (2) cognition must be hardwired with a social cognition contentful "toolkit" or "starter pack" for fueling the model-like theorising supposed in (1). The paper offers a positive alternative, one that avoids contradictions or confusions. After clarifying the profile of social cognition under enactivism, i.e. without assumptions (1) and (2), the last section advances an enactivist-dynamic model of cognition as dynamic, real time, fluid, dynamic, contextual social action, where we use the formalisms of dynamical systems theory to explain the origins of sociocognitive novelty in developmental change and active inference as a tool to explain social understanding as generalised synchronisation.

Keywords: social cognition, niche construction, active inference, theory of mind, enactivism, dynamical systems theory.

Introduction

Because time is continuous, and because touch and bodily experience form the first interaction with the world, cognition must be embodied. With bodily experience and action, infants first enact the world. They acquire simple motor skills such as learning how to walk, to reach, or to kick their legs. These tasks are learnt because infants have some motivation to reach a goal: getting across a room to grab a toy, for example. This motivation forces the exploration of the environment by both bodily experiencing it and learning of patterns: "infants come to acquire solutions through exploration: generating movements in various situations and feeling and seeing the consequences of those movements" (Thelen and Smith, 1996, p. 325; see also Barsalou et al. 2007; Sheya and Smith, 2019). Although the challenge is new when faced with a new task, the cognitive process of moving and perceiving is continuous in time. It is through everyday embodied actions, such as poking, squinching, banging, and so on, that the child gathers understanding about their movements in the environment.

All of this of course occurs before language and continues to exist after language. With language and eventually mastering of a reasoning toolkit humans come to conceptually articulate their bodily experience, what they perceive and body action. More precisely, humans can and do use propositional logic to describe, think, or picture their bodily experience of the world; even if the phenomenon logic works on is non-propositional, as it is totally made of bodily experience (Lakoff and Johnson, 2008; Maturana and Varela, 2012; Varela, Thompson and Rosch, 2016; Hutto and Myin, 2013, 2017; Gallagher, 2020). From this follows that meaning is present even before language. An infant is forced to develop a body skill by first being motivated to reach a goal that is meaningful to them, such as reaching a toy or hugging mum, and then bodily exploring and thereby experiencing the world. Meaning has its origins in action and it is through real time, fluid, dynamic, contextual action and activity that it is made explicit (Thelen and Smith, 1996).

In the real world, of course, infants and individuals never explore the environment on their own. From as early as birth, learning and understanding the world is social. Family and peers are as much part of the world as the physical objects they interact with. Unlike physical objects, though, humans create an intricate, dynamic network that is situated and evolving in time. A fundamental aspect of this network is that it is partly held together by meanings. Relationships with family, school, and other institutional communities, impart meanings in the same way meaning is made explicit by embodied action. Meaning is made explicit by the embodied actions of a specific community. By "meaning" we refer to the non-semantic natural attunements between organisms and their sociocultural environments whose historical situatedness across multiple spatial and temporal scales structures the current meaningful tendencies as a socioculturally skilled response.[1] The enculturation with meanings begins with the explorations of the world: as an agent explores, develops, and, eventually, masters a social environment they become enculturated, where meaning cannot be disentangled from the actions that make meaning explicit. From this standpoint, the understanding of the world then is permeated by meanings, themselves permeating action, thought, imagination, and language (Dewey, 1916; Wittgenstein; 1969; Hutto et al. 2020).

In cognitive science, social cognition aims to explain how we come to understand these meanings in others and the world. A traditional account is the computational theory of mind (Fodor, 1983; Sprevak and Colombo, 2018) on the foundations of cognitivism in cognitive science (Haugeland, 1978). Cognitivism is the position counteracting the behaviourist's *theoria non grata* of the mind as black box[2], that cognitive life comes to the computations of mental representations (Pylyshyn, 1980; Dennett, 1982; Fodor, 1983; Sprevak and Colombo, 2018; Smortchkova, Dołrega and Schlicht, 2020). So, explaining cognition is explaining how information is received, organized, stored and retrieved. For cognitivists, while cognition is a process of developing mental representations about the state of the world, social cognition is a process of developing mental representations of another person's mental state. The latter is known as mindreading approaches within Theory of Mind (ToM). Pushing against behaviourism, and especially their understanding of a newborn's mind as a "blank state", ToM suggests a hardwired social cognition module that computes mental representations about other people's beliefs, desires, intentions, emotions, etc. (Scholl and Leslie, 1999; Gerrans, 2002; Stone, and Gerrans, 2006; Wellman, 2018; Pesch, Semenov, and Carlson, 2020).

[1] This formulation is, we think, compatible with Hutto and Myin's (2017) notion of *UR-intentionality* and Kiverstein and Rietveld's (2015) notion of *skilled intentionality*.

[2] For a non-straw personned characterisation of behaviourism — one that sees in behaviourism the ambition to establish law-like relationships between mental states and behavior that dispense with any sort of mentalistic or intentional idiomand— see Alksnis and Reynolds (2021).

Embodied and enactive cognitive science rejects the cognitivist view that understanding others comes down to the ability to infer and attribute mental states in a manner somewhat hardwired from birth. According to the enactivist view, even engagements with the world that involve representational structures, such as thinking, deliberating, or planning, cannot reduce to stored mental objects in the mind of a disembodied spectator. The aim of this paper is two-fold: to dissect new accounts that blend enactivism with inferential accounts and explain why doing so involves contradiction. The second aim is to offer the only reasonable account linking enactivism and inferential accounts, specifically in the case of social cognition. While some inference models of social cognition seemingly take an enactive perspective on social cognition, they explain it as the attribution of mental states to other people via representational machinery in line with Theory of Mind (ToM). A recent account specifically making this link is Veissière et al. 's (2020) "Thinking through other minds" (TTOM). Holding both enactivism and ToM entails contradiction and confusion present in two ToM assumptions rejected by enactivism: (1) that social cognition reduces to mental representation and (2) that, at birth, individuals are equipped with an inference toolkit or starter pack for the fueling the model-like theorising supposed in (1). The last section advances an enactivist-dynamic model of cognition as dynamic, real time, fluid, dynamic, contextual social action, where we use the formalisms of dynamical systems theory to explain the origins of sociocognitive novelty in developmental change and active inference to explain social understanding as generalised synchronisation.

.

1 Thinking through other minds: "Enactive" inference?

The explanation of cognitive processes underlying enculturation aspects of life is a live issue in cognitive science (Colagè and d'Errico, 2020; Levinson and Enfield, 2020; Kirmayer et al. 2020; Hutto et al. 2020). The traditional philosophy of mind and cognitivism, comprehends the cognitive processes in general as a theoretical activity of applying or updating representations. More precisely as an information-based process that unfolds to the end of computing intelligible representations (Fodor, 1985; Millikan, 2017; Shea, 2018; Piccinini, 2018; Sprevak and Colombo, 2018; Smortchkova, Dołrega and Schlicht, 2020; Rupert, 2021). If the mark of the cognitive is representational processes, then, under this account, enculturation is expected to involve representational properties. This reasoning is widely known as "Theory of Mind" (ToM): the capacity to attribute mental states to other people in an accurate way (Scholl and Leslie, 1999; Gerrans, 2002; Stone, and Gerrans, 2006; Wellman, 2018; Pesch, Semenov, and Carlson, 2020).

Active inference is today a well-known theory of cognition that breaks up with the traditional computational orthodoxy (Parr et al. 2020; Hipólito et al. 2021; Baltieri and Buckley, 2018), and is increasingly brought to converge with enactivism insights (Constant, Clark, Friston, 2021; Korbak, 2021; Robertson and Kirchoff, 2019; Kirchhoff, 2018), although this compatibility has been questioned (Di Paolo, Thompson and Beer, 2021). Active inference is used to explain social cognition and the processes underwriting enculturation (Hesp et al. 2021; Smith, Ramstead and Kiefer, 2021; Bouizegarene et al. 2020;

Vasil et al. 2020; Veissière et al., 2020; Bolis and Schilbach, 2019; Constant et al. 2018; Gallagher and Allen, 2018).

Active inference is a modelling theory about how agents act in the environment in order to maximise their understanding, and thereby maintain a suitable state to their survival and experiential interests.[3] An adaptive system's action for the maximisation of their understanding can be translated into the minimisation of uncertainty, entropy, or surprisal.[4] As a "first principles" approach to understanding behaviour and the brain, it is framed in terms of a single imperative to minimise free energy given a generative model (Parr, Pezzulo and Friston, 2022). The Free Energy Principle (FEP) states how natural systems remain in non-equilibrium steady states by restricting themselves to a limited number of states. The evolution of systems, i.e. how a system interacts with the environment, is explained in terms of free energy minimization by the internal states of the system, by using variational Bayesian methods[5] (Da Costa et al. 2020). Internal states correspond to an open system's biomechanical dynamics: a living system (internal states), for example, is situated in an environment (external states).

The influences between internal and external states can be highlighted using a tool: Markov blankets. A Markov blanket is a scale-free statistical tool that allows us to interpret a natural system's behaviour as influences between a system and its environment. Because it is a statistical tool of dynamics and flows, it does not necessarily correspond to a physical boundary (e.g. external force in a moving pendulum), even if it sometimes does (e.g. cell exchanging energy in a tissue). A Markov blanket allows for interpreting the activity or behaviour of a system as influences between internal and external states, which indirectly influence one another via a further set of states: *active* and *sensory states*. These states, also directly influencing one another, are called *blanket* states (see fig.1).

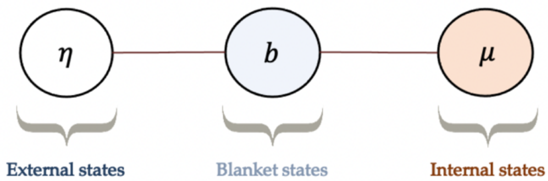

Figure 1. A Markov blanket delineates the conditionally independent internal and external states (the lines represent conditional dependencies between random variables). Considering that there is no line between μ and η, these states are conditionally independent, being indirectly influenced by blanket states comprising active and sensory states. Given its scale-free, this formalism can be applied to explain the influential flows and dynamics of any open system at any scale. (Figure reproduced from Da Costa et al. 2021).

[3] For a formal step-by-step tutorial see Smith, Friston and Whyte (2021).

[4] The term *surprisal* should not be confused with psychological surprise. It is a statistical term that refers to the "surprise" of seeing the outcome (a highly improbable outcome is very surprising). This means minimizing surprise maximizes the evidence for the agent (model).

[5] The idea of variational bayes is to construct an analytical approximation to the posterior probability of the set of unobservable variables (parameters and latent variables), given the data.

The ways in which this influence occurs is explained by supposing that internal states of a system engage in an active inference activity: that of predicting the external state. This prediction is made by a generative model i.e. a probabilistic model of how external states influence the Markov blanket that is implicit in the dynamics of internal states.

Here we arrive at an interesting philosophical bifurcation that ties up with the well-known scientific realism debate[6] in philosophy of science. It is possible to understand active inference in two ways: (1) a *realist view* that the properties of the model constructed by applying active inference tools should be also expected to exist as an ontological property in the scientific phenomenon we are trying to explain, e.g. the brain, cognitive, or cell activity, and so on; or (2) a *non-realist view* that the model used is simply an instrumental tool that, once applied to some activity, allows the scientist to draw interpretations and explanations, but the system under study does not have the properties of the model. In short, there aren't Markov blankets in the wild.[7] Elsewhere we have argued that only the latter is compatible with enactivism (**AUTHOR NAME HIDDEN**).

Veissière et al. (2020) aim to explain the processes underwriting the acquisition of culture via active inference. Taking a realist view on active inference, the authors claim that all aspects of social cognition come down to active inference. Departing from an understanding of cognition as embodied and enactive, the authors argue that individuals learn the shared habits, norms, and expectations of their culture by "thinking through other minds (TTOM)": "the process of *inferring other agents' expectations* about the world and how to behave in social context" by which "*information* from and about other people's expectations constitutes the primary domain of statistical regularities that *humans leverage to predict* and organize behaviour." (p. 1, emphasis added).[8] Veissière et al.'s (2020) argument for understanding others and the world can be formally put as follows:

> P1. Social cognition is the embodied and enactive cognitive activity for acquiring culture and understanding others.
> P2. P1 is hidden and cannot be directly grasped by the social actor.
> P3. What cannot be directly grasped must be inferred.
> Conclusion: all aspects of P1 reduce to inference.

For Veissière et al. (2020), while (P1) cognition is embodied and enactive, because (P2) all scales of social understanding are hidden, and (hidden assumption) there is information at all aspects or levels of social engagement, and (P3) what cannot be directly grasped (i.e. requires mediation by a representation) must be inferred, thus (conclusion) no doubt embodied and enactive social cognition must either be or leverage inference .

[6] See Rowbottom (2019); Agazzi (2017).
[7] For a detailed mapping of the realist *vs* instrumentalist views in the FEP, see van Es and Hipólito (2020).
[8] For a specific criticism to *Thinking through other minds* (TTOM), see Kiverstein and Rietveld (2020).

By virtue of P2 and P3 TTOM joins the ToM orthodoxy: understanding the world and others comes down to the ability to infer and attribute mental states (e.g., beliefs, desires, intentions, imagination, emotions) (Scholl and Leslie, 1999; Gerrans, 2002; Stone, and Gerrans, 2006; Wellman, 2018; Pesch, Semenov, and Carlson, 2020):

> 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. (Veissière et al., 2020, p. 2, emphasis in the original, although we would highlight the last clause).

In Veissière et al.'s (2020) theoretical model, TTOM, while cognition is understood as embodied and enactive, the social understanding of others is leveraged in mind-reading mechanisms under ToM as "the process of *inferring other agents' expectations* about the world and how to behave in social context" (p. 1). The next section critically assesses TTOM, from an enactivist point of view.

2 *Something's gotta give*: Against enactive inference through other minds

Many well-known philosophical arguments have been raised in recent literature alone by the embodied and enactive cognitive science against the mindreading ToM (Gallagher, 2001, 2006; Slors, 2010; de Bruin, Strijbos and Slors, 2011; Abramova and Slors, 2015; Hutto, 2011; Castro and Heras-Escribano, 2020; Heras-Escribano, 2020; Hipólito, Hutto and Chown, 2020; Lindblom, 2020; Heersmink, 2020; see also Menary and Gillett, 2016).

A contradiction between enactivism and ToM is found between P1 and P2 of what we laid out above as Veissière et al.'s (2020) formal argument: social cognition cannot both reduce to inference (P2) AND be embodied/enacted (P1). The contradiction between P1 and P2 results from two hidden assumptions leveraging Veissière et al.'s (2020) argument: (1) that social cognition reduces to mental representation and (2) social cognition is hardwired with an inference toolkit or starter pack for fueling the model-like theorising supposed in (1), which this section critically analyses below. Veissière et al.'s (2020) argument, laid out with its two hidden assumptions, is constructed as follows:

> P1. Social cognition is the embodied and enactive cognitive activity for acquiring culture and understanding others.
>
> P2. P1 is hidden and cannot be directly grasped by the social actor.
>
> P3. What cannot be directly grasped must be inferred.

Hidden assumption (1): social cognition reduces to information in explicit propositional form (mental representation and ascription)

Hidden assumption (2): social cognition is hardwired with the concepts and logical tools for inference.

Conclusion: all aspects of P1 reduce to inference.

In what follows we make the contradiction between premise 1 and 2 evident by critically assessing, from an enactivist perspective, what problems underlie the two assumptions and why enactivists think they should be rejected. It is worth noting that while we critically assess Veissière et al.'s (2020) TTOM formal argument, we take it as a paradigmatic case of ToM or any representationalist/cognitivist account of social cognition. Because TTOM is in perfect alignment with ToM, whatever remarks we make about TTOM will logically apply to ToM, any representationalist account of social cognition, or any account of social cognition holding assumptions (1) and/or (2).

2.1 Assumption 1: *social cognition reduces to mental representation*

Enactivism rejects the view that understanding others and the world reduces to mental representation or any form of model-like theorising (Lakoff and Johnson, 2008; Maturana and Varela, 2012; Varela, Thompsona and Rosch, 2016; Hutto and Myin, 2013, 2017; Gallagher, 2020). Because ToM-like theories, on the contrary, defend social cognition as always and everywhere a construction and ascription of a mental and/or neural representation, we find a contradiction between P1 and P2.

Across the board, in embodied and enactive cognitive science, fully enculturated agents, with conceptual and reasoning skills, engage in theorising activity. Humans can and do use propositional logic to describe, think, or picture their bodily experience of the world. They write poems, essays, measure and map things, they paint and draw how they see things, from their embodied perspective, and they also offer reasons to explain their actions. The bodily experience of the sociocultural setting is the stuff about which this theorising activity is about.

ToM-like theories suppose that all there is to social cognition is the above form of implicit or explicit theorising. In fact, this is so much so that some of the most prominent architects of ToM, explain infant development through the analogy between children and scientists: "the scientists as a child" (Bishop and Downes, 2002; Gopnik, 1996). Because the social world is hidden and mysterious, from infancy, humans ought to go around developing and testing theories to attain the most plausible explanation of the social everyday world. Social interaction thus exposed delivers a profile of social actors as if they were not active constructors of a social scene, but instead, on the outside spectators of someone else's narrative, where much is unknown and thereby requires inferring and adjudicating reasoning via the use of models, representations, and theories. Because enactivism widely rejects profiling social actors as passive inferring spectators, we reach a contradiction.

In real-time social interaction there is little to infer. Social interaction emerges from social actors *co-constructing* a social scene (the scene would not take place without them co-constructing it). Social interaction is replete with non-representational meanings that were there even before social actors were able to speak. An infant is forced to develop a body skill by first being motivated to reach a goal, such as reaching a toy or hugging mum, as something that is meaningful to them. Meaning has its origins in action and it is through real time, fluid, dynamic, contextual action and activity that it is made explicit. From this follows that meanings are present before and regardless of language. It so happens that with mastering a language, humans get to symbolically articulate their bodily, social experiences. In other words, humans get to conceptually articulate experience, i.e. explain or give reasons for the non-representational stuff they bodily experience in a social scene. But embodied non-representational meanings are regardless of language.

If these embodied meanings are non-representational, what is their profile? They emerge as a co-construction in social action. That is to say, by embodied actions within a specifically enculturated community: for example, how people respond to certain events, how they proceed from one assumption to another, how they organise word after word, the manner in which sentences are said, what reasons they give in favour of an idea, what arguments they raise in what circumstances, what they find interesting and uninteresting, and so on. Meanings are those not made explicit by language but that are grasped anyway. They emerge from our engaging in social practices and understanding others without the primacy of explicit theorising, wondering, or inferring. Meanings are, ultimately, the links holding sociocultural shared beliefs and stories together: the non-representational aspects involved in social cognition rooted upon a combination of local stories embodied in the individual practices without them being explicitly talked about.

This is so much so that individuals sharing a sociocultural background can see links between the stories that non-enculturated individuals can't. A cultural clash may result from the failure to see some culturally specific meanings by virtue of not having been enculturated in that way, launching them into a "spectator" seat. What the spectator lacks is the enculturated non explicit meanings. The spectator situation is evident in "second culture" phenomena (e.g. visiting a new culture, newly expats, refugees, etc.) (Ahmed, 2021; Taguchi, 2019). Before being specifically enculturated, they experience things from a spectator's seat. This means that, while they can, in principle, understand the reasons for enculturated practices, the space of reasons does not immediately grant the space of enculturated action. In this case the spectator must resort to theoretical activity, i.e. inference to the best explanation, where this theoretical activity is fully *permeated* by the ways in which the spectator has been otherwise enculturated. Potentially, this lack of understanding can be partly overcome by members of the community explicitly offering reasons to the spectator, i.e. conceptually articulating an explanation of the enculturation meanings the spectator fails to understand (e.g. why someone acted the way they did). But all of this occurs within the space of reasons.

The spectator does not become a (social) actor, i.e. does not leave the inference space, until they slowly and gradually start enacting these practices themselves. Confronted with a novel sociocultural setting, there is still a form of co-construction, the social actor is there operating in the same space as the locally enculturated people do, and they still participate in some sense in the practices all the while they also infer what is going on. Notably, as agents become enculturated, inference is not necessarily gone, as co-construction is inherently *negotiative*, which can take inferential forms (even if it does not need to). It is in this form that co-constructing the sociocultural niche and forms of life that agents come to be able of providing reasons for why some stories are played out as they are, as well as explain whether they are consistent or conflicting with the enculturated practice. In this respect, Anscombe (2000) remarks that individuals can justify an intention by providing reasons as to why something is done or something would be the case, as opposed to evidence for why the practice is *true*. Truthiness refers exclusively to the inference space of a spectator's logical reasoning and soundness adjudication, but not to a practice. A practice can only be either consistent or inconsistent with a cultural picture, where the enactment of a practice reinforces or modifies culturally shared meanings. Notably, culture is enacted and permeates everything that we do, including more intellectual practices such as theorising scientific and philosophical models of the world or parts of it. It is worth that this is consistent with *standpoint theory* in (feminist) philosophy of science, which says that a model of nature must representative of diverse theories, given the social and political values underlying standpoint (Halpern, 2019; Harding, 2004).

If understanding others involves enculturated standpoints and practices, social cognition cannot reduce to representational structures with truth value conditions. While representational structures may be useful when non explicit meanings fail, i.e. when someone's action is "alien" to us, meanings of the enculturated practice (the understandings that are not explicit by language) should take us a long way in our understandings and co-constructions of social scenes. As co-constructors of a social scene, social actors, from a certain enculturated standpoint, non-representational meanings are made explicit, are embodied, in everything we do. In doing so, niches are constructed as cultural niches, viz. language, rituals, beliefs, tools, and so on.

While for ToM-like theories, social cognition comes down to discovering an objective hidden world by means of engaging in inference-like modelling; for enactivists understanding others comes down to the shared non-explicit meanings, which are context-specific, modifiable, and dynamic: here lies an evident contradiction between P1 and P2. The primary issue enactivists take with ToM-like theories is that they won't be able to take social actors out of the spectator's seat. Enactivists don't think that the cultural world is mysterious, nor that culture is the *acquisition* or *transference* of mental objects. The cultural world is not hidden such that it requires understanding through intellectual achievement. Meanings are out there given, made explicit in the actions and permeating everything in between. Meanings cannot be disentangled from the enculturated practices that give rise to them: for one *cannot* simply decide not to be enculturated in a certain way. Even if one can question our enculturation structures, its shared beliefs and practices, one will be doing so from our enculturated perspective. This does not mean that non-representational meanings of our experience are not real. They are real, not in the sense of objective reality (whatever this may mean), but in that they are real experiences. Indeed, someone interacts the way they do given the very real, not hidden or mysterious, meanings explicit in the enculturated interaction. From this follows that culture is not simply *acquisition* and *transferring* of objects. Culture is enacted and thereby dynamically modifiable: a live museum preserving history, but forever reinventing itself by means of member's actions. Understanding an (enculturated) action without leaving the space of reasons will not take the social actor out of the spectator seat: for they will forever sit on the outside making inferences about things, as opposed to acting or enacting. For enactivists, cognition is enacted and embodied, where social action can involve some model-like theorising along with embodied graspings. This model-like theorising becomes more useful as a tool if one lacks the enculturation of a local community.

2.2. Assumption 2: *social cognition is hardwired from birth*

Enactivism rejects the assumption that social cognition is hardwired from birth. Within ToM's literature, there are two ways of understanding the hardwiring, both of them aligned with the Modularity of Mind, some go as far as to call it Theory of Mind *Module* (ToMM) (Gerrans, 2002), i.e. a computational system that is automatically activated, given 'social cognition stimuli', in an encapsulated manner. The first way of understanding the hardwiring is to think that the social cognition-specific module is a fixed mechanism with universal, somewhat nativist properties *a la* Fodor (1983), 'doomed' to work in a certain way given a certain stimulus, as Churchland noted (1996), i.e. a social cognition full toolkit. The second is a flexible mechanism that revises in the light of new evidence according to hardwired rules (Frith, 2019; Wellman, 2017; Scholl and Leslie, 1999), i.e. a social cognition "starter pack".

Veissière et al. (2020) do not hold a nativist position. This is made evident given their main goal to determine "how culture is *acquired*". Siding with ToM, they must hold a developmental view of ToM, in which case two challenges are in order. The first is the circular reasoning that comes from not spelling out how the starter pack is acquired (note that it is not possible to do a nativist move here). By starter pack it is meant, a flexible system whose models or representations are not universally constrained from birth but can update and upgrade given new evidence and according to hardwired rules. It is in this sense that the rules are hardwired: they are contained in a generative model with social cognition specific conceptual machinery allowing for the theorising and adjudicating of mental states to others.

Without a nativist assumption, Veissière et al. (2020) (or any ToM-like theory) need to explain how the starter pack is such that at birth new-borns understand their mother's face and gestures by means of inference. How do infants intend, move, and understand the environment by means of inference if they have not been enculturated with conceptual (i.e. developed) machinery and reasoning adjudication? Without such explanation, their theory is circular: new-borns acquire culture by inference and inference is possible by virtue of being born enculturated. In this setting, the question of how humans become enculturated remains unanswered in the shadows of nativism. What needs to be explained, without a nativist assumption, is the origins of novelty in developmental change. We will answer this question in the next section.

As developing organisms perceive and act in daily life, there must be continuity between these activities and changes over a long-time scale. No one denies the contribution of the nervous system, the hormonal system, and the genes (and so on) to human behaviour. But it would be a serious mistake to limit the contributors to those inside the biological system and exclude contributors from outside the organism, such as everyday features of the physical and social environment. Turning things on its head, the question then is how behaviour arises from a multitude of underlying contributing elements. How do these pieces come together as a whole?

ToM-like theories take it that social cognition comes down to inference, where it is not spelled out how individuals come equipped with the tools for inference. This is a problem largely diagnosed by enactivists (Gallagher, 2001, 2006; Slors, 2010; de Bruin, Strijbos and Slors, 2011; Abramova and Slors, 2015; Hutto, 2011; Fernández-Castro and Heras-Escribano, 2020; Hipólito, Hutto and Chown, 2020; Lindblom, 2020). It is thereby with surprise that we see Veissière et al.'s (2020) TTOM aligning with enactivism, as they say, "cognition as an embodied, enactive, affective process involving cultural affordances" (p. 1).

It is also with surprise that we see Veissière et al.'s (2020) TTOM aligning with Dynamical Systems Theory (DST). Veissière et al.'s (2020) claim that their TTOM:

> seeks to resolve key debates in current cognitive science, such as . . .the more fundamental distinction between *dynamical and representational accounts of enactivism*." (p. 1, emphasis added).

But this cannot be the case. It is widely known that DST categorically rejects the notion of representation or cognition as information processing (Favela, 2020). From the classics, we have the insight: "rather than computation, cognitive processes may be dynamical systems; rather than computation, cognitive processes may be state-space evolution within these very different kinds of systems" (van Gelder, 1995, p. 346).

If social cognition is a dynamical system, then social cognition is not hardwired from birth, nor in the form of representational content (nativist ToM), nor in the form of representational rules (developmental ToM). Social cognition processes are not computational, but a state-space evolution that is made explicit in the form of the niches constructed by communities in particular and the human community as a whole. From this follows that enculturation processes cannot be conceived of as the *acquisition* and *communication* of static mental objects, but instead as an enactment of the dynamics of a temporally situated social scene. In fact, dynamic theories of social cognition clearly state that:

> our commitment to a biologically consistent theory means that we *categorically reject machine analogies of cognition* and development . . . the brain may well share certain operations with a digital computer, but it is different from a machine on the most fundamental thermodynamic level. . . a developmental theory must be appropriate to the organism it serves; thus, *we deliberately eschew the machine vocabulary of processing devices, programs, storage units, schemata, modules, or wiring diagrams*. We substitute, instead, a vocabulary suited to fluid, organic systems, with certain thermodynamic properties" (Thelen and Smith, 1996, p. Xix, emphasis added).

In conclusion, because enactivism categorically rejects any form of hardwired computations, they are in clear contradiction with ToM-like theories. Because DST categorically rejects the analogy between cognition and a computer and machinery vocabulary, ToM-like theories have nothing to offer DST. By the same token Veissière et al.'s (2020) TTOM does not resolve any "key debates in current cognitive science, such as . . .the more fundamental distinction between *dynamical and representational accounts of enactivism"* (p. 1). On the contrary, it brings unnecessary confusion holding upon contradiction.

In what follows, we present an enactivist-dynamic explanation of how we understand others and the world that, while consistent with the description above – of fluid, organic systems, with certain thermodynamic properties – answers questions about the origins of sociocognitive novelty in developmental change.

3 Into the dynamics of social understanding

In the previous section we have critically assessed the incompatibility between ToM-like theories and enactivism. In doing so we rehearsed and laid out the main features of an enactivist social cognition profile. More precisely, we characterised the activity of understanding others as an activity that is not reducible to mental representations nor hardwired from birth. We explain that in understanding others we engage in forms of niche construction.

In order to offer a cohesive account, in this last section we indicate and explain the experimental tool that we think is most suitable for the enactive framework of social cognition laid out above. DST is an approach that serves to evaluate the behaviour of both abstract and physical systems as situated in and changing over time (Hirsch, 2020; van den Bosch and van der Klauw, 2020). Despite recent hype, DST is not new. In fact, its computational machinery, such as network analyses, agent-based modelling, dynamical causal modelling, or differential equations have facilitated some of cognitive science's most significant early achievements (Favela 2020) to study diverse cognitive functions (Holmes, 2020; Barfuss, 2021; Tschacher, 2021; Han and Amon, 2021), as well as neural activity in neuroimaging studies (Friston et al. 2019).

As a formalism, DST is useful to computationally study and understand cognitive behaviour for one major reason: it does not require a realist attitude about the (computational) models used to simulate a behaviour of scientific interest. That is, while DST offers mathematics as well as the computational machinery to simulate complex behaviour (such as cognitive behaviour) that would otherwise not be possible to study, it does so without supposing that the physical system under scrutiny ontologically entails, involves, or leverages the computational machinery used in the simulation model. This is precisely van Gelder's insight in his seminal 1995 paper, asking "what could cognition be if not computation?". For him, cognitive behaviour is not computation but it is what is observable: viz. that cognitive behaviour is *situated* and *changes* in time: it is dynamical. DST is an epistemic instrument in the sense that those using it refrain from making ontological assumptions about the phenomenon being studied on the basis of the computational machinery used to understand a physical system.

Organism-environment systems are complex systems. Computing the dynamics of the entire system involves too many variables and interrelations for it to be tractable. As such, how to compute on a low-level — in a mathematically and computationally tractable manner — the activity that has been generated in a high-level space (e.g. the cognitive behaviour or activity) is a typical computational problem posed in computational neuroscience. Faced with this issue, a common procedure in mainstream computational neuroscience is to adopt something called *dimensionality reduction* (DeMers and Cottrell, 1993; Beyeler et al. 2019; Tanisaro and Heidemann, 2019; Reddy et al 2020). Dimensionality reduction is an approximation or optimisation procedure that involves representing in a low dimension, i.e. a model, some meaningful properties of the data collected from the activity of interest. The data collected *lives* in high dimension because it has been generated by a complex system. A complex system is a system with high degrees of freedom, i.e. a system with a set of variables so vast that it is not mathematically tractable nor can it be computationally simulated (Garey and Johnson, 1979; Rich et al. 2020). So, the common procedure is dimensionality reduction. This procedure comes with a cost, a Laplace assumption, which "assumes a *fixed Gaussian form* for the conditional density of the parameters" (Friston, 2007, p. 220, emphasis added). This means that it assumes that the local interacting parts generating behaviour do not interact in a dynamical manner. Although this is a useful and insightful move, ubiquitous in Machine Learning and Variational Bayes[9], it is important to highlight that it is simply an instrumental move. That is, a simplification of a complex system so as to enable us to create dynamical models, and thereby make the complexity tractable.

While this is simply an instrumental move, often we are given the impression that the complex system under study is instead static, i.e. its parts do not interact in a dynamical manner (just like the instrument we used). Mainstream cognitive science commonly refrains from using change *per se*. This is either because of the problem of intractability (as explained above), or a philosophical standpoint, for example, that the world is static and hidden and therefore, its exploration and construction depend upon representing it objectively under accuracy conditions. But static systems are not dynamical systems. One way to represent a static relationship is as follows:

$$y\mathrm{i} = f(x\mathrm{i}), \qquad (1)$$

What the equation says is that, with y as a dependent variable and x as an independent variable, for any possible value of $x1$, a corresponding value will be generated for the dependent variable y. In short, the equation describes a static system of a particular value of the variable as a function of the value of another variable or a set of such variables (we critically analyse this below). A static system or model, by definition, will generate predictions without any reference to recursiveness. Making ontological claims from a static model would mean to say that the physical system, just like the model, is static, linear, and can be understood as if it were isolated in time and space. To put it otherwise, using the above model in a realist sense would entail claiming that cognitive behaviour is not situated and embedded in a dynamical environment: a ToM-like mechanism where certain social theory or model activate given a certain stimulus.

9 For niche construction see Constant et al. (2018); for ecology and sentient systems see Ramstead et al. (2019).

DST takes its instruments as epistemic instruments, not ontological predictors. Understanding cognitive behaviour, including its maturation and enculturation, through a model inherently means to simplify it. Yet this simplification must conserve the system's characterising features, one of which is *complexity*. An organism situated in its environment is an ensemble of many closely interacting, interdependent components, whose activity is more than the sum of the parts of the components — known as *nonlinearity*. Because the system's structure and organisation results from the interaction between parts, the system is self-organised. Notably, although the situated organism is constantly changing, it maintains coherence over time, i.e. it is a complex system (Phelan, 2001; Ay et al. 2011; De Domenico et al. 2019; López-Ruiz, 2021).

DST captures the system's characterising feature: complexity. It departs from the observation that things change. Phrased more radically, it makes a key assumption "that there is only process" (Thelen and Smith 1994, p. 39). Models in DST retain this specific characteristic of complex systems: change. As defined by Weisstein (1999): a dynamical model is "a means of describing how one state develops into another state over the course of time," which can be expressed mathematically as

$$y\text{t}+1 = f\,(y\text{t}), \qquad (2)$$

expressing that the next state (at time $t + 1$) is a function, f, of the preceding state, at time t. In a slightly different notation:

$$y\ /\ t = f\,(y), \qquad (3)$$

stating that the change of a system, denoted by y, over some amount of time, denoted by t, is a function f of the state of y. The function f is also referred to as the dynamical rule. It is important to note that f specifies some causal principle of change and that the current equation depicts recursive relationships (i.e. yt leads to yt+1, and accordingly, yt+1 generates yt+2 and so on).

Applying a DST model to enculturation aspects of cognition, for example, how a child comes to develop a conceptual kit (since this is not given from birth), i.e. a child's growing conceptual toolkit, we obtain the following. The equation describes the current state as a function of a preceding state in a recursive way. This means taking the result of step one in the process (conceptual toolkit today) as the starting value generating the next step (the conceptual toolkit tomorrow). f corresponds to the principle of change such that the learning of new concepts at time t depends on the concepts already known and the environment the child is situated at (e.g. the people with whom the child communicates at a time t) (van Geert, 2009; for recent dynamic approaches to education and learning see van Dijk, 2020; Kaplan and Garner, 2020; Koopmans, 2020). This recursiveness illustrates the *enaction*, i.e. the processes that happen "between one behavioural moment and the next" (Varela, 1992, p. 106; see also Di Paolo et al., 2017; and Di Paolo et al.,

2021) — which is also characteristic of the dynamical systems approach.[10] An individual's conceptual toolkit is a niche construction process itself: the language we speak, the conversation styles favoured in specific groups, the uses we give to them, i.e. how we use them to articulate our and others' practices is niche construction. Acquiring the abilities to understand and respond to the links between spoken and written patterns, we contribute to niche construction in real time. In languaging we participate in what constitutes a way of living as a human (Wittgenstein, 1953; Hintikka, 1979; Moyal-Sharrock, 2021). After all, "we are linguistic/discursive beings and not merely animals with an evolved capacity for language" (Rouse, 2015, p. 77). This can be frustratingly difficult in our language permeated environment, especially when we find ourselves learning a second language (second language acquisition, or SLA). On the matter, Soleimani (2013) argues that the "Newtonian conceptualization of SLA research cannot be comprehensive to deal with the complexities of language acquisition research", and therefore applies a dynamical systems approach. Languaging is pervasive in that it remains connected to other forms of engagement with the environment: it involves complex perceptual and practical capacities. Because linguistic exchanges are directed, responsive, and accountable to our environmental circumstances, language and languaging are better understood as self-organising dynamical systems (Hohenberger, 2011). In line with this, Elman (1995) explains language not as rule-governed, i.e. 'operations on symbols', but rather embedded in the dynamics of the system permitting movement from certain regions to others, i.e. navigating the situated environment where languaging happens (see also Patriarca et al. 2020).[11] Importantly, on a DST account, not even languaging is understood in terms of mental representation.

Because languaging is always situated within a wider practical and perceptual context, linguistic capacities are open and incorporate other sensorimotor/cognitive capacities. In this regard, evidence from Nölle et al. (2020) confirms that "subtle environmental motivations drive the emergence of different communicative conventions in an otherwise identical task, suggesting that linguistic adaptations are highly sensitive to factors of the shared task environment." Moreover, the authors speculate that "local interactional level, through processes of cultural evolution, contribute to the systematic global variation observed among different languages" (p. 1). Linguistic articulation, as an enculturated practice, thereby contributes to the material manipulations that further shape the niche we find ourselves in, i.e. the self-producing process networks of the society, or the long history of niche constructive activities we live in.

Dynamical models answer questions about culture acquisition without assuming that cognition is computational. Static models (eq. 1), by assuming that associations between variables across a sample can be used as valid approximations of the dynamic relations given the enaction, tell us very little about enculturation aspects of cognition: viz. how individuals come to explore and adapt, navigate, and socially

---

[10] It may look as though the notion of 'behavioral moment' is at odds with the dynamicist claim that 'there is only process' cited above. In the latter, 'moments' as such, do not exist: there is only movement, and movement is inherently durational, whereas moments are inherently durationless. Yet we can conceive of behavioral moments in abstract terms, as a manner of describing a time-slice in what is essentially a moving process.

[11] For another recent dynamic understanding of language see Müller-Frommeyer et al. (2020).

engage with their environments from one behavioural moment to the next. Dynamic models, for example, dynamical causal modelling, on the contrary, offer the tools to explain niche construction as the behaviour generated within the reciprocity between the environment and the organism, such that the specific way that an organism behaves does not exist without the specific way that the environment is and again *vice versa*.

4 Active Inference in an Enactive-Dynamic Setting

In order to explain how we understand others, it is necessary to highlight the changes involved in the unfolding activity or event in which people are participating, where events and activities are essentially enacted and dynamic. An enactive-dynamic account of cognition posits mental life as an emergence of activities in everyday life. It provides the biological ground for a cultural and contextual account of how humans understand others and the world. Culture permeates everyday life, where the shared non-representational aspects of that culture permeate how we understand others.

Active inference can be an insightful instrument to describe the dynamics underlying cultural co-construction, considering that it does not necessarily take agents as passive, the agents together are the authors of the states of one another (Gallagher & Allen, 2016). As mentioned in section 1, active inference is a tool that can be applied for explaining the behaviour of a complex, dynamic system by supposing that the internal states of a system predict the external states. In coupled systems, say for example two social actors, this would translate into each one predicting each other such that they *synchronise* their generative models. (While this has been taken in the realist sense by TTOM, in an enactive framework the non-realist sense is the correct one to take).

The social scene has been described as generalised synchronisation (Friston and Frith, 2015). That is, by supposing that each social actor behaves *as if* they knew the hidden states of the other. To put it more precisely, each actor behaves *as if* they have generative models that end up synchronising. This characterisation is still not compatible with enactivism nor dynamical systems theory. In social cognition conceived of as the prediction of hidden states for epistemic purposes, social actors are stuck in the spectator seat.

The subtle but crucial insight is that not all states are hidden in the social scene. They are hidden to the scientist modelling a social scene: for the scientist is the spectator in the social situation. When two social actors come to co-construct a social interaction, meanings emerge, which in turn allows each actor to directly understand the other actor's behaviour. The reason a social interaction can be described *as if* each social actor understands each other's states is because, in fact, they do grasp some or most meanings.

Typically, social interaction, under inference of ToM-like theories, is explained as having the following form:

Inference/prediction of each other's states → generative models → social understanding scene

The above form of characterisation, however, corresponds to that adopted by the scientist. It is the scientist that, while developing a model, starts by making assumptions, i.e. predictions, constructing a generative model that is common to the social agents interacting – e.g. general synchronisation of generative models in Friston and Frith, 2015 –, to attain the aim of explaining how they understand each other. The form characterised above refers to that of a spectator seat.

Crucially, the standpoint taken by the scientist is one that is different to that taken by social actors. Scientists constructing models are spectators, while agents interacting are actors. From this follows that the characterisation of the form of interaction must be distinct to the one pertaining to passive spectating. The co-construction of a social scene has thereby a different profile that can be described and understood as follows:

Social understanding scene → generative models → inference/prediction of each other's states

In the above profile, social interaction is not caused by prediction. Departing from the scientific observation that there is social understanding from social actors co-constructing a social scene, we can then scientifically characterise this social understanding in the terms of generative models, or, to put it even more in line with the active inference literature, in terms of a general synchronisation of generative models, which then will allow us to explain and make predictions about how social agents behave and social cognition as a cognitive function. Interacting social actors *give the impression that they are running the generative model*, which grants epistemic value to the generative model as tool. Note however, that the epistemic value of the model does not grant that the tools used in the model should also be ontological properties of the system being modelled (i.e. the social interaction).

The non-sensical move from epistemic value to ontology is made evident with analogy to other coupling, dynamic systems such as pendulums. Few would say that pendulums actively infer each other's states, and yet moving pendulums will, at some point, synchronize their activity (Francke et al. 2020). They do so, obviously, without literally inferring each other's states. This occurs as complex behaviour emerging from the interaction between relatively simple systems (Wolfram, 2018; Rihani, 2002), which is known as stochastic resonance (Lucarini, 2019; Gammaitoni et al. 1998). Yet they behave as if they *knew* each other's states. Other examples include synchronised neurons, individually interacting as if they *knew* their peer's behaviour, when they could not possibly (Protachevicz et al. 2021; Balconi and Fronda, 2020; Friston and Frith, 2015; Lewis et al. 2004).

The reason for this is that it is not prediction/inference that causes the behaviour of pendulums or neurons. Prediction or inference is the tool we use to make sense of the dynamics of synchronized pendulums or neurons. Likewise, it is not prediction that causes the social understanding between social actors, prediction or inference is the tool we use to make sense of the dynamics of a social scene. In short, a social scene, as well as synchronized pendulums or neurons, are real world, complex phenomena, the stuff our scientific models can be applied to make real world, complex phenomena intelligible, even if complex phenomena is not representational in nature.

Real-world dynamic behaviour	Scientific tools
Synchronised pendulums, neurons, people Collective intelligence (families, crowds, communities, swams, flocks of birds)	Dynamical and complex systems theory Active inference Free Energy Principle

It is with this causal order in mind and an instrumentalist perspective that active inference is useful to explain cognitive natural and life behaviour in general and social cognition in particular. Active inference, as a corollary of the FEP, provides the mathematical and conceptual tools that can be applied to understand real world dynamical systems. It can be applied in two ways: (1) to build up scientific models of highly complex phenomena. Variational free energy, as an information theoretical function, can be applied to solve for optimization of model evidence. This allows for model comparison analysis. And it can also be applied to offer insights over (2) the behaviour of self-organizing systems. Markov blankets allow us to interpret the social interaction as the meaningful influences between social actors co-constructing the social environment. This is possible because Markov blankets are a statistical tool that measures influences and does not necessarily correspond to a physical boundary. In a dynamical setting, they can highlight the synchronisation of pendulums or social understanding. Friston et al. (2021) have advanced the formalisms that allow us to think of the coupling between internal and external states in terms of generalised synchrony of chaos on the interface between physical and life sciences.

Notably, none of these techniques (1) or (2), under dynamical systems theory and non-equilibrium statistics, in itself, entails the realist claim that the targets use, leverage, or are the models by which they are described or explained. The FEP applies to all natural systems, or in other terms to all systems that self-organize to maintain non-equilibrium steady states; to take a realist attitude about embodied generative models would mean that all natural systems, adaptive and non-adaptive, living and non-living, behave in the ways they do by leveraging a generative model (which they don't).

Conclusion

This paper had a twofold aim: to dissect new accounts that blend enactivism with inferential accounts and explain why doing so involves contradiction. We then offered the only reasonable link between enactivism and inferential accounts, i.e. one that does not involve ToM-like assumptions, specifically in the case of social cognition. While some inference models of social cognition seemingly take an enactive perspective on social cognition, they explain it as the attribution of mental states to other people, via representational machinery, in line with Theory of Mind (ToM). We have shown that holding both enactivism and ToM entails contradiction and confusion. This is evident when we critically dissect the two hidden assumptions held by ToM-like theories such as TTOM, which are clearly rejected by enactivism: (1) that social cognition reduces to mental representation and (2) that, at birth, individuals are equipped with an inference toolkit or starter pack for fueling the model-like theorising supposed in (1). In our critical assessment we rehearsed and laid out the main features of an enactivist social cognitive profile: cognition is enacted and embodied, where social action can involve some model-like theorising if and when embodied understanding is lacking. As co-constructors of a social scene, social actors, from a certain enculturated standpoint, non-representational meanings are made explicit in everything we do. Enaction is the process that happens between one behavioural movement and the next. The formalisms of dynamical systems theory further explain the origins of sociocognitive novelty in developmental change, and active inference is a suitable, complementary tool to explain the social understanding as generalised synchronisation observed in natural and life sciences.

References

Ahmed, M. A. (2021). Cross-Cultural Adjustment and Second Language Acquisition. *International Journal of Language and Literary Studies*, *3*(2), 290-300.

Alksnis, N., & Reynolds, J. (2021). Revaluing the behaviorist ghost in enactivism and embodied cognition. *Synthese*, *198*(6), 5785-5807.

Barsalou, L. W., Breazeal, C., & Smith, L. B. (2007). Cognition as coordinated non-cognition. *Cognitive Processing*, *8*(2), 79-91.

Bishop, M. A., & Downes, S. M. (2002). The theory theory thrice over: the child as scientist, Superscientist or social institution?. *Studies in History and Philosophy of Science Part A*, *33*(1), 117-132.

Bolis, D., & Schilbach, L. (2019). 'Through others we become ourselves': The dialectics of predictive coding and active inference.

Bouizegarene, N., Ramstead, M., Constant, A., Friston, K., & Kirmayer, L. (2020). Narrative as active inference.

Churchland, P. M. (1996). *The engine of reason, the seat of the soul: A philosophical journey into the brain*. mit Press.

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

Da Costa, L., Friston, K., Heins, C., & Pavliotis, G. A. (2021). Bayesian mechanics for stationary processes. *arXiv preprint arXiv:2106.13830*.

Dennett, D. C. (1982, January). Styles of mental representation. In *Proceedings of the Aristotelian society* (Vol. 83, pp. 213-226). Aristotelian Society, Wiley.

Di Paolo, E., Thompson, E., & Beer, R. D. (2021). Laying down a forking path: Incompatibilities between enaction and the free energy principle.

Friston, K., & Frith, C. (2015). A duet for one. *Consciousness and cognition*, *36*, 390-405.

Friston, K. J., Preller, K. H., Mathys, C., Cagnan, H., Heinzle, J., Razi, A., & Zeidman, P. (2019). Dynamic causal modelling revisited. *Neuroimage*, *199*, 730-744.

Frith, C. (2019). Theory of mind in schizophrenia. In *The neuropsychology of schizophrenia* (pp. 147-161). Psychology Press.

Gallagher, S., & Allen, M. (2018). Active inference, enactivism and the hermeneutics of social cognition. *Synthese*, *195*(6), 2627-2648.

Gerrans, P. (2002). The theory of mind module in evolutionary psychology. *Biology and Philosophy*, *17*(3), 305-321.

Gopnik, A. (1996). The scientist as child. *Philosophy of science*, *63*(4), 485-514.

Halpern, M. (2019). Feminist standpoint theory and science communication. *Journal of Science Communication*, *18*(4), C02.

Harding, S. G. (Ed.). (2004). *The feminist standpoint theory reader: Intellectual and political controversies*. Psychology Press.

Haugeland, J. (1978). The nature and plausibility of cognitivism. *Behavioral and Brain Sciences*, *1*(2), 215-226.

Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., & Ramstead, M. J. (2021). Deeply felt affect: The emergence of valence in deep active inference. *Neural computation*, *33*(2), 398-446.

Hipólito, I. Hutto, D., Chown, N. (2020) Understanding Autistic Individuals: Cognitive Diversity not Theoretical Deficit. in *Neurodiversity Studies: A New Critical Paradigm*. Routledge.

Hutto, D., Gallagher, S., and Ilundain-Agurruza, J. Hipólito, I. (2020). Culture in Mind - An Enactivist Account: Not Cognitive Penetration But Cultural Permeation.In L. J. Kirmayer, S. Kitayama, C. M. Worthman, R. Lemelson, & C. A. Cummings (Eds.), *Culture, mind, and brain: Emerging concepts, models, applications*. New York, NY: Cambridge University Press.

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), 20170792.

Kiverstein, J., & Rietveld, E. (2015). The primacy of skilled intentionality: on Hutto & Satne's the natural origins of content. *Philosophia*, *43*(3), 701-721.

Korbak, T. (2021). Computational enactivism under the free energy principle. *Synthese*, *198*(3), 2743-2763.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.

Maturana, H. R., & Varela, F. J. (2012). *Autopoiesis and cognition: The realization of the living* (Vol. 42). Springer Science & Business Media.

Pylyshyn, Z. W. (1980). Computation and cognition: Issues in the foundations of cognitive science. *Behavioral and Brain Sciences*, *3*(1), 111-132.

Robertson, I., & Kirchoff, M. D. (2019). Anticipatory action: Active inference in embodied cognitive activity. *Journal of Consciousness Studies*, *27*(3-4), 38-68.

Scholl, B. J., & Leslie, A. M. (1999). Modularity, development and 'theory of mind'. *Mind & language*, *14*(1), 131-153.

Sheya, A., & Smith, L. (2019). Development weaves brains, bodies and environments into cognition. *Language, cognition and neuroscience*, *34*(10), 1266-1273.

Smith, R., Friston, K., & Whyte, C. (2021). A step-by-step tutorial on active inference and its application to empirical data.

Smith, R., Ramstead, M. J., & Kiefer, A. (2021). Active inference models do not contradict folk psychology.

Smortchkova, J., Dołrega, K., & Schlicht, T. (Eds.). (2020). *What are Mental Representations?*. Oxford University Press.

Sprevak, M., & Colombo, M. (Eds.). (2018). *The Routledge Handbook of the Computational Mind*. Routledge.

Taguchi, N. (Ed.). (2019). *The Routledge handbook of second language acquisition and pragmatics*. Routledge.

Thelen, E., & Smith, L. B. (1996). *A dynamic systems approach to the development of cognition and action*. MIT press.

Varela, F. J., Thompson, E., & Rosch, E. (2016). *The embodied mind: Cognitive science and human experience*. MIT press.

Vasil, J., Badcock, P. B., Constant, A., Friston, K., & Ramstead, M. J. (2020). A world unto itself: Human communication as active inference. *Frontiers in psychology*, *11*, 417.

Wellman, H. M. (2017). The development of theory of mind: Historical reflections. *Child Development Perspectives*, *11*(3), 207-214.

Wittgenstein, L., Anscombe, G. E. M., von Wright, G. H., Paul, D., & Anscombe, G. E. M. (1969). *On certainty* (Vol. 174). Oxford: Blackwell.