PhiMiSci
Philosophy and the Mind Sciences | Vol. 6 | 2025

What active inference still can't do

The (frame) problem that just won't go away

Darius Parvizi-Wayne[1,*]

Abstract

The frame problem, or problem of relevance, concerns the capacity of cognitive agents to zero in on relevant information during action and perception, whilst intelligently ignoring everything else. Although this is an ability that such agents realise even in the most seemingly novel of situations, it is generally accepted that no comprehensive explanatory account of it has been provided by cognitive-scientific researchers. However, a new account deriving from the popular active inference framework purports to solve the problem of relevance, an achievement which, if realised, would give strong evidence for the framework's claim to be an all-embracing theory of cognition. Unfortunately, this account, alongside previous active inference attempts to respond to the problem of relevance, is explanatorily inadequate. This means that, as is the case for all other frameworks in cognitive science, active inference has heretofore failed to resolve the problem of relevance.

Keywords: Active inference • Frame problem • Predictive processing • Relevance

[1] Macquarie University
[*] Primary contact: darius.parviziwayne@hdr.mq.edu.au

Parvizi-Wayne, D. (2025). What active inference still can't do: The (frame) problem that just won't go away. *Philosophy and the Mind Sciences*, *6*.
https://doi.org/10.33735/phimisci.2025.12118

1 Introduction

It has long been noted that any theory that purports to explain how the mind (or, given certain materialist assumptions, brain) works ought to have a clear account of how cognitive behaviour exhibits intelligent, fluid sensitivity to context-bound relevance — that is, to that which is *truly* important to the system in question within its regimes of action and perception. This is typically labelled the frame problem.

The frame problem, in fact, originated in logicist artificial intelligence (AI), where it concerned the problem of writing formulae that describe the effects of action without having to add extra formulae describing the obvious non-effects of that action (Genesereth and Nilsson, 1990; Lormand, 1990; McCarthy and Hayes, 1969). It quickly became apparent, however, that this difficulty was just one of *many* that any mechanistic account of intelligence, in attempting to account for context-bound relevance-sensitivity, faces (Pylyshyn, 1987). As such, the AI researchers' term was appropriated by cognitive scientists and philosophers of mind, who identified the frame problem with the broader problem of relevance (Dreyfus, 2007; Glymour, 1987; Rietveld, 2012; Wheeler, 2005; 2008; 2010).[1] Again, this broader issue concerns the explanation of the ostensibly rather quotidian fact that humans, as well as other intelligent creatures, can home in on particular features of reality — that is, track *certain* facts about the world and select *certain* actions — rather than others (Vervaeke et al., 2012).[2] In other words, we demonstrate "*adequate* sensitivity to context-dependent relevance", as Rietveld (2012, p. 2, italics added) puts it, such that there is a normative dimension to this capacity (see also Rietveld, 2008).

This is a facet of our lives that is utterly ubiquitous. In fact, it has been claimed that the frame problem doesn't present an overt issue that human agents resolve *per se* in every living moment. Rather, it putatively "doesn't occur" for us: we, *thrown* into an already meaningful context, just manage to zero in on salient aspects of our world in which we have been enculturated and are now embedded (Dreyfus, 2007, p. 262; see also Dreyfus, 1992). Nevertheless, this kind of account does not explain *how* we achieve this relevance realisation (Wheeler, 2008). The challenge for the cognitive scientist, therefore, is to provide a *subpersonal explanation* of this agent-level ability (Dennett, 1969; 1978; Drayson, 2012; 2014); that is, an explanation which proceeds "by analyzing a person into an organization of subsystems [...] and attempting to explain the behaviour of the whole person as the outcome of the interaction of these subsystems" (Dennett, 1978, p. 154). All these years later, this is *still* one of the central challenges in cognitive science, which, as a discipline, trades on this kind of functional analysis.

As such, the problem of relevance presses directly upon perhaps the most prominent project in all modern cognitive science: active inference. Roughly speaking, the active inference framework (AIF) holds that cognition is a probabilistic, predictive process (Friston, 2009, 2010; Parr et al., 2022). Importantly, its proponents have often heralded it as *the* one grand unifying theory to rule them all (see e.g., Clark, 2016, p. 2; Hohwy, 2013, p. 1; although cf., Bruineberg et al., 2022; Colombo and Wright, 2021; Williams, 2022). If active inference is, indeed, all that it sometimes claims to be, it surely ought to give us an explanatory, subpersonal account of this foundational, psychological sensitivity to relevance. As Reynolds (2024, p. 23) puts it: active inference theorists "either need to provide a formal and computational answer to the frame problem [...], or else they need to

[1] In fact, it has been claimed that the frame problem, in its original, narrow, AI guise has been solved (see Lifschitz, 2015; Shanahan, 1997).

[2] This is a very broad description of the relevance problem. It is worth recognising that it has been described in a multitude of ways (Chow, 2013). For example, for Fodor (1987, p. 140), it has little to do with action and is about finding a "nonarbitrary strategy for delimiting the evidence that should be searched in rational belief fixation" — that is, solving Hamlet's problem. For others, it is linked to *ceteris paribus* reasoning and the non-formalisability of background common-sense (Dreyfus, 1992, p. 57; Wittgenstein, 2009). It is also spoken about in relation to embodied skills (Dreyfus, 1992; 1996; Wheeler, 2005). Since my focus here is on critiquing proposed active inference solutions to the problem of relevance, and given that these have not played out at a particularly pronounced level of specificity, I will keep my discussion at a generally broad level, focussing more on the viability of what active inference theorists have *already* said, rather than trying to resolve every variation of the problem of relevance through active inference.

moderate some of their stronger rhetorical claims regarding being the sole cognitive kind." Importantly, there have been attempts by these theorists to attend to aspects of the problem, albeit without focussing exclusively on its wholesale resolution (see e.g., Andersen et al., 2022; Kiverstein et al., 2022; Linson et al., 2018). Such projects have culminated in a recent paper by Darling et al. (2025, p. 1), which responds to Reynolds (2024) and proposes that active inference "has the tools to comprehensively solve the problem of relevance", thus targeting the phenomenon in its entirety.

In this paper, I will argue against this claim and hold instead that there is no account that has heretofore been issued from the AIF which can adequately explain how cognitive creatures like humans solve the problem of relevance. This is not to say that active inference cannot (at least somewhat) *describe*, and has not described, the cognitive encoding of relevance in terms of its formal mechanisms. What it is lacking at this point is a powerful *explanation* for exactly *how* relevance is rapidly and flexibly realised, including in novel situations.

To outline more precisely what I am referring to here when I speak of the explanatory inadequacy of active inference with respect to the problem of relevance, we can start by noting that this type of expository weakness is certainly on show at the theoretical level of *physical* implementation. Indeed, within the AIF more broadly, neuronal details are arguably light on the ground and not entirely conclusive (Kogo and Trengove, 2015; Miłkowski and Litwin, 2022; Walsh et al., 2020; cf., Ficco et al., 2021; Kanai et al., 2015; Keller and Mrsic-Flogel, 2018), and with respect to the problem of relevance, they are wholly absent. This notwithstanding, the focus of this paper roughly spans two other levels of analysis, sketched out by David Marr (2010, p. 25) as the "computational" and "algorithmic". Admittedly, we can very broadly sketch the goal of a putative computation (or set of computations) which underwrites relevance realisation — that is, to realise relevance. At the same time, as will be shown, active inference fails to provide a convincing account, with respect to the problem of relevance, firstly, of "the logic of the strategy by which it [the goal] can be carried out", with this belonging to the Marrian (ibid.) computational level, nor of "how […] this computational theory [can] be implemented?", with this concern belonging to the so-called algorithmic level (ibid.). What's more, Marr's (2010, ibid.) three level story is overtly directed at the understanding of "any machine carrying out an information-processing task". Since active inference theorists fundamentally take the cognisor to be an information processor, abiding by certain computational principles, this broadly Marrian perspective, directed at the explanatory credentials of active inference with respect to the problem of relevance, is not out of place.

Adding a touch more here, note that what I am focussing on, with respect to the explanatory weakness of the active inference account vis-à-vis the problem of relevance, is not so much its incapacity to elucidate *why* the algorithms it posits, operating over specified inputs and outputs (see Marr, 2010, ibid.), would give rise to relevance-realising behaviour, but *how* these processes are said to function in the first place. It is this *how* question, and the adequacy of one's *how* solution, that I think is at stake in the problem of relevance, and it is, thus, from that perspective that this paper unfolds.

For comparison, this concept of the explanatory *why* underpins what philosophers often see as the hard problem of consciousness: the challenge is to explain *why* a certain neural correlate gives rise to phenomenality at all (Chalmers, 1995), rather than *how* that neural correlate processes information. These concerns are, of course, connected, but the point is that one could, in principle, provide a potentially adequate, *true* algorithmic and physical-implementational account of consciousness in terms of neural activity, whilst leaving the *why* question untouched.

With respect to this paper, although I acknowledge that there remains a legitimate question as to *why* some computational process ultimately gives rise to relevance-realising behaviour — and perhaps this would be tied into concerns about consciousness — the concern I am raising is directed at the inability of AIF theorists to provide a genuinely comprehensive, satisfying computational theory of *how* relevance realisation is achieved in the first place. To reverse the logic laid out above then, it could be argued that, in principle, one could imagine a potentially adequate, true account of *why* some neural algorithm gives rise to relevance-realising behaviour, without attending directly to *how* that algorithm in fact operationally functions.[3]

To be clear, I am not saying AIF theorists don't *attend* to this concern. If they did not, I would not have a paper to write, insofar as I am considering the efforts made thus far to deal with the problem of relevance. As such, they have *tried* to answer the *how* question. As we shall see, they have pointed to certain mechanisms, operating on algorithmic principles, with defined inputs and outputs, that they take to underwrite our capacity to realise relevance. My argument is that these explanations fall short, not exclusively because they bypass the relevant *why* question — for this is not their target — but also because they inadequately tackle the relevant *how* question. To return to Marr (2010), they fail to provide an explanatorily adequate story of the "logic of the strategy" by which relevance is realised, nor do they offer a convincing account of *how* their computational theory can be algorithmically implemented such that it is shown *how* the active inference agent *actually* realises relevance. It is thus this concept of the explanatory *how* which I will be working with in this paper when I speak of the explanatory inadequacy of the AIF with respect to the problem of relevance.

This paper proceeds as follows. In Section 2, I will briefly outline the AIF. In Section 3, I will describe the problem of relevance, stressing how it is an issue for any explanatory account, whether in AI or psychology, of cognition. In Section 4, I will offer a first AIF account of relevance, which tries to ground the realisation of relevance in the priors a Bayesian agent holds. I will propose that although this may, in principle, be able to account for relevance-realising *perceptual* inference, it cannot explain *selective* decision-making. Furthermore, how it putatively accounts for relevance-realising perceptual inference commits AIF theorists to an overly bloated conception of the internal generative model which many, if not all, would find hard to accept, and for good reason. In Section 5, I will consider the account of Andersen et al. (2022), who propose that the realisation of relevance is achieved via so-called precision-weighting. I will conclude that this proposal ends up being explanatorily weak, insofar as it does not explain the actual performance of context-sensitive precision-weighting given the agent's goals and preferences. In Section 6 and Section 7, I will examine explanations of relevance realisation which root precision-weighting in a certain class of always highly weighted priors, known as first priors. I will argue that this mechanism fails to fulfil the task it assumes, since it directly blocks the capacity for flexible, context-sensitive behaviour that any relevance-realising creature should demonstrate. It will thus be shown that AIF theorists are presented with a dilemma on either horn of which they suffer impalement. On the one hand, without invoking first priors, they risk positing an infinite regress of policies and preferences, thereby running into a canonical version of the problem of relevance. On the other, the invocation of first priors directly prohibits the target cognitive system from exhibiting the type of context-sensitive flexibility that it must demonstrate if it is to deserve the title of relevance-realising.

[3] Whether or not this is actually possible is peripheral. My point is that there is a difference, at least with respect to a kind of lens of analysis or perspective, between the explanatory how and the explanatory why. I focus on the former.

2 Active inference

Given the plethora of introductions to the AIF, I will not dive too deeply into the details here, instead recounting the essentials needed for this upcoming analysis (see Friston, 2019; Parr et al., 2022; Smith, Friston, and Whyte, 2022 for more). Active inference is often spoken of as — and can be historically taken to be — a corollary of the free energy principle, which holds that, in order to persist through time "adaptive agents must occupy a limited repertoire of states and therefore minimize the long-term average of surprise associated with sensory exchanges with the world" (Friston, 2010, p. 127). This is because the long-term average of surprise — or better, surprisal — is entropy. The agent cannot directly measure how surprising some outcome is. Therefore, in order to minimise surprisal, it is claimed that the agent minimises its proxy, variational free energy (VFE), which is an upper bound on surprisal. VFE can be decomposed as follows — see Figure 1:

$$Q(s,u) = \arg\min_Q F$$

$$F = \mathbb{E}_Q\left[\ln \underbrace{Q(s,u)}_{\text{posterior}} - \ln \underbrace{P(o \mid s,u)}_{\text{likelihood}} - \ln \underbrace{P(s,u)}_{\text{prior}}\right]$$

$$= \underbrace{D_{KL}[Q(s,u) \,\|\, P(s,u \mid o)]}_{\text{divergence}} - \underbrace{\ln P(o \mid m)}_{\text{log evidence}}$$

$$= \underbrace{D_{KL}[Q(s,u) \,\|\, P(s,u)]}_{\text{complexity}} - \underbrace{\mathbb{E}_Q[\ln P(o \mid s,u)]}_{\text{accuracy}}$$

Figure 1: VFE Equation

Here, F stands for the VFE, Q is the recognition density which is encoded within the agent's generative model and essentially stands in for the true, but computationally intractable, posterior P (s,u|o) of latent states s and action policies u, given observations o. The divergence term here refers to the Kullback-Leibler (KL) divergence between the recognition density and the true posterior. Minimising this divergence term is generally understood to be subsumed by the process of perceptual inference (Parr and Friston, 2019a). However, perception is not the only way the agent can minimise VFE; rather, it can actively intervene in the world and change the sensory inputs it receives. In fact, minimising the KL divergence between the recognition density and the true posterior technically just lowers the VFE upper bound on surprisal, but does not minimise surprisal itself (see Bruineberg et al., 2018). To minimise surprisal *per se*, the agent must encounter (by actively altering its sensorium) "expected" observations (see Parr et al., 2022, Chapter 2). Indeed, Figure 1 shows that another way to minimise VFE is to encounter observations which are highly likely given a

$$G(u) = \mathbb{E}_{Q_u}[\ln Q(s_\tau|u) - \ln Q(s_\tau|o_\tau,u) - \ln P(o_\tau|c)]$$

$$= -\underbrace{\mathbb{E}_{Q_u}[\ln Q(s_\tau|o_\tau,u) - \ln Q(s_\tau|u)]}_{\text{expected information gain}} - \underbrace{\mathbb{E}_{Q_u}[\ln P(o_\tau|c)]}_{\text{expected value}}$$

$$= \underbrace{D_{KL}[Q(o_\tau|u) \,\|\, P(o_\tau|c)]}_{\text{risk}} - \underbrace{\mathbb{E}_{Q_u}[\ln Q(o_\tau|s_\tau,u)]}_{\text{ambiguity}}$$

Figure 2: EFE Equation

generative model (a prior and a likelihood) of how that observation was generated. Observations expected under the *preferred* (or *typical* or *expected*) states associated with the agent — e.g., for a human, to be in an oxygenated room — will have a high probability, insofar as the generative model is akin to a statistical description of that agent, ensuring that persisting agents can be read as seeking out observations that provide evidence for a model that their continued existence entails. In the AIF, this process is known as self-evidencing (Hohwy, 2016).

Importantly, AIF theorists propose that adaptive agents can plan so as to minimise the "free energy expected in the future" (Parr and Friston, 2019b, p. 495). This quantity is known as expected free energy (EFE) and in the AIF it is used to score action policies (i.e., possible sequences of actions — Smith, Friston, and Whyte, 2022); see Figure 2 (Friston et al., 2015; 2016; Kaplan and Friston, 2018).

Here, G stands for the EFE expected under a policy u, where the goodness of a policy is scored by the negative EFE expected under it. More precisely, the agent will employ the policy with the lowest EFE as that becomes (or is "set to be" Parr et al., 2022, p. 53) the most expected policy and thus the one they pursue (see also Parr et al., 2022, p. 33, pp. 72-73). This summation of EFE can be achieved since each successive state associated with an action is expected to generate specific observations, allowing one to calculate (posterior predictive) beliefs about the observations predicted under a given policy. What Figure 2 demonstrates is that EFE can be decomposed into an epistemic term — the KL divergence between the prior and posterior beliefs about future states predicated on a policy, i.e., (expected) Bayesian surprise or information gain — and a pragmatic term, in which value is conditioned on preferred outcomes encoded in the so-called "c" model parameter.

Now, numerous philosophers have pointed out here that the free energy principle is essentially limited to a descriptive or instrumental account: if things continue to persist as they are, they can be *described* as if they minimise free energy, both of the variational and expected type (see e.g., Andrews, 2021; Williams, 2022; see also Nave, 2025 for a thorough discussion). However, in terms of strict entailment, nothing ontologically follows from this — i.e., one cannot derive *a priori* a mapping from the fact that self-persisting things look as if they minimise VFE to any fact that they actually *do,* nor to any process theory of how this would be achieved (Williams, 2022). As a philosopher, I do believe these are grave concerns which ought to be taken seriously, but I believe these are concerns I can in fact park in this paper.

I believe this because, as will be outlined below, active inference is an explanatory project which goes far *beyond* a description of the self-persistence of entities. For example, it claims to provide subpersonal explanations for such psychological phenomena as information-seeking (Friston et al., 2015), attention (Feldman and Friston, 2010), learning (Friston et al., 2016), and much more (see Parr et al., 2022; Sprevak and Smith, 2023). Given this, I think I can focus on the types of explanatory resources active inference brings to bear on these types of psychological phenomena — especially because it has brought these types of resources to bear on the problem of relevance (see e.g., Darling et al., 2025) — without worrying too much about the (perhaps tenuous) connection between active inference and the free energy principle.

These types of resources are standardly taken to belong to what are called the "process theories" of active inference (Parr et al., 2022). Work in this domain differs from work conducted on the free energy principle not only because they have different explanatory targets, but also because, whilst the free energy principle does not commit itself to any claims about what agents are actually doing in order to self-persist, these process theories do generally adopt a strong *realism* about the mechanisms involved in their subpersonal explanations. More precisely, as I will demonstrate, when these theories have been used to account for relevance realisation, this is done with a genuine ontological commitment to the existence and operation of certain mechanisms within certain neuropsychological systems (see e.g., Andersen et al., 2022; Darling et al., 2025). It is this standard of realism to which I will hold the active inference researcher (of the process-theory ilk), and, thus, myself, in this critique.

With the notion of a process theory outlined, it is important to note that the AIF, as a broad explanatory project, encompasses, in fact, different process theories, which either jointly or alone issue what come known to be active inference accounts of psychological phenomena (see Parvizi-Wayne et al., 2024 for an example). PP (or more narrowly, predictive coding) "remains a promising model of perceptual inference in the continuous state domain" (Sprevak and Smith, 2023, p. 14), whilst, in the discrete state space domain, the agent is assumed to "represent the world as a partially observable Markov decision process (POMDP) with discrete time and discrete states" (ibid., pp. 1-2). This second activity does not have a consistent label. Smith et al. (2022, p. 5), for one, call it "decision active inference". Given the broad overarching framework we call the AIF, I will refer to the activity of the agent, which they achieve by way of POMDPs, Markov decision processing (MDP), Note that there are so-called mixed models in the broader active inference framework, which link MDP and PP hierarchically (Friston et al., 2017). The details of such models are broadly irrelevant here, but what I will add is that the active inference accounts which tackle relevance realisation, both directly and indirectly, use the resources of both MDP and PP as means for their end, where the mechanisms and structures belonging to both frameworks are spoken of as genuinely existing. I therefore think it is appropriate to examine whether the explanatory resources belonging to *both* process theories can jointly prove to resolve the problem of relevance.

PP views the brain as a multilevel, hierarchical prediction machine whose job is to minimise prediction error, the mismatch between a prediction and the received input (Clark, 2013; 2015; Hohwy, 2013; Friston, 2005; 2008). Importantly, not every prediction error is given the same amount of neuronal weight — thus, more precisely, the job of the brain is to minimise *precision-weighted* prediction error, where precision refers to the inverse variance of the prediction error, and more precise signals are said to be given greater weight in neural processing (Feldman and Friston, 2010; Friston, 2009). Given certain Gaussian assumptions, prediction error can be shown to be equivalent to VFE, hence the reason that they are often aligned in the literature, insofar as it is presumed that one way cognitive creatures like us minimise VFE is through some PP schema (again, cf., Williams, 2022). Moreover, recent work has shown that so-called "outcome prediction error" is equivalent to EFE, indicating that such errors drive policy selection (Smith, Ramstead, and Kiefer, 2022; see also Champion et al., 2021; Friston et al., 2018). Technically, then, the prediction errors equivalent to VFE are "state" prediction errors, which drive belief updating with respect to states and can be modelled by explicit message passing algorithms.

However, outcome prediction errors are not directly associated with message passing schemes, and, thus, for the most part, when discussing policy selection, active inference researchers refer to MDP. Importantly, in active inference, preferences are formalised as prior probability distributions, not as separate variables labelled "rewards" or "values". As mentioned above, the variable c is generally used to denote these preferences, such that, given $p(o|c)$, observations which are highly probable given the agent's model are treated as rewarding. The specific type of model accessed by the agent here is a POMDP, which "describes beliefs about abstract states of the world, how they are expected to change over time, and how actions are selected to seek out preferred outcomes or rewards based on beliefs about states" (Smith, Friston, and Whyte, 2022, p. 10). POMDPs are Markovian insofar as the agent is able to use its model, in combination with beliefs about the current state they are in, to select actions related to preferred future states and observations. Recall that these states, both present and future, are hidden to the agent — they are "partially observable". Within the POMDP, other beliefs are encoded in addition to the agent's prior preferences. The agent possesses beliefs about how one state will evolve into another, encoded in a "transition matrix" (B); a likelihood function $p(o_t|s_t)$ encoded in the A matrix; prior beliefs about policies encoded in the E vector which are used to model the influence of habits; a precision estimate for the EFE of policies (γ); and beliefs about initial hidden states encoded in the D vector (Hesp et al., 2021; Sandved-Smith et al., 2021; Smith, Ramstead, and Kiefer, 2022).

3 The problem of relevance

Much ink has been spilled on the so-called frame problem, and what started as a somewhat narrow issue for engineers in classical AI[4] has morphed into a much deeper problem for cognitive scientists and philosophers of mind. For many, the question now concerns the capacity for any intelligent system (not least us) to be sensitive to that which is *relevant* to it, whilst ignoring what is not (Andersen et al., 2022; Darling et al., 2025; Dreyfus, 2007; Jaeger et al., 2024; Rietveld, 2012; Vervaeke et al., 2012). Thus, in many circles, the frame problem has become equivalent to the problem of relevance.

To bypass or resolve the problem of relevance at the agentive level[5] is to demonstrate such sensitivity flexibly and swiftly in a world of volatility, novelty and complexity like our own, rather than a "simplified and sterile world", artificially engineered such that it can be dealt with by an AI system

[4] By classical AI, I am referring to a domain of work that has cast intelligence as the rule-based manipulation of atomic symbols. This is the domain in which the narrow frame problem originated, and from which the broader problem of relevance emerged (see e.g., Dennett, 1984). I will not be discussing whether connectionist AI is able to avoid or resolve the problem of relevance (Haselager and van Rappard, 1998).

[5] I add "at the agentive level" here because, as mentioned in the introduction, the frame problem is sometimes cast as an issue for the agent, sometimes for the theory trying to explain the agent's behaviour. Of course, the two are linked, so I take steps to distinguish these formulations.

containing merely a few "frame" axioms, encoded in the predicate calculus, which describe the general context and the effects of each action type for the system in question (Dennett, 1984, p. 7). We know, of course, that, as Haugeland (1987, p. 83) bluntly puts it, "the world isn't as nicely structured as all that". Being able to account for this ability is thus a problem for cognitive scientists — because they do not know how humans, who clearly do have this ability, realise it — and artificial intelligence engineers — because they do not know how to engineer it.

In an attempt to solve this problem, classical AI theorists such as Marvin Minsky and Roger Schank proposed that similar situations — which humans do, and an AI agent could, in principle, encounter — should be grouped together into a single class (Minsky, 1975; Schank, 1975; Schank and Abelson, 1977). More precisely, they suggested these scenarios or "microworlds" — such as living rooms, birthday parties, football games, classrooms, restaurants and so on — could be summarised and condensed into a limited set of stereotypical themes or scenarios, which would then be encoded within data structures that represent those stereotypes (see Minsky, 1975, p. 212; Schank, 1975, p. 39). Minsky called these structures "frames"; Schank called them "scripts".

Within these structures what is relevant is *pregiven*, insofar as each "situation" is defined in terms of a limited set of relevant facts which act as a "predetermined causal chain of conceptualizations that describe the normal sequence of things in a familiar situation" (Schank, 1975, p. 131). Importantly, an artificial "agent" armed with such frames would come into situations yielding "expectations" about the relevant facts and how they normally change as a result of an event, and would thus be sensitive to how what they are "experiencing" diverges from the stereotypes they started with (cf., Dreyfus, 1992, p. 34). This can be read as an attempt by these theorists to capture what Dreyfus (1992, p. 56) calls "the background of commonsense know-how" that was left unaccounted for by the frame axiom approach. Interestingly, certain AI theorists explicitly stated that the development of these "micro-worlds" has deep psychological relevance, contributing to our way of "understanding human intelligence" (Winograd, 1976, p. 3; see also Winston, 1975, p. 2; Schank, 1975, p. 132).

Unfortunately, it quickly became apparent that whilst frames might work fairly well for simple scenarios, they could not reflect the idiosyncrasies of everyday life, thus rendering any AI agent armed with them deprived of real intelligence and relevance sensitivity. As Dreyfus and Dreyfus (1987, p. 106; see also Dreyfus, 1992, p. 42) point out: "in restaurant-going we do not simply enter, order, eat, and leave. After we enter, we might or might not wait to be seated, and then we might or might not study a menu, for example". In the face of this charge of skeletalism, researchers in AI suggested that frames could always be made more specific and discriminatory (Schank and Childers, 1984, p. 121). This led them to conclude, for example, that "the restaurant script contains all the information necessary to understand the enormous variability of what can occur in a restaurant" (Schank and Childers, 1984, p. 11). However, adding the amount of complexity necessary to model human behaviour is not only a hopelessly difficult endeavour, but also means that such engineered systems simply come to reflect in their own structure the very complexity they are modelling, and it seems highly unlikely — nigh-on impossible, actually — that the brain (or an AI system) harbours an encyclopaedic knowledge of every potential and actual situation it might meet, especially since each situation can be defined and redefined by way of a potentially infinite regress of qualifications as to what that situation exactly is (Dreyfus, 1992, p. 59). For example, in restaurant-going in the real-world, I may be disinclined to go for a good-old British fry up because I am watching my weight, but then I realise that I, in fact, ought to go because it's my friend's birthday and that particular local cafe is her favourite, but then I realise that my friend has been particularly encouraging of me during my weight-loss, so that maybe I could just get a drink, but that might seem rude, but who cares what other people think, *and so on*. Predetermining all *this* and setting it in place in a situation-specific slot is no small order, to say the least.[6]

[6] Furthermore, we know that in the real world the intelligent cognitive agent *is* able to zero in on these relevant concerns. They are not distracted by irrelevant issues, such as the fact that their friend will continue to exist even after their drink has been finished. How *this agent-level activity* is achieved is left unexplained by the simple frame/script approach. This is because this approach, originally directed at

In fact, this kind of solution no more removes the problem of relevance than repositions it, for, in this case, it is not just the world that the system has to deal with, but also the very internal model it possesses of the world, now composed of potentially millions of frames (Haugeland, 1987, p. 85; see also Fodor, 1983). In short, the problem is that situations are never entirely stereotypical — no two birthday parties are alike (perhaps we sing Happy Birthday in French, perhaps there are only drinks, perhaps the food is gluten-free, perhaps the birthday boy has a tantrum, and so on) — and to adjust one's pre-given frames to deal with their idiosyncrasies is to strip them of their stereotypical nature, thereby rendering them void of explanatory value, as well as computationally infeasible. However, if there are "slots" of such stereotype-structures left open for idiosyncratic details, then we must invoke a mechanism for the assignment and updating of these variables, and these are the variables that are particularly relevant, thereby recapitulating the whole problem of relevance once more (Haugeland, 1987, p. 86).

Imagining that the problem of relevance might be solved through Minskyan frames also misses an additional, crucial aspect of *human* intelligence: namely, that we are capable of grasping what matters and what doesn't not only in situations which unfold as expected, but also those that *do not*. Unlike humans, these Minskyan systems, "when their worlds turn perverse [...] typically cannot recover gracefully from the misanalyses they are led into" (Dennett, 1984, p. 9). The task, therefore, is not only to be sensitive to deviation from what is anticipated, but also to cope aptly with such changes, essentially by always already (if you'll pardon the Heideggeranese) being tuned to what is relevant in that new context. Returning to the original formulation of the frame problem, this sensitivity includes an appreciation for the effects and non-effects of one's actions, which are "strongly conditioned by situation-specific details" (Haugeland, 1987, p. 86). Again, one could rebut that all we need to do is refine our frames such that all possible changes are accounted for, but then we are back in a position where we have to be sensitive not only to what is relevant in the world but also in our indefinitely complex, isomorphic internal model.

This point in fact brings us back to the fact that no two birthday parties are entirely alike. What we might now add is that although Minskyan systems might, in principle, be able to traverse the stereotypical aspects of contexts like birthday parties, they are doomed to evince a type of unintelligence when something *un*-stereotypical happens. The point here is not that something un-stereotypical *always* happens. It is that when it does, humans are able to adapt accordingly. If my cake turns out to be made of toothpaste because of my friend's practical joke, I will respond in light of this development in an adroit, context-sensitive manner. Perhaps, for example, I'll say "yuck!". The Minskyan system, essentially *stuck* expecting the stereotypical scenario, cannot do the same. Thus, it is not *just* the case that the variability across situations like birthday parties might strip frames of their stereotypical nature. It is also the case that, since they encode stereotypical features of situations, systems made up of frames *must* fail when the un-stereotypical happens, unlike genuinely intelligent, relevance-sensitive creatures like humans.

Thirdly, as Dreyfus (2007) points out, such a system must first recognise the situation it is in in order to apply the relevant frame. For this, it will presumably need a higher-order frame to pick out a frame that picks out a set of facts, and, in turn, a higher-order frame for that frame, and so on, *ad infinitum*. As we shall see throughout, this type of infinite regress is one of the relevance problem's most pernicious weapons. Moreover, as I have already pointed out, we know that situations change, and we stay sensitive to relevance regardless. Aptly moving from frame to frame in the Minskyan sense seems to require another frame to supervise that process, and another frame for that, and so we find ourselves stuck once more.

classical AI systems, takes an overtly *external* approach to the problem of relevance, insofar as they try to tackle the issue of encoding *worldly* contingencies into a given system, such that it can recognise the context it is in, as well as the changes which occur therein. In adopting this externalist perspective, the frame/script approach also failed to give an account for the agent's dynamic, unfolding, *personal-level* sensitivity to what is *actually* relevant to *them* within those situations, such as, for example, that they are on a diet thanks to their friend's encouragement. I thank a reviewer for pointing out this additional issue of the frame/script paradigm to me.

I could go on, but for the sake of time, let me just say that what we need, in essence, is a system that, as Dennett (1984, p. 7; see also Andersen et al., 2022; Wheeler, 2005) nicely puts it, "genuinely ignores most of what it knows, and operates with a well-chosen portion of its knowledge at any moment. Well-chosen, but not chosen by exhaustive consideration." This sense of genuine ignorance means that the system does not go through every possible fact and decide whether it matters. Rather, to be intelligently ignorant is in fact to *avoid* that decision-making process, zeroing in only on what matters to it at that moment, without going through the time-consuming and intractable process of exhaustively assigning relevance to each fact (Haugeland, 1987, pp. 82-83).

Importantly, this type of intelligence needs to be applied to two domains of cognition. Firstly, it needs to occur in perceptual belief updating, insofar as I need to be able to discern not only what has changed and what has stayed the same from moment to moment, but also what *now* matters for me and what does not. It is sometimes said that achieving this would be to resolve, at the agentive level, the so-called *epistemological* version of the frame or relevance problem (Dennett, 1978; Fodor, 1983). Secondly, relevance realisation must occur in action selection, insofar as what I need to be able to zero in on what is most relevant for me to do (in short, I need to decide what to do) (Dreyfus and Dreyfus, 1987; Wheeler, 2008). Importantly, as we have seen in Section 2, action selection in the AIF is also said to be driven by prior preferences (i.e., beliefs), rendering the problem of relevance faced by an active inference agent — and active inference theorists trying to explain that agent — epistemological *across* the action/perception divide.

A couple more points of clarification. When discussing the problem of relevance, it is perhaps useful to distinguish between two scenarios in which it might appear, a distinction which I have only implicitly alluded to thus far. This has been most explicitly laid out by Michael Wheeler (2008; 2010), who argues that, on one hand, any system of putative intelligence faces an *intra-context* problem, "which challenges us to say how a naturalistically discharged system is able to achieve appropriate, flexible and fluid action within a context" (Cappuccio and Wheeler, 2012, p. 6). On the other hand, according to him, it also faces an *inter-context* problem, "which challenges us to say how a naturalistically discharged system is able to flexibly and fluidly switch between an open-ended sequence of contexts in a relevance-sensitive manner" (ibid.). To return to Minskyan frames, the point is not only that we can act intelligently *within* a frame, but also that we can flow *between* frames, resting on what Dennett (1984, p. 9) calls "a deeper understanding to fall back on". Nevertheless, it is worth pointing out that humans, as canonical relevance-realising agents, are never outside or ignorant of — that is, insensitive to — the broader context in which they are embedded (Dreyfus, 2007; Rietveld, 2012). To realise proper relevance, therefore, is to retain an inter-context sensitivity even when wholly absorbed within a given context — that is, a genuine neurophysiological openness to peripheral affordances, as well as a broader recognition of the background state of affairs — such that both problems of relevance never, in fact, appear detached (Rietveld, 2012; see also Dreyfus, 1992, p. 14 for a similar argument).[7]

Finally, although the frame problem, and the broader problem of relevance, has captivated AI researchers, it is not a given that the putative solutions they provide, whether this be McDermott and Doyle's (1980) "non-monotonic logic" or Reiter's (1980) "logic for default reasoning", tell us *anything* about what is actually happening in *humans* — that is, the processes that underwrite our ability to be sensitive to relevance (see Dennett, 1984, for further discussion). Notice that I am *not* saying that if we figured out how to construct a robot which showed this same capacity in a non-toy environment, this would necessarily have no import for psychology. Rather, it is just that this is not a given. With the distinction between the two domains set up, let me stress that my focus in this paper is on assessing claims which have been made by active inference theorists regarding how *we*, as organisms that exhibit relevance sensitivity, realise this clearly quite monumental achievement, not on the attempts of roboticists to encode this capacity. If these claims turn out to be feasible or, speaking ambitiously, *true*, it is plausible that the likelihood of our eventually constructing an intelligent robot, or artificial agent of any kind, would go up enormously

[7] A discussion of exactly what falls within and without a particular context, as well as whether the intra-context/inter-context distinction is legitimate, is unfortunately beyond the scope of this current paper (cf., Rietveld, 2012).

(Da Costa et al., 2022; Lanillos et al., 2021; Pezzulo, Parr, Cisek, et al., 2024).

As mentioned above, I am picking the AIF because of its ambitious claim to provide an explanatory account of psychological reality. The problem of relevance thus acts as a litmus test for these aspirations. I am particularly interested in the recent account provided by Darling et al. (2025, p. 1), because they "suggest predictive processing has the tools to comprehensively solve the problem of relevance." Moreover, they are not just saying that predictive processing — which, as we have seen, is a process theory belonging to the broader explanatory project of the AIF — will help us *describe* relevance realisation, but rather that it will help us "*explain* relevance" (ibid., italics added). Like me, they recognise the bold ambitions of the AIF as a "leading framework for understanding brain, mind and behaviour" and that it is therefore incumbent on theorists working within the AIF to "explain how an agent solves the problem of relevance" (ibid., p. 7) in terms of systems, structures, processes and mechanisms which are taken to genuinely exist in the brain and body. I will follow Darling et al. (2025), who exploit the whole toolbox of the AIF, in going step by step through each potential solution, where each step will *also* incorporate previous attempts by AIF theorists to account for the problem of relevance. In line with what I said in Section 2, these previous attempts also incorporate the resources of MDP and so claims from that domain will also be studied and scrutinised. What's more, some of the formal concepts in PP also belong to MDP — e.g., "precision". Therefore, some of the claims made by Darling et al. (2025) will call for rebuttals that go beyond PP. What I will show is that even when AIF theorists have pooled together all the resources belonging to both process theories, they have failed to resolve the problem of relevance, insofar as they have been unable to comprehensively explain how cognitive creatures like us realise relevance.

4 Priors

Let us start, as Darling et al. (2025) do, with the proposal that what is relevant to an agent will be connected to (or "relationally define[d]" by (ibid., p. 9)) the content of an agent's Bayesian prior probability distributions in a given situation, whatever their content might be. We can see this kind of account in Linson et al. (2018, p. 15), who claim that "relevance can be construed in relation to prior preferences about ultimate actions". Recall that these priors encode the expectations, or predictions, the agent has about the states it finds itself in. When the input that agent receives is different from what it expected, so-called prediction error emerges, and it is the minimisation of prediction error that PP theorists hold to be the central function of the brain.

Locating relevance in these priors, *prima facie*, might seem plausible. In order to resolve the epistemological version of the problem of relevance, the agent "must somehow fluidly leverage prior knowledge in their moment-to-moment interaction with the environment" (Darling et al., 2025, p. 5). By always ensuring that the agent brings *prior* expectations to a situation against which evidence is weighted, one could claim that the agent always starts off with something akin to a Minskyan frame[8], insofar as they have preloaded expectations about the events they are encountering and will encounter, deviations from which they will be sensitive to. These features of expectation and sensitivity to deviation from expectation were, indeed, constitutive of frames as I defined them earlier (see Dreyfus, 1992, p. 34). The AIF adds that such expectations are probability distributions, divergences from which are measured in terms of prediction error or VFE. With this frame-esque picture in place, the AIF theorist contends that by having a prior with a high probability, and many others which have a low probability, the agent is able to "delimit the possible considerations" about what it is encountering (Darling et al., 2025, p. 8), whether that be a window, a door, or a laptop screen, as well as about how it is going to act, given that, as revealed in Section 2, and stressed by Linson et al. (2018), actions are understood in the AIF as driven by prior preferences. In this way, these AIF theorists have, albeit implicitly, transposed the idea of a frame from classical AI into their account of relevance realisation, updating it, of course, with the relevant Bayesian or quasi-Bayesian elements.

[8] Notwithstanding, of course, that sets of priors and frames are not identical, in part because they belong to different explanatory projects.

To add a little more meat onto the explanatory bone here, let's recall that, according to AIF theorists, the algorithm on which the brain runs is roughly (or "approximately") Bayesian, whereby a posterior hypothesis about the cause of sensory data is given by the multiplication of the prior with the likelihood (of the sense data given the cause) distribution, divided by the marginal likelihood, which encodes the probability of data over all possible hidden causes (Clark, 2015; Friston, 2010; Hohwy, 2013). Importantly, the derived posterior becomes the prior for the next time step, such that, by combining prior expectations with sensory evidence, the agent keeps track of what is changing in its environment. In sum, priors act as a kind of starting point for inference: they ensure that both present and future inference is not ungrounded (Darling et al., 2025; Hohwy, 2018). This is because, from moment to moment, sensory input only acts as a source of feedback for the generative model, ensuring that it does not need to be wholesale tweaked in each instance. Notably, this notion of feedback is, at least *prima facie*, resonant with earlier, non-PP analyses of the problem of relevance, in which context-sensitivity requires a tight coupling — such as the one guaranteed by prediction error minimisation — between an agent and its world (see e.g., Wheeler, 2008; 2010).

This argument does not quite seem to be enough, however. Firstly, as we have already seen, the problem of relevance is not solved by the simple imposition of frames à la Minsky and Schank, and the feasibility of grounding relevance in priors is that they could, feasibly "act like a frame" (Darling et al., 2025, p. 8). As Darling et al. (2025, pp. 8-9) nicely point out, the relevance problem is "not necessarily that an agent could not be provided [...] particular frame axioms. Rather, the relevance problem questions how an agent or researcher comes to the right (relevant) frames". The issue is no different from the one that Minsky and Schank faced. Moreover, if our holistic sensitivity to relevance is achieved entirely through the proliferation of priors (frames), those priors would have to be as rich as the world they are taken to model. Although the sensory feedback encoded in prediction errors means that the agent does not have to instance-by-instance inaugurate a new model of the world in every passing moment, the epistemological issue of selecting and updating *only those beliefs which matter* (and which have genuinely changed) is still afoot.

To be fair, this complaint arguably does not land *if* the PP theorist claims that the world with *all* its spatio-temporal variegated dynamics, both those potential and actual, is continuously encoded in a vastly rich internal model, in which case prediction error minimisation just *does* the job of tweaking the model without the need for inferential reasoning about belief updating. In other words, if the fullness of the world is mapped from within and updated through the single route of prediction error minimisation, where prediction error directly *specifies* the mismatch between predicted and actual input, there is no further action (at least with respect to perceptual inference) that needs to take place. This is a move many would hasten to reject for *many* reasons, a thorough explication of which is beyond the scope of this paper,[9] (see e.g., Clark, 2015; Dreyfus, 1992; Facchin, 2021a; 2021b; Hutto and Myin, 2013), but one *could*, in principle from an AIF perspective, bite the bullet and suppose that the world is modelled as such.[10] Hohwy, a prominent figure in the AIF, for example, argues that prediction error minimisation "makes room for the changing world by retracting farther away from the world, into a vast internal model that seeks to represent the *full richness of the world* and the way it changes *over many time scales*" (2018, 141, italics added), and that "there is continuous

[9] For example, Facchin (2021b, p. 11609) argues that "no structure within predictive processing systems can be rightfully identified as a representational vehicle". This, *a fortiori*, blocks the possibility of a rich internal model which representationally tracks all possible and actual contingencies. Unfortunately, for the sake of space, I will be unable to engage fully with Facchin's (2021a, 2021b) rich critique of representationalism with respect to active inference systems, recognising, however, that if he is right, then there can be no richly representational priors to serve as frames, because they don't exist in the first place. In this paper, I will accept that AIF theorists could, at least in principle, posit this type of rich, internal, representational model — it is the problems that arise from this picture with *direct respect* to the problem of relevance which I focus on here. For more on the debate about representationalism in the AIF, see, in addition to Facchin (2021a, 2021b), Gładziejewski (2016); Kiefer & Hohwy (2018); Wiese (2016) and Williams (2018a, 2018b).

[10] Although this commitment is by no means necessitated within the AIF (see e.g., Clark, 2015; Tschantz et al., 2020).

modelling and tracking of *all* the relevant potential causal interactions across all contexts" (2019, p. 5, italics original; see also Kiefer and Hohwy, 2018; Wiese, 2018, pp. 191-193, p. 202, Footnote 22; Williams, 2018a). If we take this view seriously, we might be led to conclude that perceptual inference, unfolding in this manner, achieves an accurate internal model of the environment in such a way that it thereby always already tunes the agent to what has changed and what has stayed the same in their environment (cf., Wheeler, 2005, p. 67).

Of course, one could respond by saying that not *all* potential interactions can possibly be modelled simultaneously — for this would lead to combinatorial explosion and insurmountable issues for the finite human brain (see Facchin, 2021a, p. 294) — and so at any given moment the agent must still be selecting a subset of perceptual beliefs from the holistic background to update, upweight, downweight and so on.[11]

This is problematic, however, insofar as this selection leads one *back* into the problem of relevance, given that one must account for how the system knows, in some sense, what is relevant and what is not such that it selects the *relevant* subset of perceptual beliefs. Hohwy (2019, p. 5, italics added), in fact, seems to concede this, suggesting that what is tracked are "*all* the *relevant* potential causal interactions across all contexts", not all potential causal interactions tout court. However, in making this concession, Hohwy, as we can see, *assumes* the resolution of the problem of relevance in the above (2019) quotation, without further explication of how such relevance-sensitivity is achieved. For instance, it is unclear *how* such a holistic, internal, structural model could yield perceptual relevance-realisation without any further inferential reasoning — which would thereby reinstate the problem of relevance Hohwy and others are trying to avoid — nor, if it is the case that potential hidden causes "that could obfuscate the current situation or the execution of our quick and dirty strategy for achieving our goals" (Hohwy, 2019, p. 7) are also tracked, how the model does not become computationally intractable, given the infinite regress of qualifications that threatens the modelling of the unknown (Dreyfus, 1992). Thus, by recognising that we cannot track *everything*, Hohwy assumes the resolution of the problem of relevance, leaving us no better off explanatorily.

Furthermore, even if we allow that a particularly rich set of priors would, in principle, skirt the perceptual side of the epistemological problem, the action side is still open to investigation. Again, the retort could be that the action/perception divide here is weak: prediction error grows with respect to expected states and action can be taken to quash it, entailing that the worldly feedback that derives from prediction error, whether attended to by perception or action, keeps the agent tied to what is most relevant. However, the problem here is one of prioritisation: with prediction error emanating from all layers of the generative hierarchy, the agent must decide which errors to quash and a way to do so. However, this just repackages the problem of relevance insofar as the relevance of these foundational issues is under enquiry, and their explanation would, in turn, ignite a related problem of relevance. This means that Linson et al.'s (2018, p. 15) claim that "given that expected free energy scores the epistemic affordance of alternative policies on models, there is an inbuilt imperative to select significant or relevant actions" ends up being explanatorily weak with respect to the problem of relevance, since the question is *how* those relevant policies, geared towards the actualisation of relevant goals, are in fact chosen and carried out. It is interesting that Linson et al. (2018) stress the fact that EFE scores the epistemic value of a policy, and that it is this which allows the active inference agent to solve the problem of relevance. They make similar claims elsewhere: for example, they write, "by incorporating epistemic imperatives into the (Bayesian model) selection of policies in AIF, the broad frame problem never manifests" (ibid.). We should not get sidetracked by the distinction of epistemic and pragmatic affordances within the EFE equation, however (cf., Figure 2). Again, the question of relevance concerns *how* those policies — whether epistemic or pragmatic — are chosen, not just that they are, in fact, chosen, which amounts to saying that humans, somehow, surpass the frame problem.

Notice that, as depicted here, the problem of relevant updating is con-

[11] Indeed, by stressing that prediction errors are in fact *prioritised* via schemes of precision-weighting (see Section 5), AIF theorists all seem to concede this point.

nected to the so-called problem of holism (Dreyfus, 2007; Fodor, 1983; Lormand, 1996; Samuels, 2010; Wheeler, 2008). Roughly speaking, this amounts to the issue of how the agent can arrive at *particular* (updated) beliefs about the situation it is in and what it ought to do, given the indefinitely complex network of information, including the information encoded in other beliefs, which *could* be relevant at any one point (Darling et al., 2025, p. 5). In other words, *tractably* determining what is truly of relevance — i.e., without scouring one's entire nexus of beliefs, or even the total universe — such that at least adequately successful, timely perception and action is achieved, seems problematic and is not clearly resolved by the mere insistence that what is relevant is encoded in one's priors (Dietrich and Fields, 1996).[12] Again, by having prior updating tied to feedback from the world in the guise of prediction errors, the PP theorist could, in principle, respond by proposing that the very act of prediction error minimisation ensures the realisation of relevance, such that the problems of holism and updating are never in fact in play, because the model itself is i) holistic and ii) updates itself. However, this just supposes once again that the entire world, both in its actual and potential form, is recapitulated in the generative model, a view even the most representationalist of AIF theorists, such as Hohwy reject, and rightly so. But in eschewing this account, we have seen that these theorists have resorted to assuming the resolution of the problem of relevance, which is the very issue they ought to be addressing.

Furthermore, even if it we grant that that the system is able to globally minimise prediction error via this gigantic internal model, what is left unexplained is how the agent is able to *prioritise* the minimisation of some prediction errors over others, a type of prioritisation one would assume is in play unless we console ourselves with the idea that total prediction error is minimised in one fell swoop. Again, this is at least imaginably the case in perceptual inference, which, recall, involves arriving at a posterior hypothesis that best explains the currently incoming sense data. This is because the emergence of such a posterior just is the representation of a univocal selection of a prioritised belief, a belief which comes to be prioritised, in essence, because it best minimises prediction error. In other words, what is prioritised in terms of model updating just falls out of prediction error minimisation, which can be plausibly achieved given an immensely rich generative model (although see Footnote 11, this paper). However, this doesn't seem to be quite right from the perspective of action and decision-making, for it is an intrinsic feature of a relevance-realising creature to *prioritise* what it does (including what it pays attention to — see Parvizi-Wayne, 2024; Watzl, 2017). Indeed, only through such prioritisation can a long-term average of prediction error minimisation occur, something that PP theorists have argued the cognitive system is ultimately directed at (Hohwy, 2020, p. 213; cf., Sprevak and Smith, 2023). Thus, even if the holistic adjustment of the model is sufficient *in principle* to ensure context-sensitive perceptual inference, the mere imposition of priors fails to offer a sufficiently viable account of action selection.

In sum — and note that I am in accord with Darling et al. (2025) here — grounding relevance in priors alone is explanatorily inadequate. Admittedly, one could always describe what is relevant to an agent as (or related to) that which is encoded in its priors. Furthermore, by ensuring that posteriors are always informed by sensory input, the PP story is able to claim, at least on some descriptive level, that the agent will, just by minimising prediction error, home in on an accurate model which also facilitates their goal-directed behaviour, insofar as that activity *just* is necessitated by the imperative to minimise prediction error. However, any account of relevance realisation according to which the putatively relevance-sensitive agent is armed with priors alone fails at a deeper explanatory level, insofar as the imposition of priors alone does not elucidate *how* the agent *in fact* zeroes in on just those beliefs that are relevant at that moment. To skirt this issue, *at least with respect to perceptual inference,* is to rely on the deeply troubling idea of a generative model that tracks *everything,* whereby the effect of zeroing in on an accurate internal model just is a consequence of successful

[12] Note that the problem of holism I am discussing here is broader than the inferential problem of holism, which states that "given appropriate background beliefs, (almost) any belief can be rendered relevant to the assessment of (almost) any other belief" (Samuels, 2010, p. 284). This is because the information that is potentially relevant to one's beliefs derive not only from the other beliefs the agent holds, but from the environment more broadly.

perceptual inference. This view is unpalatable for multiple reasons, a fact which AIF theorists seemingly concede. Furthermore, without invoking further features of the PP architecture, the story heretofore presented cannot account for how the right subset of beliefs are in fact prioritised in real time in the service of action. In fact, to account for this context-sensitive prioritisation, the PP theorist invokes so-called precision-weighting. It is to this mechanism that we now turn.

5 Precision-weighting

Recall that, according to PP, the central role of the brain is to minimise not just prediction errors, but *precision-weighted* prediction errors (Clark, 2013; Hohwy, 2013; Feldman and Friston, 2010; Friston, 2009). This is generally taken to be the case because some sensory signals are going to be more ambiguous than others, and the brain ought to prioritise less ambiguous (i.e., precise) signals more as it carries out perceptual inference and model learning. In other words, prediction errors are weighted according to their reliability — measured in terms of precision — and this determines the degree to which predictions are updated by influencing the posterior distribution (Friston and Kiebel, 2009). Importantly, the precision of prediction errors is context-sensitive and thus can come to be expected (Hohwy, 2012). In other words, the predictive agent is said to possess beliefs about expected precision, and this allows them to assign confidence to prediction error signals in real time, thereby ensuring that only the signals with the highest signal-to-noise ratio affect perceptual processing (Clark, 2015). Finally, attention is taken to correspond with precision optimisation, insofar as it is understood to be a gain mechanism that upweights signals which are expected to have a high precision (although cf., Parvizi-Wayne, 2024; Ransom and Fazelpour, 2020).

With all this in mind, Andersen et al. (2022, p. 366, original emphasis) propose that "assigning high precision is thus the perceptual system's way of labelling an error signal as being particularly *relevant* [...] All of this suggests that precision-weighting simply is how relevance is realized within the PP framework". So, are they right? Does the imposition of precision-weighting resolve the relevance problem? I suggest that it does not.

In seeing one reason why, let's return to a now famous frame in which we humans (seemingly effortlessly) cope with the world and exhibit spectacular sensitivity to relevance: a birthday party. From the story outlined above, it seems Andersen et al. (2022) are proposing that what is relevant is that stimulus that is least expected to be noisy — perhaps the part of the room where it is quietest or brightest (Clark, 2015). But this cannot be right: what is most likely to be relevant is probably exactly where it's noisiest and the prediction errors are unable to pick out the underlying cause of sense data — e.g., under the dazzling disco ball where the music is playing full blast (see Ransom et al., 2020).[13] This is an issue for PP theorists that goes beyond the scope of this paper (see Parvizi-Wayne, 2024 for a full outline). What I will say is that, henceforth in this paper, we should follow Clark (2017, p. 118; see also Andersen et al., 2022, p. 367) in holding that "precision should not be thought of simply as encoding the reliability of a signal, but also its estimated value" (Clark, 2017, p. 118; although cf., Parvizi-Wayne, 2024, pp. 19-20 for a slightly different approach).

Back to the birthday party. With this updated account of precision, I can now say that whatever I estimate to be relevant will be that which I pay attention to. Crucially, precision-weighting predicated on expected precision also seems to afford the type of context-sensitivity we are after (Hohwy, 2020). Thus, the argument goes, whatever is most relevant to me is that to which I assign most precision-weight — ergo, relevance is precision-weighting.

[13] Indeed, when PP theorists discuss the noisiness of a prediction error, they are generally referring to how unambiguously an observation and a state map onto one another, such that a noisy signal is ambiguous (and thus taken to be imprecise) if, from the perspective of the signal receiver, there are many things which could have caused it (see e.g., Clark, 2015, pp. 56-57; Hohwy, 2012, p. 6; 2013, p. 195, p. 124; Friston, 2012, p. 238). Thus, a noisy signal is one which, if it were to contribute to perceptual inference, "would lead to less accurate predictions and therefore less accurate representation of the external world" (Ransom and Fazelpour, 2020, p. 120).

Unfortunately, this position does not constitute an adequate explanation of relevance. Rather, it simply kicks the can down the road. To see why, consider that there are (at least) three questions we can ask Andersen et al. (2022) here, to probe whether their account truly deals with the problem of relevance. Firstly, why are these specific stimuli relevant to the actual agent in question? Secondly, what happens when the context changes — e.g., a clown bursts out of the birthday cake? Thirdly, and related to the second question, how does the agent decide between stimuli of competing relevance?

Starting with the first question, it appears Andersen et al. (2022, p. 367; cf., Ransom et al., 2020) have an easy route out. As they put it:

"Why would we direct our attention towards a barking dog? Because we have some high-precision proprioceptive prediction which states something like 'I will avoid bodily injury' and the barking dog (and any other unexpected loud noise) reliably provides information which is likely to be relevant to that prediction?"

The essential point here is that, with our updated account of precision, what is weighted in terms of gain control is that information which pertains to one's goals and preferences (Parvizi-Wayne, 2024). Recall that in active inference, such goals, preferences and desires are conceptually elided and are cast as the prior beliefs the agent has about the states it expects to be in (Friston et al., 2014; 2015; 2022).[14] With this in mind, the sceptic could press on and ask why a given agent holds these priors in the first place (Klein, 2018). As Clark (2020, p. 6) rightly points out, "the PP-mechanism itself offers no concrete story here." That said, it strikes me that this is not necessarily a death knell for Andersen et al. (2022); one could simply retort that however those priors are bequeathed, either by one's evolutionary or experiential past, they are the beliefs that govern one's realising of relevance (Pezzulo et al., 2022).

Now onto the second and third questions. To my eyes, these strike at the heart of the relevance problem, insofar as they probe *how* the agent decides what is relevant now and *how* they remain acutely sensitive in real time to changing sources of relevance. In essence, they ask *mechanistic* questions. It is all well and good to propose that precision-weighting underlies relevance; it is another thing to be able to explain how precision-weighting is manipulated on the fly to ensure that real relevance — i.e., a veridical tracking of reality *and* adequate goal achievement — is actually realised (Reynolds, 2024). Again, by this I am *not* calling for a how explanation at the Marrian (2010) level of *physical implementation*; rather, I am calling upon (previous) AIF theorists to explain how relevance realisation occurs at a level which touches upon both Marr's computational and algorithmic levels.

Unfortunately, Andersen et al. (2022) offer no answers to these types of questions. They thus offer, in my eyes, a simply descriptive active inference account of relevance, falling short of explanatory adequacy with respect to the problem of relevance. Again, on one hand, precision-weighting does — now with a slight adjustment to its original conceptualization — mark what is relevant to the agent. This means that an external observer, given a suite of an active inference agent's precision-weighting, would be able to predict what is most relevant for that agent. On the other hand, how precision-weighting is actually enacted flexibly and in novel situations is left unexplained. Thus, to truly explain *how* relevance is encoded, it is not enough to just insist that it is via precision-weighting; rather, one needs

[14] Even if we continued to construe the precision-weighting of errors as resting on expectations about the noisiness of those prediction errors, this would still render that process just as representationally heavy as the first-order inner model governed by perceptual inference, insofar as expectations over noise rest on wider context tracking — e.g., whether I am at a cocktail party and should therefore down-weight auditory signals, or in the dark and should therefore down-weight visual signals (see Clark, 2015; Hohwy, 2013; Kiefer and Hohwy, 2018, p. 2400). Again, the AIF theorist could either postulate a (surely unattractive) gigantic internal model where precision expectations are sensitive to all tracked, external contingencies, or they could assume the resolution of the problem of relevance, where precision-expectations always already pertain to a relevant subsection of the internal model. I think, for reasons outlined above, either option is wholly undesirable. Furthermore, from the perspective of MDP, it will be shown in Section 6 that the policy of assigning high precision-weight to the likelihood "A" matrix, which is taken to be equivalent to a prediction error, suffers from a regress problem. I thank an anonymous reviewer for encouraging me to press this point further.

to explain how precision-weighting itself works within the whole active inference economy.

On this note, we have also seen that precision-weighting is subservient to the status of higher-order expectations, which deem certain stimuli to be relevant or not. If we want to answer mechanistic questions, therefore, about how precision-weighting is algorithmically realised, it appears that it is the status and function of *those* priors that really ought to be examined. To these priors we now turn.

6 Expected precision

The second and third questions raised to Andersen et al. (2022) end up targeting the same issue, albeit subtly. Let's imagine that while my friend is blowing out her candles, my other friend, who I have not seen for years (she lives in Australia) walks through the door. I am now conflicted, caught in a dilemma of frames. Who is more relevant to me?

Somehow or another, I will end up making the decision. A decision must, in fact, be made, even if that is to close my eyes and make no decision. That, as Sartre taught us, is a decision. So, does the active inference story explain how this form of decision-making works? If it does, it seems we are on the brink of resolving the problem of relevance.

Since we are now talking about precision-weighting governed by one's preferences — do I look at my blowing-out-candles-friend or my just-come-home-from-Australia friend or do I just close my eyes?— we ought to bring in the resources of MDP, which, recall, is used to explain the decision-making and action selection processes of active inference agents. The story is essentially as follows. Some preference, let's say not to upset anyone, suddenly becomes more pronounced — technically, precision is deployed over the C tensor encoding this belief — and this constrains my action selection, including my mental action of paying attention, such that I shut my eyes and keep them closed (Pezzulo et al., 2018; Sandved-Smith et al., 2021). This is because that sharpened (via precision over C) preferred outcome to not upset anyone enters into the EFE calculation, such that policies with the lowest EFE are those that are biased towards the achievement of that outcome, rather than towards the type of epistemic foraging that would ensue if low precision was deployed over C, which would lead to the flattening of the preference distribution encoded therein (Smith, Friston, and Whyte, 2022, p. 7; Sprevak and Smith, 2023, p. 17).

Let me provide a bit more detail. Given that the preference not to upset anyone is threatened, precision-weight over the C tensor encoding that particular preferred outcome goes up, which ensures that I take the appropriate action — i.e., close my eyes — to achieve that preference. More precisely, through policy selection associated with this *higher* level (i.e., not to upset anyone), my cognitive system will deploy weight ("precision") over *lower-order* preferences (i.e., C). This will drive, through policy selection, weight over even lower-order preferences and so on and so forth until precision weight is deployed over preferences encoded in the lowest sensorimotor levels of the generative hierarchy, leading to the actual execution of motor action, such that the minimisation of EFE at each level, where each level tracks different spatio-temporal dynamics, ensures the realisation of preferred outcomes ultimately designated by the preference of not upsetting anyone (Pezzulo et al., 2018; see also Clark, 2020; Friston et al., 2014; Parr et al., 2022, p. 204; Pezzulo, 2012; Pezzulo et al., 2015). There is, in other words, hierarchical motivated control geared by dAI, whereby "the higher levels of the control hierarchy [...] prescribe the initial states [...] and the prior preferences over the evolution of their future sequences of states (via C) of the lower levels. Crucially, the influence from higher- to lower-level state sequences is precision weighted [...] this allows the motivational hierarchy to optimize the precision of prior preferences" (Pezzulo et al., 2018, p. 299). Again, this relies on predictions of (or precisely weighted preferences about) precision, whereby "descending predictions of precision in hierarchical inference can be construed as a form of attentional selection. In the present setting, these predictions play the role of intentional or goal selection by, effectively, applying an attentional bias to *prior preferences*." (ibid., p. 301, italics added)".

Voilà: it seems like the relevance problem has been solved, insofar as the prior that defines what is relevant is the one that is highly weighted, since that weighting has been determined by the agent's uppermost concerns and goals. In a similar vein, Darling et al. (2025, p. 16) write that "the precision over action policies also constrains the potentially relevant policies, which helps avoid another version of the relevance problem, where an agent faces an unbounded policy space." The claim in question here, however, is even stronger: it is not just that precision over policies — if read as precision over C — constrains the policy space, but that it strongly biases which policies are chosen![15] With this, it appears that the action selection, decision-making, planning side of the problem of relevance is solved by active inference. This is because this account purports to tell us *how* precision-weighting (again, described as how relevance is encoded) onto priors and prediction errors (where relevance is encoded) is, in fact, enforced.

But there's a problem. Again, why did precision over the C tensor encoding the preferred outcome of keeping my eyes shut get upweighted? The active inference story here, as we have seen, is that there was a higher-order policy driving precision over that C tensor. But why is that policy chosen? Well, the active inference theorist says, there must have been high precision over a C tensor (with its encoded preferences) associated with that policy. And onwards we go, into an infinite regress, in a brain which, of course, we do not expect to regress infinitely.

More broadly, the argument is that if, to assign a C tensor precision-weight, thereby leading to a particular selection of overt policies (P1), I must employ a higher-order policy (P2) (predicated on its own set of preferences encoded in C) to enforce the assignment of that precision-weight, I will then need another policy (P3) that leads to the assignment of precision-weight over the C tensor associated with P2, and so on and so forth. This type of infinite regress is a canonical feature of the frame problem (Dreyfus, 1992; Wheeler, 2005; 2008).

7 First priors

One *prima facie* plausible way to prevent this infinite regress is to posit a set of priors — or, more technically, C tensors that encode them — that are permanently highly precision-weighted and thus do not require a higher-order policy to set them in place.[16]

For example, in the case of shutting my eyes, perhaps my prior preference to not upset anyone is permanently upweighted. When that preference is "perturbed", for want of a better word, it sets off the sequence of hierarchical motivated control which ends in me shutting my eyes, although without this first, permanently highly weighted prior of not upsetting anyone needing to be upweighted itself. The permanently upweighted prior to not upset anyone just sets high precision-weight over a lower-order preference, which enforces a policy to set high precision-weight over an even lower-order preference, and so on and so forth, until I shut my eyes.

This type of approach, which gestures at a set of permanently highly precision-weighted, prioritised priors, is not exactly novel (see Allen and Tsakiris, 2018; Clark, 2020; Darling et al., 2025, p. 17; Kiefer and Hohwy, 2025, pp. 12-13; Kiverstein et al., 2024; Pezzulo et al., 2022; 2024, p. 5; Sladky et al., 2024, p. 226). As Clark (2020, p. 5) puts it: "one response at this point has been to depict some predictions as simply more deeply ingrained than others", where deeply ingrained means "with high action-entraining precision" (ibid, p. 6; see also Allen and Tsakiris, 2018, p. 31; Kiverstein et al., 2024, p. 6). This is an approach, in fact, directly advocated by Darling et al. (2025, p. 17),

[15] Darling et al. (2025) might well be referring here to precision over beliefs about state transitions, encoded in the B matrix, which policies model. This precision-weighting is also under the control of hierarchical precision-weighting (and thereby falls prey to the same infinite regress problem which befalls the hierarchical weighting of preferred outcomes). In this case, precision over B can be read as the "confidence that the selected policy will achieve the desired goals" (Pezzulo et al., 2018, p. 304). They might also be referring to the inverse temperature parameter γ, which scales EFE estimates.

[16] Note I will continue to speak here in terms of permanently unweighted priors, rather than C tensors, as this is how this phenomenon is generally discussed in the AIF. This is of little significance. Permanent weight over C tensors encoding said preferences would still lead to those preferences dominating action selection, insofar high precision would ensure that one preferred outcome in a C tensor would come to be even more prioritised over others.

who claim that "agents also have various beliefs about preferred states, and some of those beliefs are held with extremely high precision. For biological agents, these probably include high-precision beliefs about homeostatic states, some of which may be"given" to us (say, through evolution or during prenatal development".[17]

I point this out because Darling et al. (2025, p. 16) use this argument to support their putatively *strong* account of relevance realisation, where it is not just the case "that agents may have various (sometimes strange) goal-states, which result in a range of real relevance determinations." Rather, as they say, "the relevance problem questions *how* the agent (itself) determines relevance; how does the robot, strange goals and all, ignore the endless irrelevant consequences of action, and zero in on the relevant actions that make progress toward its goal?" (Darling et al., 2025, pp. 16-17, italics added). This, as I have been stressing throughout, is exactly the question an adequately explanatory account of relevance ought to be answering. Thus, the explanation that context-sensitive precision-weighting is ensured by having a certain set of priors permanently upweighted constitutes an attempt by active inference theorists to offer a genuinely explanatory, subpersonal account of relevance. In the case of the birthday party, perhaps my preference to not upset anyone is already so highly weighted that when it is under threat — i.e., when I really am at risk of upsetting somebody — I ensure that all my behaviour conforms with that ultimate preference. Following Allen and Tskarisis (2018) and Kiverstein et al. (2024), I will call these always highly weighted preferences *first priors*.

Unfortunately, I am not convinced that introducing first priors solves the problem of relevance. In fact, I am not convinced that there even are permanently upweighted first priors. Again, to solve the problem of relevance presented by the birthday party, it would be the case that my prior to not upset anyone is permanently upweighted, guaranteeing that, when the threat of that prior not being fulfilled is raised, no higher-order policy is needed to ensure its enslaving effects on downstream action. Firstly, the active inference theorist needs to explain how the threatened status of the prior — in this case not to upset anyone — is realised by the agent in the first place, such that this modulation, in turns, sets off the enslaving effect of that particular preference at that moment.

Even if we grant that the active inference theorist could provide us with this story in such a way that we can genuinely hold that *no* extra precision is deployed over that preference and yet it manages to determine what is relevant in the moment, there is a further problem. The prediction derived from the putative presence of first, highly weighted priors, such as one encoding our desire not to upset anyone, is that we should always, *and only*, be seeking to not upset anyone — to stay with this example. One could, of course, respond by saying that many people do live their lives with this desire in play. However, this retort misses the mark: the point is that if the weight of this preference is essentially fixed, no flexibility *at all* is afforded to the agent, a flexibility which we know humans evince, such that, for example, at one moment they are more selfish than at others. Furthermore, it is unclear how the agent could negotiate between *all* the first priors they are said to hold; presumably, at points, they would have to prioritise satiation over caregiving, sleep over satiation and so on. However, reinstantiating prioritisation just reinstantiates the problem of relevance and the threat of infinite regress: one, would, for example, require a higher-order policy *prioritizing* sleep over satiation. Thus, the two horns of what one might call "the prior dilemma" arise: on one side, an infinite regress; on the other, rigidity.

To the best of my knowledge, that having permanently weighted preferences would *not* afford the type of context sensitivity and flexibility which intelligent systems display, has only been recognised, among the AIF theorists, by Clark (2020, p. 6), who writes:

"Such deep-set predictions get us only so far. At some point, the PP- theorist needs to accommodate the ordinary shifting webs of (as we would ordinarily say) desire—the ebbs and flows of intention that sometimes lead us to play the piano, then to work on a paper, then to order a Chinese rather than Indian takeaway, watch a certain movie, and so on."

[17] Recent work by Kiverstein et al (2024, p. 4) proposes that these first priors are not simply those that encode biological variables, but "relate more generally to the lifestyle the agent leads in its niche".

One response here could be to say that such deep-seated preferences always start with a fairly high degree of precision, which then becomes accentuated in relevant contexts. The problem with this suggestion is that it, again, reintroduces the threat of an infinite regress: in order to deploy that extra bit of precision over my preference to not upset my friends at the birthday party, I must employ a higher-order policy with its own associated preferences, whose precision require their own associated higher-order policy, and so on and so forth. In other words, as soon as there is some "context-sensitive adjustment of precision-weighting" (Kiverstein et al., 2024, p. 11), the frame problem seems to re-emerge (see also Dreyfus, 1992, p. 221). This is the case too if we imagine that the members of the set of first priors are constantly changing, perhaps for affective reasons. All of this *presupposes* a relevance-sensitivity: for example, if the claim is that affect tunes context-sensitive modulation of precision-weighting, the cognitive system would need to know what priors (now cast as first priors) and prediction errors (related to those first priors) to modulate, assuming prior knowledge of relevance. Thus, since AIF theorists who have explored the problem of relevance describe the assignment of precision-weight in model-based policy-selection-based terms (see e.g., Pezzulo et al., 2018, p. 298), they fail to surpass the problem of relevance. Once model-based policy-selection is in play, the infinite regress charge is always afoot — perhaps indicating that non-model-based solutions might be a fertile avenue for active inference theorists to explore (Kotler et al., 2025; Maisto et al., 2019; Parvizi-Wayne et al., 2024).

Consequently, when Kiverstein et al. (2024) propose that it is not only first priors which are "always privileged" (p. 15), but also that any prediction error related to a first prior will "automatically be conferred a high precision" (p. 6; note that they attribute this secondary claim to Allen & Tsakiris (2018) – see e.g., p. 31), they run into two problems, at least from the perspective of the problem of relevance. Firstly, they postulate first priors. Secondly, invoking the notion of automatic precision-assignment induces an infinitely regressive hierarchy of policies and preferences. The only solution is to cap off any potential regress via the imposition of first priors all over again. But we have seen that this cognitive architecture would *also* not yield intelligent relevance-sensitivity between different contexts — i.e., within a whole life.

In sum, unlike other active inference theorists, I do not think that the introduction of a special class of always highly weighted priors is the way to resolve the problem of relevance. With this, we have finished our survey of extant AIF attempts to resolve the problem of relevance. We are left to conclude that it is a problem which still plagues them.[18]

8 Concluding remarks

In sum, it is not simply enough to say that an "agent that conforms to active inference inherently optimizes toward realizing strong relevance" (Darling et al., 2025, p. 19). In my eyes, this amounts to no more than saying an agent with goals that acts to achieve those goals will behave in a way that is goal-directed. The question, ultimately, is *how* the agent does so. Here, I find active inference accounts to still be lacking (Reynolds, 2024). This is a concession that some of the active inference theorists who try to tackle the problem of relevance make. For example, Kiverstein et al. (2022, p. 8) write, "How then does the agent constrain the search space to only action policies of relevance (i.e., those expected to minimize free energy)?

[18] An anonymous reviewer helpfully suggested that rather than treating behaviour as the output of a computational, policy-selection system, we might begin with the low-level, continuous operation of prediction error minimisation (PEM), and treat relevance and policy-like structures as post hoc conceptualisations. In principle, I could imagine that context-sensitive behaviour could ultimately be cashed out at a descriptive level in terms of PEM. My concern is that an account which states this alone is explanatorily weak, a weakness which mechanisms like policy selection via MDP do, to be fair, tackle. And yet, as we have seen, these mechanisms lie at the centre of the AIF's inability to resolve the problem of relevance, rendering the AIF, in this domain, explanatorily inadequate. Finally, as I have stressed throughout, I am eager to consider the efforts made thus far to tackle the problem of relevance from within the AIF. With respect to those efforts, policy selection, precision-weighting and the like are not post-hoc conceptualisations. On the contrary, they are mechanisms which AIF theorists take to be genuinely operative in the brain and body.

Most active inference models up until now have avoided this question by pre-specifying the search space."

What I think has happened is that the modeller's pre-specification of the search space has translated into a pre-specification of the search space *within* the generative hierarchy via the imposition of permanently upweighted first priors. This achieves, at least *prima facie*, the type of context-sensitive constraints over decision-making that accounts for adequate hierarchical motivated control (Pezzulo et al., 2018). Unfortunately, it is an inadequate solution, and results in a dilemma both horns of which AIF researchers would be inclined to avoid. They thus ought to return to the drawing board to deal with this one problem that just won't go away.

As a very final point, it ought to be recognised that I have not strictly relied here — although I hint at it in Section 4 — on a stronger Dreyfus-style argument (e.g., 1992, 2007), which holds that if a system is strictly computational (in the sense of realising algorithmic, state-dependent, sequential processes which are sensitive to the syntactic properties of representations which are thereby accessed, manipulated and/or transformed), it will be unable to overcome the frame problem (see also Fodor, 1987; 2000; Haugeland, 1987; Samuels, 2010; Wheeler, 2010). Roughly put, Dreyfus holds this position because he believes that the background, common knowledge we hold cannot be reduced to a set of context-free facts, such as those represented in a computer. He argues, instead, that our background knowledge is constituted by an embodied, non-propositional know-how, a know-how that allows us to respond directly to worldly relevance, as past experience gears us to respond to context-bound solicitations, not context-insensitive epistemic primitives (Dreyfus, 1990; 1992; 2007; 2017; see also Wheeler, 2005, p. 168; 2010).

It is generally held that PP[19] and MDP are computational process theories which permit talk of an internal, representational model which stands in, at least to some degree, for the external world, as the organism's locus of control (see e.g., Corlett et al., 2020; Shipp, 2024). Again, for Dreyfus (see e.g., 1992, pp. 299-300), a cognitive system defined by its internal model of the world will never be able to skirt the problem of relevance, because it is faced with the issue, firstly, of storing this encyclopaedic array of facts, and, secondly, of accessing those facts in a context-sensitive manner. He concludes that "this problem is avoided by human beings because their model of the world is the world itself" (Dreyfus, 1992, p. 300).

As I suggested in Section 4, the purported richness of this recapitulated model is open to interpretation, however, and it is unclear whether the AIF is thus committed to the type of radical representationalism with its foundational, so-called metaphysical assumption — namely, that background knowledge and common-sense is systemically encoded within the cognitive architecture or internal model — to which Dreyfus is so opposed (Dreyfus, 1992, p. 56, p. 266; cf., Clark, 2015, 2016; Hohwy, 2018; 2019). Even the Clarkian approach, however, according to which the internal model (or "economy") is action-orientated, fast, frugal, situated, and in which "sensing, thinking, and acting conspire, overlap, and start to merge together as whole perceptuo-motor systems engage the world" (Clark, 2015, p. 12), seems unable to outright deal with the problem of relevance. This is because it presupposes, rather than addresses, the issue of just how the cognitive system can zero in on that which is most relevant to it. Unfortunately for the AIF, we have also seen that the Hohwyian approach, which is so often juxtaposed to the Clarkian, is liable to fall foul of the same issue, rendering both approaches explanatorily inadequate with respect to the problem of relevance.

In sum, what I have tried to show in this paper is that attempts to resolve the frame problem using the toolbox of active inference have failed *thus far*. Whether such attempts were doomed to fail because of the very computational basis of the process theories which belong to active inference is a far more damning suggestion which ought to be taken seriously. If those investigations are, in fact, seen to be indicting, it might be wise to turn to alternative, non-computational accounts of cognition for an explanation of relevance realisation.

To offer just one example, Michael Wheeler (see e.g., 2005; 2008) has approached the problem of relevance from the perspective of dynamical systems theory, proposing — just to highlight a single aspect of his rich

[19] Interestingly, and somewhat ironically, PP follows computational principles despite being used to model continuous state spaces.

account — that inter-context relevance sensitivity emerges through the non-representational, non-computational continuous, large-scale, holistic reconfiguration of the cognitive system's dynamics (see also Clark, 1998; Dreyfus, 2007; Freeman, 2000). If he is on the right track, then subpersonal explanations of relevance realisation need not revert to the presence of rigid structures underlying all context-sensitivity — such as permanently upweighted first priors — to circumvent the type of regress that arguably befalls all computational accounts of action selection, and it can eschew the image of an indefinitely bloated internal model, employed in order to track all actual and potential worldly affairs, which, for reasons outlined above, I believe we should firmly resist. That said, as Wheeler (see Wheeler and Di Paolo, 2014, p. 9) concedes, "the cognitive-scientific story here is incomplete", and is incomplete because, as he admits, even his non-computational approach falls short of thoroughly explaining relevance-realisation, although for different reasons than the ones which confront the AIF theorist.

Acknowledgments

I would like to thank two anonymous reviewers, Neil Levy, Jack Reynolds, Thomas Parr, Karl Friston, Pat Sweeney and Jaime Ruiz Serra for their helpful feedback on the ideas expressed in this paper.

References

Allen, M., & Tsakiris, M. (2018). The body as first prior: Interoceptive predictive processing and the primacy of self-models. In M. Tsakiris & H. De Preester (Eds.), *The interoceptive mind: From homeostasis to awareness* (p. 0). Oxford University Press. https://doi.org/10.1093/oso/9780198811930.003.0002

Andersen, B. P., Miller, M., & Vervaeke, J. (2022). Predictive processing and relevance realization: Exploring convergent solutions to the frame problem. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-022-09850-6

Andrews, M. (2021). The math is not the territory: Navigating the free energy principle. *Biology & Philosophy*, *36*(3), 30. https://doi.org/10.1007/s10539-021-09807-0

Bruineberg, J., Dołęga, K., Dewhurst, J., & Baltieri, M. (2022). The emperor's new Markov blankets. *Behavioral and Brain Sciences*, *45*, e183. https://doi.org/10.1017/S0140525X21002351

Bruineberg, J., Kiverstein, J., & Rietveld, E. (2018). The anticipating brain is not a scientist: The free-energy principle from an ecological-enactive perspective. *Synthese*, *195*(6), 2417–2444. https://doi.org/10.1007/s11229-016-1239-1

Cappuccio, M., & Wheeler, M. (2012). Ground-level intelligence: Action-oriented representation and the dynamics of the background. In Z. Radman (Ed.), *Knowing without thinking: Mind, action, cognition and the phenomenon of the background* (pp. 13–36). Palgrave Macmillan UK. https://doi.org/10.1057/9780230368064_2

Chalmers, D. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies*, *2*(3), 200–219.

Champion, T., Grześ, M., & Bowman, H. (2021). Realizing active inference in variational message passing: The outcome-blind certainty seeker. *Neural Computation*, *33*(10), 2762–2826. https://doi.org/10.1162/neco_a_01422

Chow, S. J. (2013). What's the problem with the frame problem? *Review of Philosophy and Psychology*, *4*(2), 309–331. https://doi.org/10.1007/s13164-013-0137-4

Clark, A. (1998). *Being there: Putting brain, body, and world together again* (Revised). MIT Press.

Clark, A. (2013). Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, *36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, A. (2015). Radical predictive processing. *The Southern Journal of Philosophy*, *53*, 3–27. https://doi.org/10.1111/sjp.12120

Clark, A. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. OUP USA.

Clark, A. (2017). Predictions, precision, and agentive attention. *Consciousness and Cognition*, *56*, 115–119. https://doi.org/10.1016/j.concog.2017.06.013

Clark, A. (2020). Beyond desire? agency, choice, and the predictive mind. *Australasian Journal of Philosophy*, *98*(1), 1–15. https://doi.org/10.1080/00048402.2019.1602661

Colombo, M., & Wright, C. (2021). First principles in the life sciences: The free-energy principle, organicism, and mechanism. *Synthese*, *198*(14), 3463–3488. https://doi.org/10.1007/s11229-018-01932-w

Corlett, P. R., Mohanty, A., & MacDonald, A. W. (2020). What we think about when we think about predictive processing. *Journal of Abnormal Psychology*, *129*(6), 529–533. https://doi.org/10.1037/abn0000632

Da Costa, L., Lanillos, P., Sajid, N., Friston, K., & Khan, S. (2022). How active inference could help revolutionise robotics. *Entropy*, *24*(3), 361. https://doi.org/10.3390/e24030361

Darling, T., Corcoran, A. W., & Hohwy, J. (2025). Solving the relevance problem with predictive processing. *Philosophical Psychology*, 1–26. https://doi.org/10.1080/09515089.2025.2460502

Dennett, D. C. (1969). *Content and consciousness*. Routledge; Kegan Paul.

Dennett, D. C. (1978). *Brainstorms*. MIT Press.

Dennett, D. C. (1984). Cognitive wheels: The frame problem of AI. In C. Hookway (Ed.), *Minds, machines and evolution*. Cambridge University Press.

Dietrich, E., & Fields, C. (1996). The role of the frame problem in Fodor's modularity thesis: A case study of rationalist cognitive science. In K. M. Ford & Z. W. Pylyshyn (Eds.), *The robot's dilemma revisited: The frame problem in artificial intelligence* (pp. 9–24). Ablex.

Drayson, Z. (2012). The uses and abuses of the personal/subpersonal distinction. *Philosophical Perspectives*, *26*(1), 1–18. https://doi.org/10.1111/phpe.12014

Drayson, Z. (2014). The personal/subpersonal distinction. *Philosophy Compass*, *9*(5), 338–346. https://doi.org/10.1111/phc3.12124

Dreyfus, H. L. (1990). *Being-in-the-world: A commentary on Heidegger's being and time, division I*. Bradford.

Dreyfus, H. L. (1992). *What computers still can't do: A critique of artificial reason*. MIT Press.

Dreyfus, H. L. (1996). The current relevance of Merleau-Ponty's phenomenology of embodiment. *Electronic Journal of Analytic Philosophy*.

Dreyfus, H. L. (2007). Why Heideggerian AI failed and how fixing it would require making it more Heideggerian. *Artificial Intelligence*, *171*(18), 1137–1160. https://doi.org/10.1016/j.artint.2007.10.012

Dreyfus, H. L. (2017). *Background practices: Essays on the understanding of being*. Oxford University Press.

Dreyfus, H. L., & Dreyfus, S. E. (1987). How to stop worrying about the frame problem even though it's computationally insoluble. In Z. W. Pylyshyn (Ed.), *The robot's dilemma: The frame problem in artificial intelligence* (pp. 95–112). Ablex.

Facchin, M. (2021a). Are generative models structural representations? *Minds and Machines*, *31*(2), 277–303. https://doi.org/10.1007/s11023-021-09559-6

Facchin, M. (2021b). Predictive processing and anti-representationalism. *Synthese*, *199*(3), 11609–11642. https://doi.org/10.1007/s11229-021-03304-3

Feldman, H., & Friston, K. (2010). Attention, uncertainty, and free-energy. *Frontiers in Human Neuroscience*, *4*. https://doi.org/10.3389/fnhum.2010.00215

Ficco, L., Mancuso, L., Manuello, J., Teneggi, A., Liloia, D., Duca, S., Costa, T., Kovacs, G. Z., & Cauda, F. (2021). Disentangling predictive processing in the brain: A meta-analytic study in favour of a predictive network. *Scientific Reports*, *11*(1), 16258. https://doi.org/10.1038/s41598-021-95603-5

Fodor, J. A. (1983). *The modularity of mind*. MIT Press.

Fodor, J. A. (1987). Modules, frames, fridgeons, sleeping dogs, and the music of the spheres. In Z. W. Pylyshyn (Ed.), *The robot's dilemma: The frame problem in artificial intelligence* (pp. 139–149). Ablex.

Fodor, J. A. (2000). *The mind doesn't work that way: The scope and limits of computational psychology*. The MIT Press.

Freeman, W. (2000). *How brains make up their minds*. Columbia University Press.

Friston, K. J. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *360*(1456), 815–836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. J. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Friston, K. J. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293–301. https://doi.org/10.1016/j.tics.2009.04.005

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

Friston, K. J. (2012). Predictive coding, precision and synchrony. *Cognitive Neuroscience*, *3*(3–4), 238–239. https://doi.org/10.1080/17588928.2012.691277

Friston, K. J. (2019). A free energy principle for a particular physics [arXiv:1906.10184]. https://doi.org/10.48550/arXiv.1906.10184

Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience and Biobehavioral Reviews*, *68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022

Friston, K. J., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1521), 1211–1221. https://doi.org/10.1098/rstb.2008.0300

Friston, K. J., Parr, T., & deVries, B. (2017). The graphical brain: Belief propagation and active inference. *Network neuroscience*, *1*(4), 381–414.

Friston, K. J., 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

Friston, K. J., Rosch, R., Parr, T., Price, C., & Bowman, H. (2018). Deep temporal models and active inference. *Neuroscience and Biobehavioral Reviews*, *90*, 486–501. https://doi.org/10.1016/j.neubiorev.2018.04.004

Friston, K. J., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., & Dolan, R. J. (2014). 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

Genesereth, M. R., & Nilsson, N. J. (1990). Logical foundations of artificial intelligence. *Journal of Symbolic Logic*, *55*(3), 1304–1307. https://doi.org/10.2307/2274491

Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, *193*(2), 559–582.

Glymour, C. (1987). Android epistemology and the frame problem: Comments on Dennett's 'cognitive wheels'. In Z. W. Pylyshyn (Ed.), *The robot's dilemma* (pp. 65–76). Ablex.

Haselager, W. F. G., & van Rappard, J. F. H. (1998). Connectionism, systematicity, and the frame problem. *Minds and Machines*, *8*(2), 161–179. https://doi.org/10.1023/A:1008281603611

Haugeland, J. (1987). An overview of the frame problem. In Z. W. Pylyshyn (Ed.), *The robot's dilemma: The frame problem in artificial intelligence*. Ablex.

Hesp, C., Smith, R., Parr, T., Allen, M., Friston, K. J., & Ramstead, M. J. D. (2021). Deeply felt affect: The emergence of valence in deep active inference. *Neural Computation*, *33*(2), 398–446. https://doi.org/10.1162/neco_a_01341

Hohwy, J. (2012). Attention and conscious perception in the hypothesis testing brain. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00096

Hohwy, J. (2013). *The predictive mind*. OUP Oxford.

Hohwy, J. (2016). The self-evidencing brain. *Noûs*, *50*(2), 259–285. https://doi.org/10.1111/nous.12062

Hohwy, J. (2018). The predictive processing hypothesis. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The oxford handbook of 4E cognition* (pp. 129–145). Oxford University Press.

Hohwy, J. (2019). Quick'n'lean or slow and rich?: Andy Clark on predictive processing and embodied cognition. https://doi.org/10.1093/oso/9780190662813.003.0015

Hohwy, J. (2020). New directions in predictive processing. *Mind & Language*, *35*(2), 209–223. https://doi.org/10.1111/mila.12281

Hutto, D. D., & Myin, E. (2013). *Radicalizing enactivism: Basic minds without content*. MIT Press.

Jaeger, J., Riedl, A., Djedovic, A., Vervaeke, J., & Walsh, D. (2024). Naturalizing relevance realization: Why agency and cognition are fundamentally not computational. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1362658

Kanai, R., Komura, Y., Shipp, S., & Friston, K. (2015). Cerebral hierarchies: Predictive processing, precision and the pulvinar. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *370*(1668), 20140169. https://doi.org/10.1098/rstb.2014.0169

Kaplan, R., & Friston, K. J. (2018). Planning and navigation as active inference. *Biological Cybernetics*, *112*(4), 323–343. https://doi.org/10.1007/s00422-018-0753-2

Keller, G. B., & Mrsic-Flogel, T. D. (2018). Predictive processing: A canonical cortical computation. *Neuron*, *100*(2), 424–435. https://doi.org/10.1016/j.neuron.2018.10.003

Kiefer, A., & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, *195*(6), 2387–2415. https://doi.org/10.1007/s11229-017-1435-7

Kiefer, A., & Hohwy, J. (2025). A predictive architecture for the attitudes. https://doi.org/10.31219/osf.io/4n27k_v1

Kiverstein, J., Kirchhoff, M. D., & Froese, T. (2022). The problem of meaning: The free energy principle and artificial agency. *Frontiers in Neurorobotics*, *16*. https://doi.org/10.3389/fnbot.2022.844773

Kiverstein, J., Miller, M., & Rietveld, E. (2024). Desire and motivation in predictive processing: An ecological-enactive perspective. *Review of Philosophy and Psychology*. https://doi.org/10.1007/s13164-024-00757-6

Klein, C. (2018). What do predictive coders want? *Synthese*, *195*(6), 2541–2557. https://doi.org/10.1007/s11229-016-1250-6

Kogo, N., & Trengove, C. (2015). Is predictive coding theory articulated enough to be testable? *Frontiers in Computational Neuroscience*, *9*. https://doi.org/10.3389/fncom.2015.00111

Kotler, S., Parvizi-Wayne, D., Mannino, M., & Friston, K. (2025). Flow and intuition: A systems neuroscience comparison. *Neuroscience of Consciousness*, *2025*(1), niae040. https://doi.org/10.1093/nc/niae040

Lanillos, P., Meo, C., Pezzato, C., Meera, A. A., Baioumy, M., Ohata, W., Tschantz, A., Millidge, B., Wisse, M., Buckley, C. L., & Tani, J. (2021). Active inference in robotics and artificial agents: Survey and challenges [arXiv:2112.01871]. https://doi.org/10.48550/arXiv.2112.01871

Lifschitz, V. (2015). The dramatic true story of the frame default. *Journal of Philosophical Logic*, *44*(2), 163–176. https://doi.org/10.1007/s10992-014-9332-8

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: General information dynamics for natural and artificial embodied cognition. *Frontiers in Robotics and AI*, *5*. https://doi.org/10.3389/frobt.2018.00021

Lormand, E. (1990). Framing the frame problem. *Synthese*, *82*(3), 353–374. https://doi.org/10.1007/BF00413881

Lormand, E. (1996). The holorobophobe's dilemma. In K. Ford & Z. Pylyshyn (Eds.), *The robot's dilemma revisited: The frame problem in artificial intelligence* (pp. 61–88). Ablex.

Maisto, D., Friston, K., & Pezzulo, G. (2019). Caching mechanisms for habit formation in active inference. *Neurocomputing*, *359*, 298–314. https://doi.org/10.1016/j.neucom.2019.05.083

Marr, D. (2010). *Vision: A computational investigation into the human representation and processing of visual information*. MIT Press.

McCarthy, J., & Hayes, P. (1969). Some philosophical problems from the standpoint of artificial intelligence. In B. Meltzer & D. Michie (Eds.), *Machine intelligence 4* (pp. 463–502). Edinburgh University Press.

McDermott, D., & Doyle, J. (1980). Non-monotonic logic I. *Artificial Intelligence*, *13*(1–2), 41–72. https://doi.org/10.1016/0004-3702(80)90012-0

Miłkowski, M., & Litwin, P. (2022). Testable or bust: Theoretical lessons for predictive processing. *Synthese*, *200*(6), 462. https://doi.org/10.1007/s11229-022-03891-9

Minsky, M. (1975). A framework for representing knowledge. In P. Wilson (Ed.), *The psychology of computer vision* (pp. 211–277). McGaw Hill.

Nave, K. (2025). *A drive to survive: The free energy principle and the meaning of life*. The MIT Press.

Parr, T., & Friston, K. J. (2019a). Attention or salience? *Current Opinion in Psychology*, *29*, 1–5. https://doi.org/10.1016/j.copsyc.2018.10.006

Parr, T., & Friston, K. J. (2019b). Generalised free energy and active inference. *Biological Cybernetics*, *113*(5), 495–513. https://doi.org/10.1007/s00422-019-00805-w

Parr, T., Pezzulo, G., & Friston, K. J. (2022). *Active inference: The free energy principle in mind, brain, and behavior*. The MIT Press. https://doi.org/10.7551/mitpress/12441.001.0001

Parvizi-Wayne, D. (2024). How preferences enslave attention: Calling into question the endogenous/exogenous dichotomy from an active inference perspective. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-024-10028-5

Parvizi-Wayne, D., Sandved-Smith, L., Pitliya, R. J., Limanowski, J., Tufft, M. R. A., & Friston, K. J. (2024). Forgetting ourselves in flow: An active inference account of flow states and how we experience ourselves within them. *Frontiers in Psychology*, *15*. https://doi.org/10.3389/fpsyg.2024.1354719

Pezzulo, G. (2012). An active inference view of cognitive control. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2012.00478

Pezzulo, G., Parr, T., Cisek, P., Clark, A., & Friston, K. (2024). Generating meaning: Active inference and the scope and limits of passive AI. *Trends in Cognitive Sciences*, *28*(2), 97–112. https://doi.org/10.1016/j.tics.2023.10.002

Pezzulo, G., Parr, T., & Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*(1844), 20200531. https://doi.org/10.1098/rstb.2020.0531

Pezzulo, G., Parr, T., & Friston, K. (2024). Active inference as a theory of sentient behavior. *Biological Psychology*, *186*, 108741. https://doi.org/10.1016/j.biopsycho.2023.108741

Pezzulo, G., Rigoli, F., & Friston, K. (2015). Active inference, homeostatic regulation and adaptive behavioural control. *Progress in Neurobiology*, *134*, 17–35. https://doi.org/10.1016/j.pneurobio.2015.09.001

Pezzulo, G., Rigoli, F., & Friston, K. J. (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

Pylyshyn, Z. W. (Ed.). (1987). *The robot's dilemma: The frame problem in artificial intelligence*. Ablex.

Ransom, M., & Fazelpour, S. (2020). The many faces of attention: Why precision optimization is not attention. In S. Gouveia, D. Mendonça, & M. Curado (Eds.), *The philosophy and science of predictive processing*. Bloomsbury.

Ransom, M., Fazelpour, S., Markovic, J., Kryklywy, J., Thompson, E. T., & Todd, R. M. (2020). Affect-biased attention and predictive processing. *Cognition*, *203*, 104370. https://doi.org/10.1016/j.cognition.2020.104370

Reiter, R. (1980). A logic for default reasoning. *Artificial Intelligence*, *13*(1), 81–132. https://doi.org/10.1016/0004-3702(80)90014-4

Reynolds, J. (2024). Framing the predictive mind: Why we should think again about Dreyfus. *Phenomenology and the Cognitive Sciences*. https://doi.org/10.1007/s11097-024-09979-6

Rietveld, E. (2008). Situated normativity: The normative aspect of embodied cognition in unreflective action. *Mind*, *117*(468), 973–1001. https://doi.org/10.1093/mind/fzn050

Rietveld, E. (2012). Context-switching and responsiveness to real relevance. In J. Kiverstein & M. Wheeler (Eds.), *Heidegger and cognitive science* (pp. 105–134). Palgrave Macmillan UK. https://doi.org/10.1007/978-1-137-00610-3_3

Samuels, R. (2010). Classical computationalism and the many problems of cognitive relevance. *Studies in History and Philosophy of Science Part A*, *41*(3), 280–293. https://doi.org/10.1016/j.shpsa.2010.07.006

Sandved-Smith, L., Hesp, C., Mattout, J., Friston, K., Lutz, A., & Ramstead, M. J. D. (2021). Towards a computational phenomenology of mental action: Modelling meta-awareness and attentional control with deep parametric active inference. *Neuroscience of Consciousness*, *2021*(1), niab018. https://doi.org/10.1093/nc/niab018

Schank, R. C. (1975). The primitive ACTs of conceptual dependency. In B. L. Nash-Webber & R. Schank (Eds.), *Theoretical issues in natural language processing*. https://aclanthology.org/T75-2008/

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding: An inquiry into human knowledge structures*. Lawrence Erlbaum.

Schank, R. C., & Childers, P. (1984). The cognitive computer on language, learning, and artificial intelligence [Available at: https://www.osti.gov/biblio/5756491].

Shanahan, M. (1997). *Solving the frame problem: A mathematical investigation of the common sense law of inertia*. MIT Press.

Shipp, S. (2024). Computational components of visual predictive coding circuitry. *Frontiers in Neural Circuits*, *17*. https://doi.org/10.3389/fncir.2023.1254009

Sladky, R., Kargl, D., Haubensak, W., & Lamm, C. (2024). An active inference perspective for the amygdala complex. *Trends in Cognitive Sciences*, *28*(3), 223–236. https://doi.org/10.1016/j.tics.2023.11.004

Smith, R., Friston, K. J., & Whyte, C. J. (2022). A step-by-step tutorial on active inference and its application to empirical data. *Journal of Mathematical Psychology*, *107*, 102632. https://doi.org/10.1016/j.jmp.2021.102632

Smith, R., Ramstead, M. J. D., & Kiefer, A. (2022). Active inference models do not contradict folk psychology. *Synthese*, *200*(2), 81. https://doi.org/10.1007/s11229-022-03480-w

Sprevak, M., & Smith, R. (2023). An introduction to predictive processing models of perception and decision-making. *Topics in Cognitive Science*, tops.12704. https://doi.org/10.1111/tops.12704

Tschantz, A., Seth, A. K., & Buckley, C. L. (2020). Learning action-oriented models through active inference. *PLOS Computational Biology*, *16*(4), e1007805. https://doi.org/10.1371/journal.pcbi.1007805

Vervaeke, J., Lillicrap, T. P., & Richards, B. A. (2012). Relevance realization and the emerging framework in cognitive science. *Journal of Logic and Computation*, *22*(1), 79–99. https://doi.org/10.1093/logcom/exp067

Walsh, K. S., McGovern, D. P., Clark, A., & O'Connell, R. G. (2020). Evaluating the neurophysiological evidence for predictive processing as a model of perception. *Annals of the New York Academy of Sciences*, *1464*(1), 242–268. https://doi.org/10.1111/nyas.14321

Watzl, S. (2017). *Structuring mind. the nature of attention and how it shapes consciousness*. Oxford University Press.

Wheeler, M. (2005). *Reconstructing the cognitive world: The next step*. Bradford.

Wheeler, M. (2008). Cognition in context: Phenomenology, situated robotics and the frame problem. *International Journal of Philosophical Studies*, *16*(3), 323–349. https://doi.org/10.1080/09672550802113235

Wheeler, M. (2010). The problem of representation [Available at: http://dspace.stir.ac.uk/handle/1893/11451]. Springer.

Wheeler, M., & Di Paolo, E. (2014). Existentialism and cognitive science. https://doi.org/10.5040/9781474227759.ch-013

Wiese, W. (2016). What are the contents of representations in predictive processing? *Phenomenology and the Cognitive Sciences*, *16*(4), 715–736.

Wiese, W. (2018). *Experienced wholeness: Integrating insights from Gestalt theory, cognitive neuroscience, and predictive processing* [Available at: https://muse.jhu.edu/pub/6/monograph/book/57448]. The MIT Press.

Williams, D. (2018a). Predictive minds and small-scale models: Kenneth Craik's contribution to cognitive science. *Philosophical Explorations*, *21*(2), 245–263. https://doi.org/10.1080/13869795.2018.1477982

Williams, D. (2018b). Predictive processing and the representation wars. *Minds and Machines*, *28*(1), 141–172. https://doi.org/10.1007/s11023-017-9441-6

Williams, D. (2022). Is the brain an organ for free energy minimisation? *Philosophical Studies*, *179*(5), 1693–1714. https://doi.org/10.1007/s11098-021-01722-0

Winograd, T. (1976). *Artificial intelligence and language comprehension*. National Institute of Education.

Winston, P. (1975). *The psychology of computer vision*. McGaw Hill.

Wittgenstein, L. (2009). *Philosophical investigations* (4th). Wiley-Blackwell.