



LUND UNIVERSITY

Rethinking intelligent behaviour through the lens of accurate prediction

Adaptive control in uncertain environments

Poth, Nina Laura; Tjøstheim, Trond A.; Stephens, Andreas

Published in:
Philosophy and the Mind Sciences

DOI:
[10.33735/phimisci.2025.11780](https://doi.org/10.33735/phimisci.2025.11780)

2025

Document Version:
Publisher's PDF, also known as Version of record

[Link to publication](#)

Citation for published version (APA):
Poth, N. L., Tjøstheim, T. A., & Stephens, A. (2025). Rethinking intelligent behaviour through the lens of accurate prediction: Adaptive control in uncertain environments. *Philosophy and the Mind Sciences*, 6. <https://doi.org/10.33735/phimisci.2025.11780>

Total number of authors:
3

Creative Commons License:
CC BY

General rights

Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: <https://creativecommons.org/licenses/>

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00



Rethinking intelligent behaviour through the lens of accurate prediction

Adaptive control in uncertain environments

Nina Laura Poth^a  (ninalaura.poth@ru.nl)

Trond A. Tjøstheim^b  (trond_arild.tjostheim@lucs.lu.se)

Andreas Stephens^b  (andreas.stephens@fil.lu.se)

Abstract

While recent cognitive science research shows a renewed interest in understanding intelligence, there is still little consensus on what constitutes intelligent behaviour and how it should be assessed. Here we propose a refined approach to biological intelligence as accurate prediction, according to which intelligent behaviour should be understood as adaptive control driven by the minimisation of uncertainty in dynamic environments with limited information. Central to this view is the concept of accuracy, which we argue is key to determining the success of predictions. We identify tensions in applying this framework to contemporary artificial systems such as large-language models, which, despite their impressive capacities for abstract prediction, show deficits in terms of context-sensitive knowledge transfer.

Keywords

Accuracy · Adaptive control · Artificial intelligence · Biological intelligence · Embodied cognition · Intelligence · Intelligent behaviour · Predictive processing

1 Introduction

While recent cognitive science research shows a renewed interest in understanding intelligence, there is still little consensus on what constitutes intelligent behaviour and how it should be assessed. Legg and Hutter (2007) review definitions of intelligence across different contexts, including dictionary and encyclopaedic definitions, as well as definitions from psychology and AI research. They summarise their synthesis definition as “[i]ntelligence measures an agent’s ability to achieve goals in a wide range of environments.” (Legg and Hutter, 2007, p. 9), thus implicitly including *goal achievement* and *context independence*. However, these contemporary characterisations say little explicitly about the role of the body and

^a Department of Philosophy of Mind and Language, Radboud University.

^b Department of Philosophy and Cognitive Science, Lund University.

environment – surprisingly, as it was central already to the cybernetics tradition (Wiener, 1948), and, to some, is the source of a remaining conceptual disconnect between AI modelling and the study of robotics (Rajan and Saffiotti, 2017, pp. 2–4). Biological views are now more frequently considered “highly relevant for AI researchers striving to build accurate models of natural cognition” because “the biological foundations of enactive cognitive science can provide *the conceptual tools* that are needed to diagnose more clearly the shortcomings of current embodied AI” (Froese and Ziemke, 2009, p. 466, emphasis added).

In this paper we explore a view of biological intelligence as a form of accurate prediction, as one perspective that we believe offers fruitful insights on adaptive behaviour. This view roots in the predictive processing framework (PP) of cognitive and biological systems, and suggests that we should rethink intelligent behaviour as adaptive control driven by prediction and the minimisation of uncertainty in dynamic environments with limited information (see Geary, 2009; Tjøstheim and Stephens, 2022).¹ In drawing out the view, we clarify the important role that the notion of *accuracy* plays in generating *adaptive control*. In particular, we emphasise the importance of embodied prediction, where organisms rely on their interactions with the environment to optimise their actions and conserve energy. Accurate predictions are not only about forecasting the future but also about transferring those predictions across different contexts. The purpose of this transfer is to enhance an organism’s ability to effectively prepare for energy-optimising action and control the anticipated environmental effects. We argue that one way of achieving adaptive control happens through a process of *re-concretisation*, where a learned abstraction is reified into a novel situation through the direct experience with the world that embodiment affords. This explains why AI models often generalise at the expense of detail, while biological systems can tailor predictions to specific environments and changes over time. Thus, prediction is not everything to adaptive behaviour, which also relies on embodied models that align predictions with real-world constraints.

The structure of the paper is as follows. In section 2, we present and discuss the predictive processing framework of cognition and its ties to cybernetics. In section 3, we argue that biological intelligence should be interpreted as a form of accurate prediction that generates adaptive control. We highlight general principles of feedback control and self-regulation in complex systems as being at their

¹ Following this view, we understand ‘adaptive behaviour’ ecumenically, in terms of a property associated with a trait that is likely to be selected in a specific type of environment. Being well adapted to an environment means, for biological systems, being likely to survive and reproduce in that environment. For artificial systems, we understand this analogically to mean that such a system is likely to persist and replicate or propagate information in conditions typically imposed by that sort of environment. More generally, being well adapted to an environment means being likely to resist decay in that environment. Propagating information is important for this because it provides flexibility and resilience in dynamic environments that require the system to go beyond meeting narrow pre-defined design goals and instead change its own goals relative to the changing environmental conditions over time.

core influenced by the context-sensitivity and the embodiment of that system. In section 4, we apply this view to draw distinctions between biological intelligence and AI such as large-language models (LLMs). In section 5, we present concluding remarks.

2 Taking a predictive perspective

2.1 Basics of PP

At its core, PP assumes that cognition follows a single ‘imperative’ to minimise a quantity called ‘prediction error,’ on average and in the long run (Friston, 2005; Hohwy, 2013). Prediction errors can be characterised as evaluations of the difference in the information content between incoming sensory signals at some point in time and the cognitive system’s previous predictions or ‘guesses’ (Figdor, 2021). There remains much controversy about how information content in predictions is to be understood exactly but a general characterisation is in terms of a system’s estimate of a particular environmental state occurring (Sprevak, 2020). In its formulation as a version of the Bayesian Brain Hypothesis (Knill and Pouget, 2004), the framework claims that the brain represents information at the subpersonal level in terms of probability distributions over possible sensory states at various levels of a hierarchical generative model. The model is cortically implemented by a hierarchical message-passing schema in which top-down connections of neural networks carry predictions about activities at lower levels of the cortical hierarchy and bottom-up connections carry information about errors in those predictions, and prediction error signals encode information about the discrepancy between incoming signals and prior predictions at each level (Wiese and Metzinger, 2017). Generally, cognitive systems are characterised as creating models (internal representations) of the external world (Craik, 1944) and then updating their models given feedback in the form of prediction errors. Message passing is seen as a way of gaining information by minimising uncertainty; the more uncertainty there is, the more information is gained by removing the uncertainty (Shannon, 1948).

Hierarchical information processing is particularly efficient when much data needs to be processed, as is the case for complex systems such as the brain.² A hierarchical organisation allows that only information about the error, as opposed to the complex incoming signal, is processed at higher levels, and this minimises the amount of data to be processed (Rao and Ballard, 1999; Bubic et al., 2010, p. 10).

PP refines this rationale in the context of cognition research by suggesting that prediction errors can be minimised in multiple ways. In perception, prediction error minimisation is achieved by updating the generative model that produces top-

² The central nervous system from this perspective may be seen as the steersman, where the brainstem controls lower-level functions, the midbrain processes sensory data, while the neocortex can govern complex mental models. This implicit notion of hierarchical control can be traced back to seminal work in control theory (Lefkowitz, 1966; Mesarovic, 1970)

down predictions at higher levels of the hierarchy in proportion to the magnitude of incoming error signals, in such a way that the resulting predictions approximate information from novel incoming sensory signals at lower levels. An alternative way of minimising prediction error is by changing the world through actions to accommodate internal predictions. This ‘active inference’ perspective is at times embedded within a broader view on the important role that cognitive principles might play for the possibility of life. This view starts from the assumption that organisms obtain certain objective goals, such as to stay alive and to reproduce. And to satisfy these goals, they need to self-regulate to maintain homeostasis.

The second way proposes that organisms self-regulate by minimising free energy (the analogue of minimisation of prediction error in neural systems) in the long run (Friston and Stephan, 2007; Friston et al., 2010). The quantity of free energy has been proposed as a representation of the misfit between an embodied agent and its niche, where the embodied agent (i.e., not only the brain) models the relevant environment statistics (Bruineberg and Rietveld, 2014). Simultaneously, the agent’s behaviour may impact its ecological niche, for example, an agent can learn the most efficient path to a pre-specified foraging location that may also emerge as a function of the niche-constructing activities of the agent itself (Bruineberg et al., 2018). Niche construction here refers to the way a cognitive system actively shapes and manipulates its external environment to enhance its cognitive functioning. The key idea is that cognition is not merely a product of the brain’s internal workings but is shaped by how cognitive systems interact with and alter the environment to support thinking and problem-solving (Clark, 2008). The minimisation or prediction error here refers to the optimisation of an agent-environment system. Instead of reactively customising the internal world-model in response to environmental effects, the agent actively changes its environment to make future processes of updating and interactions with the environment more efficient. Niche construction can thus reduce the long-term energetic cost that the agent must bring up to adapt its internal model to match the environment. Although the tendency to utilise external structures to support cognitive processing is an evolved trait, not all forms of cognitive niche construction are necessarily adaptive, that is, it could also lead to maladaptive behaviours. The concept of niche construction in embodied PP emphasises the interactive relationship between cognition and the environment and thus illustrates how cognitive systems shape and are shaped by the environmental conditions they engage with.

It is important to note that while PP and the free-energy principle (FEP) often co-occur, they are conceptually distinct frameworks with distinct assumptions and scope. PP starts from the assumption that the brain faces uncertainty about incoming sensory inputs. The aim is to explain why and how the brain generates and changes predictions to reduce sensory error. The FEP starts from the broader assumption that organisms aim to maintain homeostasis in dynamically changing, uncertain, environments. To do this, organisms minimise free energy to stay in predictable, stable, states. That is, they align their internal model, which also repre-

sents how sensory input is generated, with external world states (e.g., as through niche construction). Free energy quantifies the difference between the system's predictions and reality. As a first principle in the life sciences, the FEP should apply to any living adaptive system, not just nervous systems. There are ongoing discussions about how PP and the FEP relate. While some suggest that PP is a specific implementation of the FEP, i.e., a specific approach to how the brain processes sensory data in line within the FEP's demand of minimising free energy to maintain a stable and coherent model of the world (see Piekarski, 2023), others are more cautious. For example, Sprevak and Smith (2023, p. 7) argue that "... there is no simple logical equivalence between variational free energy and prediction error. Identifying which assumptions most plausibly connect free energy minimisation and prediction error minimisation within the brain remains an open issue."

The significance of both PP and the FEP are often stated in terms of their integrative power, as noted by Pezzulo and Sims (2021):

[...] the FEP is an integrative proposal on adaptive self-organizing systems that suggests living organisms manage to survive by forming internal generative models of the causes of their sensations and using them to minimize a measure of (roughly) surprise—or in other words, to ensure that they remain in the ecologically “adaptive” states that they should inhabit. (Pezzulo and Sims, 2021, p. 7807)

It has been argued that prediction error minimisation provides an elegant abstract schema to unify diverse aspects of cognition such as perception, action, learning, attention, memory, motivation, social cognition, psychopathology, language, consciousness, and other phenomena (Friston, 2010; Clark, 2008, 2013, 2016; Hohwy, 2013; Gładziejewski, 2016; Sprevak, 2024; Poth, 2022). Some authors, such as Beni (2018) and Gładziejewski (2019), claim that the framework also provides unified mechanistic explanations of mind and brain, but this thesis remains controversial.

However, this integrative character is challenged by a remaining tension between computational versus embodied interpretations of the framework when using it to understand mind and cognition.

2.2 Computational versus embodied PP

Computationalist approaches to PP follow the standard model of efficient coding (Section 2.1), wherein, whether it is appropriate to minimise prediction error in the long run via perception or via action in either case depends on internal aspects of the model. Specifically, it depends on the precision in proportion to which error signals are weighted. Precision is defined as the inverse of the variance of the subjective probability distribution involved in prediction. On one view, these probability distributions are interpreted as subjective degrees of belief. This interpretation illustrates the internalist view of PP advocated by, for instance, Frith (2007), Hohwy (2013) and Kiefer and Hohwy (2018), according to which predictive agents

internally model their environments. This view interprets the Markov Blanket formalism (Pearl, 1988) commonly associated with the FEP to endow the FEP with implicit representational assumptions by suggesting an interpretation of agents as models of the world. The reasoning is that internal and external states cannot influence each other directly but only via acting and sensing, and so these processes form a blanket that separates the agent's internal model from the immediate environmental causes (formally, the external environmental and internal sensory processes become statistically independent). On Hohwy's internalist view, it is due to the segregation of internal states from the external world via the Markov blanket that the brain's processing should be understood as a form of internal, secluded, inference, for a lack of direct access (see also Parr et al., 2020).

Other groups of PP researchers interested in embodied cognition shift focus away from the internalist view and transfer the maxim of prediction error minimisation to the domain of adaptive behaviour in living organisms relative to their ecological niche (Bruineberg and Rietveld, 2014; Clark, 2016), thus proposing a non-intellectualist perspective. From this perspective, predictions in physiology are typically interpreted as cybernetic setpoints that allostatic processes do work to achieve (Seth, 2014, 2015). PP proponents of this view suggest that prediction is a necessary condition of action and environmental control. These connections between cybernetics and PP are often displayed as contrastive to the standard model, describing how "perception emerges as a consequence of a more fundamental imperative towards homeostasis and control, *and not* as a process designed to furnish a detailed inner 'world model' suitable for cognition and action planning." (Seth, 2015, p. 3, emphasis added). Wiese (2015, p. 3) correspondingly labels this view "action as predictive control", contrasting it with "action as hypothesis-testing".³ The prioritised target is to understand how living beings are able to go on living, and internal representations are a means to this end.

Some descriptions characterise perception as 'controlled hallucination' where stored knowledge is used to make top-down guesses concerning our incoming sensory stimuli (Clark, 2016; see also Frith, 2007). However, Clark insists that it is a mistake to see this as a form of "fantasy" or *merely* the brain's model or hypothesis. Moreover, it is something more than just indirect realism (cf. Tiehen, 2023). Rather, Clark highlights the importance of understanding how we are "embedded" within the world which "enables us to *see through the veil of surface statistics* to the world of distal interacting causes itself." (Clark, 2016, p. 170, emphasis in original). Clark's major disagreement with the view of perception as a form of hallucination (understood as fantasy or virtual reality) lies in Clark's strong commitment to embodied cognition and the idea that perception is essentially bound to action and physical constraints (e.g., the biological fact that I cannot bend my elbow backwards beyond a certain degree limits the possible actions I predict to do). Clark instead thinks of

³ Wiese (2015, p. 11), referring to Hohwy, also highlights the importance of the depth of a generative model for improving predictive control. Our discussion below brings forth an underlying rationale, that increased depth comes from increased environmental complexity.

perceptual inference as adaptive because it is shaped by the interaction between the agent's needs and the world's properties. The role of the agent is to be active in exploiting environmental structures according to its needs (e.g., the possible ways of moving an elbow), and moreover shaping them itself (e.g., building furniture to accommodate those needs – a process of niche-construction, see Bruineberg et al., 2018). There is no fundamental 'barrier' between agent and world.

2.3 Prediction and adaptive control

Many of these ideas originate from the cybernetics tradition, which provides us a more ecumenical perspective on the complexity of adaptive behaviour and the task of maintaining stable internal states in the face of uncertain, changing, environments. Furthermore, according to Heylighen and Joslyn (2001) there is a continuation where many fields today use perspectives introduced by cybernetics, sometimes without proper acknowledgement, and for example the fields of complex adaptive systems and artificial life research "seems to have taken over the cybernetics banner in its mathematical modelling of complex systems across disciplinary boundaries, however, while largely ignoring the issues of goal-directedness and control" (Heylighen and Joslyn, 2001, p. 157, emphasis added). PP's connection to cybernetics is even more obvious given the participative nature of perception and action in both PP and cybernetics (Wiese, 2015, pp. 2–3), but while PP focuses predominantly on the adjustment of internal brain processes and prediction of future states to meet goals, cybernetics emphasises the function of behavioural regulation via perception to service action, which is conditioned on the outside world. In doing so, cybernetics brings to the fore two key ideas.

The first idea is that adaptive control requires feedback. The cyberneticists suggested that in order to enable self-regulation and environmental-effects control, feedback is required in the form of information delivered to the system in focus. This mirrors the view we find in the PP framework of an agent constantly updating her internal model – through feedback – and then her taking action in the world as a result of prediction error minimisation on average and in the long run. So, both cybernetics and PP let feedback play a central role in how agents handle information processing in order to gain control. In particular, 'feedback' can be understood as one of the mechanisms underlying 'prediction error.' Other mechanisms include a sensor that can provide a measurement, and a comparator process that can yield the difference between the prediction and the measured actual state of the system which constitutes the error; some kind of effector process is also necessary such that the prediction error can be reduced.

For example, consider a simple artificial system like a temperature controller, with the thermoregulation process that goes on in a mammalian body. In the first case, the control system consists of a temperature sensor that yields an electrical voltage depending on temperature; this voltage is fed into a comparative amplifier along with a voltage that represents the desired temperature. In this context the

setpoint, or desired temperature, can be interpreted as a prediction. The difference between them is then fed into an effector, that is, something that can do work to change the temperature, such that the error is reduced. Now, in the case of a mammalian thermoregulation system, the temperature sensor is realised by neurons throughout the body that expresses temperature sensitive proteins. Signals from these neurons are fed into nuclei in the hypothalamus, which coarsely works as a comparator process. The difference between these signals is typically produced by means of inhibition, such that setpoint values inhibit incoming sensory signals. The resulting prediction error is then fed into nuclei that can engage a variety of effects throughout the body. For example, vasoconstriction or vasodilation in the skin can up- or down-regulate blood flow in tissues to conserve or dissipate heat. For larger prediction errors, muscle contractions in terms of shivering or sweating can respectably generate heat or dissipate larger amounts of it. In addition, mammals can engage in behaviours to move their bodies actively towards, or away from heat sources or -sinks. Feedback loops enable a system to update its internal model or actions in response to ongoing changes in the environment.

The second idea is that regulation, or control, is key to the system's capability to adapt to its environment. It is possible to characterise the process of prediction error minimisation in terms of the cyberneticist's notion of cognition as an intertwined relation between perception (categorisation), choice, and action. This expanded perspective on prediction error minimisation meets an embodied-PP view of the world as affording actions. On such a view, some prediction errors can only be minimised by discovering and exploiting particular affordances. Affordances are typically understood as 'opportunities for action' that an embodied cognitive system recognises in the world. In simple terms, this is perhaps most easily understood as something like 'graspability' and 'chewability.' That is, an ape can grasp, bite and chew a banana, but not a boulder. The boulder affords 'climbability' though.

Cognition on this view is considered part of dynamic interactions with the environment, where the recognition or discovery of affordances is highly relevant: a system must actively explore the world to discover how it can be employed in the process of allostasis⁴; these discoveries can then be exploited in the service of allostasis. Motivation for this is the idea that organisms must maintain homeostasis, and one way of achieving this is via information-processing. Von Uexküll (2001) provides an example of a tick predicting the availability of nutritious blood based on the presence of airborne chemicals (butyric acid) secreted by mammals; when the tick detects these chemicals, it releases its hold on the plant on which it is sitting, dropping down and, likely, landing on the animal. This process is risky (it will require significant energy expenditure for the tick to crawl back up), but it is

⁴ By 'allostasis,' we mean all physiological processes that predictively spend energy in maintaining the system's integrity, that is avoidance of decay or dissipation. Allostasis is an instance of a cybernetic process, that is a process that uses feedback and various kinds of effectors to maintain a set-point, e.g., body temperature.

statistically sufficiently successful to result in energy gain for procreation. In this case, the likelihood that the tick successfully lands on the animal will increase with an increasing accuracy in the tick's prediction of the chemical's source location.

The above reference to affordance discovery illustrates that allostasis also encompasses explorative processes, and explorative behaviours may be necessary to learn how a particular environmental niche affords the ingestion of energy rich foods or water, and e.g., the maintenance of a steady body temperature (e.g., finding shelter). This comes in addition to exploitation behaviours that have already been established through evolutionary biases or previous exploration (Stephens and Krebs, 1986). Predictive processes, and successful prediction in particular, thus can find a natural role in allostasis: e.g., the conservation of energy related to trial and error in affordance discovery (i.e., exploration) and -exploitation (Schulkin and Sterling, 2019).

In sum, cybernetics recommends two ingredients for adaptation: *feedback* and *control*. These ingredients are central to biological systems because they allow organisms to maintain homeostasis and effectively interact with their environment. In the following, we refine this more recent embodied narrative. Specifically, we rethink intelligent behaviour in biological systems as *adaptive control* coming out of a capacity of *accurate prediction*. The accuracy of the predictions in turn presupposes a form of embodiment or interaction with the outside world.

3 Sophisticating adaptive control through accurate prediction

3.1 Spatiotemporal depth

In complex systems, control requires prediction. Adjusting behaviour in response to direct feedback from the environment alone is insufficient, as the environmental conditions may be such that regulation requires the capacity to anticipate future possible outcomes. We consider this argument in two parts. The first part identifies the relevant complexity with the internal spatiotemporal depth of a hierarchically organised predictive system. The second part considers a more ecumenical perspective, where having internal depth is rational relative to the complexity of the environment that the biological system adapts to. Thus, while embodied PP continues ideas from cybernetics, its core appeal to 'prediction' offers a unique contribution to understanding the possibility of control in complex systems.

In such systems, predictions fulfil an important functional role: to simulate the consequences of possible actions and events, before they actually occur. Pezulo and Sims (2021, p. 7803) call this functional role "vicarious use before or in the absence of external events", and it is commonly known under "vicarious trial and error" in the neuroscience literature (see e.g. Redish, 2016). Having a capacity for prediction has been considered useful for systems that face complex practical

problems. Predictions, in this sense, may fulfil an important role for instrumental reasoning in both human and nonhuman animals as well as AI (Halina, 2021).⁵

Pezzulo and Sims argue that to fulfil this functional role, predictions must be generated autonomously, that is, not “determined or sustained by external stimuli” (Pezzulo and Sims, 2021, p. 7818). They suggest two ways in which this functional role of predictions can be realised: 1) by *variational free energy minimisation*, which requires a constant comparison between prior predictions and the direct perceptual effects in the light of continuously incoming sensory information, or 2) via *expected free energy minimisation*, which supports the selection of action policies based on future – i.e., purely anticipated and non-sensory – information.

Which strategy is appropriate to use depends on the complexity of the system at hand. In PP, ‘increasing complexity’ amounts to the ‘increased spatiotemporal depth’ that is accommodated by PP’s commitment to inferential *hierarchies* – it is the hierarchical layers of the predictive process that capture different degrees of spatiotemporal grain or levels of abstraction. According to both Clark (2013, 2016) and Hohwy (2013), spatiotemporally coarse-grained predictions – encoding information about large and slowly changing phenomena – occur higher upwards in the PP hierarchy, and spatiotemporally more fine-grained predictions – about small and rapidly changing phenomena – occur lower down in the PP hierarchy.⁶ On Sims and Pezzulo’s view, strategy (2) presupposes more complexity – in the form of spatiotemporal depth – to allow agents to use their internal generative models *offline* (i.e., decoupled from sensory input) for the representation of future observable (but unobserved) states. Only through increased spatiotemporal depth can organisms make decisions or choices to select between different possible future outcomes.

The advantage of thinking in foresight, in this view, is that it obtains higher flexibility because it is bound merely by what is possible, not by what is actually the case. This allows predictions to obtain more degrees of freedom due to their

⁵ Camp (2009) has independently argued that the possibility to utilise representations offline bears an adaptive advantage for animals because it makes it possible to entertain thoughts about *absent* states of affairs. She argues that conceptual thought is essentially stimulus-independent, and this allows the animal to engage with abstract representations that are not bound to the current environment or sensory context. Camp takes stimulus-independence to be (a) essential for flexible and abstract thinking, (b) a precondition for instrumental reasoning and (c) the best explanation of the complex problem-solving behaviour of intelligent animals such as New Caledonian crows. Additionally, Camp highlights that conceptual thought involves grasping the conditions under which the belief would turn out to be true or accurate, and this grasp requires the thinker to be decoupled from present sensory experiences. Camp’s much earlier contribution is worth noting since it independently supports, or at least is externally coherent with, the claim that there is a rational relation between the development of a capacity for offline-use of representations of possible states of appearances (what we may want to call predictions) and flexible, adaptive behaviour.

⁶ Although the notion of hierarchical levels is predominant in the predictive-mind literature, we think that the notion of predictive or anticipatory control is general enough to also apply to contexts in which system levels may be heterarchically organised (for discussion see Bechtel, 2022).

being further removed from the constraints governing the physical world, or from the agent's immediate perception-action cycle. Predictions thus enjoy greater freedom in the possible contents they can express and, in consequence, the varieties of inferences and counterfactuals to draw to prepare for action. This aligns with a common-sense understanding of intelligence as a matter of planning, foresight, and abstract, counterfactual thinking – capacities that form a dominant part of psychological tests of intelligence or cognitive ability (IQ). For example, the fourth version of the Wechsler Adult Intelligence Scale tests people on verbal comprehension, perceptual reasoning, working memory and processing speed, and the results in these cognitive domains tend to highly correlate with each other to form a 'general cognitive ability' (Deary, 2020). People are tested on their ability to say what two words have in common, to explain the meaning of a word, to answer questions about every-day life problems, or to detect patterns, numbers or letters, and match, arrange or complete them. A lot of these tasks involve instrumental reasoning skills, but many of them are verbal and abstract. In fact, the test tests precisely subjects' ability to abstract away and think forward in completing the task. Performing well on tests like this corresponds to a high 'g' score, which is the statistical correlation of performance. That is, if someone is good at solving one of these tasks, then they are likely to be good at solving another task in the test. While it remains unclear why 'g' occurs, the measure indicates an ability to think or reason towards the abstract.

It would be misleading, however, to identify biological intelligence purely with the complexity of internal computations (i.e., the capacity to predict with greater spatiotemporal depth) and resulting capacities for abstraction and generalisation of a system. As we argue next, in complex environments, control requires accurate transfer across contexts.

3.2 Accuracy

We have thus far highlighted the flexibility that is gained from being able to predict future states and abstract away from current perceptual stimuli. These capacities allow intelligent systems to adapt to their environments by representing and choosing among future possibilities. However, biological organisms must not only generate predictions in novel situations. They must also be able to accurately transfer those predictions across different contexts to increase their adaptive survival chances. This perspective helps us to refine the role of prediction, from its contribution to flexible thinking in foresight (i.e., vicarious trial and error), towards its contribution to efficient goal-directed behaviour in real-world contexts.

In an attempt to elucidate the processes that underlie goal-achievement in biological systems, Tjøstheim and Stephens approximate general intelligence as "the ability to abstract away context information, identify patterns, and transfer accurate predictions across contexts, as well as the ability to perform mental transformation and comparison" (Tjøstheim and Stephens, 2022, p. 476). This focus high-

lights that organisms' need to stay alive (stay in homeostasis) can be met to the extent that they can track changes in the environment by making accurate predictions. If they are able to accurately transfer their predictions between contexts they will be less "tied" to their niche. On the one hand they are freer to imagine possibilities in adjacent niches as indicated by Pezzulo and Sims (2021), on the other, as we argue here, they can make use of abstraction to spot commonalities and make use of the knowledge they already have. To illustrate, the New Caledonian crow demonstrates remarkable cognitive flexibility by using tools in novel ways across different contexts (Weir et al., 2002). Unlike many other animals that rely on instinctual behaviours within a narrow ecological niche, these crows can manufacture and modify tools to extract food from crevices, even improving upon previous designs. Experiments have shown that they can transfer their problem-solving skills to new situations, such as bending a wire to create a hook when no pre-made tool is available. This ability to abstract key features of a problem, recognise patterns across contexts, and apply learned solutions in novel environments exemplifies how abstraction and accurate prediction supports adaptive survival beyond rigid niche constraints. In this example, the pattern that food is partly hidden but retrievable with a tool can occur both in a forest in the form of grubs in trees, but also in a city in the form of food in garbage baskets – or in an experimental setup made by humans.

Accuracy, in this framework, can be defined as a composite of 'trueness' and 'precision'. The former refers to how close an average observation is to an accepted reference value, like the bullseye of a target. The latter, on the other hand, refers to how close repeated independent measures are to each other; this is like the tightness of repeated shots at the target. Applied to cognitive systems, trueness corresponds to the correctness of categorisation – how well a system identifies or classifies phenomena – while precision reflects the level of detail or specificity, represented by the narrowness of a statistical distribution in its predictions. Together, these dimensions capture the dual demands of categorisation accuracy and predictive detail that underpin intelligent behaviour. In terms of probabilistic models, trueness is the match between the mean value associated with the internal probability distribution given by the predictive model and the actual (objective chance) value in the world. The system can approximately assess the divergence between the physical/objective value and the mean of its own internal model distribution. This definition of trueness builds on the assumption that the system is equipped with reliable ways of detecting real values, i.e., being appropriately tuned to the environment statistics.

We understand the notion of generality as the ability to recognise commonalities across contexts. This allows predictions to be generated in novel environments by abstracting away surface detail from a familiar context and 'filling in' suitable details with the use of imagination (Tjøstheim and Stephens, 2022). On this view, the ability to abstract away from the immediate effects of perceptual stimuli, is key to achieving embodied control, particularly in novel situations, for the pur-

pose of keeping the waste of energy low. “Waste” here refers to the metabolic cost of using physical activity to solve problems by trial and error, compared to doing trial and error by mental simulation as discussed under “vicarious trial and error” in section 3.1. The typical experimental setup to test vicarious trial and error involves an animal positioned in a Y maze such that it has to choose whether to go left or right to get a reward. In the beginning the animal will usually explore each arm of the maze physically. But after it has formed an internal model of the task structure, it will begin to pause at the decision point (the point at which the rat stops to decide whether to go left or right), engaging in vicarious trial and error by mentally simulating the consequences of each choice before deciding. The measurement of the predictive process typically involves invasive measurement of the activities of hippocampal neurons. These neurons are typically described to “play motion forward”, becoming active as if the animal was actually moving (Redish, 2016). This shift in behaviour reflects the transition from reliance on overt exploration to internalised predictive modelling. Crucially, this optimises energy expenditure by reducing unnecessary movement. In more general terms, the ability to anticipate and plan actions based on internal models allows organisms to adapt efficiently to environmental challenges, reinforcing the link between spatiotemporal depth in representation and homeostatic regulation. In other words, the ability to anticipate environmental changes helps keep the waste of energy low because it reduces the need for constant updating through trial and error (which costs energy or computational power) and instead allows actions to be planned effectively to avoid unnecessary movement.

We see the rationality of accurate prediction as resulting from the assumption that the complexity associated with the internal organisation of the predictive system is mirrored in environmental complexity. Mobus and Kalton (2015, p. 374) point out three factors that “apply to all control mechanisms.” These are temporal factors (e.g., timing), factors governing the system’s ability to change (e.g., costs involved to affect change), and computational factors (e.g., complexity of computation and involved costs). Organisms need to interact with their environment in a manner that respects the flows that affect them. That is, the rationale for introducing accurate prediction is that it might not be possible to wait too long before acting, and one’s actions might need to be precise and true enough for them to actually work, or else acting might waste energy. Up to an individual threshold, systems might be able to cope with such waste, although it is suboptimal. For example, a foraging animal must predict not only the location and timing of food availability but also adjust its behaviour quickly enough to capture resources before a competitor does. A delay or choice of a suboptimal action can result in wasted energy. This balance between maintaining a complex predictive model (allowing accurate prediction) and executing timely, energy-efficient actions (allowing actual goal accomplishment) is thus necessary to drive adaptive control in dynamic environments.

An important aspect of reality that organisms need to tackle is that environments change. This is true both on shorter and longer timeframes. Something immediate can happen that the system must address, but there might also be a slower change in the overall environment that gradually changes what is beneficial to do. Organisms that live in less complex environments can afford to be less complex. While organisms that need to ‘handle’ multi-level stochastic processes, since their environments are prone to faster change, must themselves be more complex. This adaptiveness that is required of organisms *relative to their environmental niche* was central to cybernetics. We have viewed PP as expanding this approach by suggesting that it is the constant updating of the agent’s internal model that allows for successful – adaptive – action and thus survival. Viewing intelligent behaviour as adaptive control, coming out of a capacity for accurate prediction, thus complements internalist perspectives in highlighting not only features (e.g., temporal depth) of mental models, but also highlights the need for internal predictions to be appropriately (i.e., adaptively) related to environmental conditions. Interpreting intelligent behaviour in biological systems as accurate prediction highlights that it is, at least in part, the environmental conditions that determine which rewards and subsequent observations reward learners receive, and which response patterns will be selected for by the learning or adaptive process.

In sum, from the perspective of accurate prediction, adaptive behaviour consists not only in vicarious use of static stimuli, but more specifically in transferring predictions accurately across dynamically changing environmental contexts. That is, predictions are not just about what is next, but about what solutions might work across different scenarios. Key to this new perspective is the claim that adaptive behaviour requires a good balance between making *true* and *precise* predictions. This balance is necessary to avoid unnecessary movements or actions that might waste energy (metabolic cost), especially in environments that change over time. Accurate prediction is not only about predicting the future but doing so in a way that allows for efficient action under conditions of urgency as when foraging for food. Our view thus rethinks biological intelligence as a fundamentally relative property. It is an attribute that is not just associated with an individual agent where the depth of its generative models promotes the capacity for flexible, perceptually decoupled, prediction. Instead, it is inherently an attribute that must be evaluated in terms of how these predictions align with environmental conditions.⁷

⁷ Our approach is thus distinct from others that take prediction to be important for intelligence. Hutter (2021), for example, argues that a good compression method can help a system to represent informative patterns in the data. In our approach, adaptive behaviour requires more than compression; it also requires the capacity to represent *relationships* in the data. This aligns with neurobiological approaches to intelligence. Hawkins (2004, 2022), for example, emphasises that the brain encodes information not about single objects but about relationships among objects, and this requires the brain to situate itself relative to other objects and changing dynamics of the world it interacts with. Our relational view provides a broader rationale: processing information about environmental complexities requires a balance between generalisation (avoiding overfitting) and learning those features in the data that are relevant to a specific task and a specific

4 Artificial versus natural intelligence

While the notion of accurate prediction applies to biological intelligence in the first place, it can be useful for identifying possible limits of AI systems. Identifying such limits becomes relevant when simply scaling up the size of computational models and data does not seem promising to improve performance, especially on tasks that require sensorimotor competence (cf. Zador et al., 2023). Some AI researchers – for example, LeCun and colleagues at *Meta AI Research* – already claim that “[t]he essence of intelligence is the ability to predict” (Henaff et al., 2016, p. 1; LeCun et al., 2015). However, it remains unclear how exactly this claim should be understood.⁸ In the following, we illustrate how rethinking intelligence as accurate prediction accommodates more fine-grained distinctions between the intelligence expressed by biological systems in contrast to artificial systems.

Considering our insights from section 3.1, we would agree that prediction characterises aspects such as abstraction, foresight and planning well in artificial systems. This is because we may well draw analogies in terms of the functional role of predictions as fulfilled by the spatiotemporal depth achieved by hierarchical organisation between some biological (e.g., human-like) and artificial systems. Recent arguments support this analogy by highlighting that AI systems such as deep convolutional networks, due to their hierarchical organisation, are capable of performing transformational abstractions by convolution and max pooling—processes taken to resemble those performed by simple and complex cells in the mammalian visual cortex (Buckner, 2024, Chapter 3). However, while hierarchical organisation may be responsible for both prediction and transformative abstraction, other aspects of deep-learning systems reveal significant disanalogies to human-like intelligence. For example, according to Halina, AI systems such as AlphaGo, which relies on hierarchical Monte-Carlo tree-search, can “... transform a conceptual space in ways that do not appear available to human minds” (Halina, 2021, p. 316). A possible explanation is that the process of self-play (Samuel, 1959; Bansal et al., 2017) used when training AlphaGo allows it to explore combined areas of the game space that no single human would be likely to encounter; it is like collecting the lifetime experience of thousands of humans in one system, thus enabling superhuman game playing abilities. In this sense, there is a significant ‘brute force’ aspect of what AlphaGo does that, while exceptional, does not resemble human cognition well.

context in which that task must be carried out. For biological systems, the relevant features will be those affording action, and what affords action is sensitive to context.

⁸ Gamez (2021) also supports the idea that prediction is a core component of intelligence, both in natural and artificial systems. On this view, an accurate measure of a system’s predictive skill can effectively measure its intelligence. However, this account is insensitive to the fact that biological systems face different accuracy demands that do not hold for AI.

4.1 Generalisation at a cost

We highlighted in section 3.1 that intelligence in biological systems is a matter of the ecological conditions imposed on an agent. In section 3.2, we argued these conditions function as accuracy conditions that constrain how well predictions in biological systems can contribute to appropriate action. Action possibilities are constrained through the embodied structure of an organism and interactions with the environment. Crucially, this structure remains lacking in current AI systems. These systems are dependent in their accuracy on the quality of static training sets instead. Although natural systems can provide a valuable resource of insights and inspiration when designing and building artificial ones (Hassabis et al., 2017), they remain different in terms of their accuracy constraints.

Take AlexNet as an example. Perconti and Plebe (2020) illustrate how this deep convolutional network architecture vindicates the classical disembodied ideal of ‘pure’ vision. The pure-vision paradigm holds three core tenets: First, the visual system’s task is to create a detailed model of the world in front of the eyes. Second, the visual system achieves this task by processing information in a hierarchical manner, with each stage extracting increasingly specific features. Third, higher levels of processing depend on the information processed at lower levels, but not the other way around, and so lower levels operate independently of the higher-level interpretations. The pure-vision paradigm directly opposes the alternative paradigm of embodied and interactive vision, according to which “[t]he visual processing of objects and their attributes is driven by the kind of task the subject is performing, and object affordances are transformed into specific motor representations” (Perconti and Plebe, 2020, p. 6). AlexNet vindicates all three commitments of pure vision, and shares little with embodied and interactive vision. Firstly, the ImageNet benchmark on which the network is trained is organised according to the hierarchy of nouns in the lexical dictionary WordNet, and each lexical entry is associated with hundreds of static images. Secondly, convolutional layers are organised in a hierarchical way, so that earlier convolutions extract low-level features, which in turn become the input of other convolutions that extract features at progressively higher levels. Finally, AlexNet is feedforward, and so higher levels depend on lower levels in information processing, but not vice versa. In other words, the net learns purely from images and lexical descriptions of the category of objects in the image. AlexNet’s success in object classification is certainly not due to its embodiment or dynamic interaction with the environment.

More generally, the failures of such deep-neural network systems are a good illustration of our claim that having predictive depth and abilities for transformational abstraction, which we are willing to ascribe to them, are insufficient to recreate biological capacities as those illustrated by the human visual system. Mitchell (2019, pp. 120–123) describes a case where AlexNet correctly classified images into ‘contains an animal’ versus ‘contains no animal’, but after rigorous tests, her team found that the net failed on images that did not contain a blurry background. The net in fact was a ‘bokeh’ detector of a photographic artefact that

arises when zooming in on the target of a portrait, as opposed to a landscape. Illustrations like these support the hypothesis that transformative abstraction alone, while praised by many, is insufficient to obtain the level of accuracy in prediction that is comparable to the mammalian visual system.

Looking to an embodied approach, as opposed to pure prediction, makes a divergence between human-like biological and artificial systems expectable. In particular, it suggests that these systems are subject to different accuracy conditions, with biological systems underlying embodiment, interaction and thermodynamic constraints that AI is not subject to. The lack of embodiment in AI systems⁹ reveals that the process of training these models does not guarantee that predictions will be accurate outside the training sets used to develop AI systems. The embodied and interactive paradigm suggests one reason why such a guarantee cannot be given: disembodied AI systems do not couple with their environment in the way biological organisms do (sometimes referred to as ‘grip’ e.g. Bruineberg and Rietveld, 2014). That is, they are not constantly processing information from and interacting with the environment. AI systems instead produce an output for a discrete input and are inert afterwards. This is also true for systems such as AlphaGo, where self-play only applies to closed domains with preset rules. This static property of AI learners stands in stark contrast to biological brains, which constantly interact with the environment. Evolved mechanisms for prediction error minimisation allow biological brains to adapt and improve the accuracy of predictions only because biological brains are embedded in dynamically changing environments.

We suggest the concept of effective ‘re-concretisation’ as key to supporting adaptation to novel situations and contexts, allowing an organism to maintain “grip” (Bruineberg and Rietveld, 2014), or coupling, to the environment. Re-concretisation can be seen as part of a predictive cybernetic control mechanism. In this mechanism, previous experience aids prediction in novel situations by supplying situated and embodied concrete information. This information is obtained from sensory and motor sources through direct interaction with the world.¹⁰ Dynamic feedback information will in this case continuously refine predictions through free energy minimisation. Because AI systems do not interact dynamically and couple with their environment in the same way, they currently do not do re-concretisation well.

The human brain treats details differently, due to the way biological neural networks work. Biological neural networks can save information and learn at different levels of detail (Navon, 1977). They can then draw on these details in their knowledge base in order to re-concretise into new contexts. In human brains, this ability

⁹ The LLMs that control chatbots can be configured to be multimodal and predict both sensory information as well as motor sequences to control a robot. While this might make them, at least in some sense, embodied, the way in which they process multimodal information remains static. These LLM variants are referred to as Vision-Language-Action models (Brohan et al., 2023; Kim et al., 2024).

¹⁰ The coupling between organisms and environment means that they tend, on average, to make good enough predictions for surviving, at least over evolutionary time.

of re-concretisation may be mediated by anti-Hebbian learning (i.e., units inhibit each other instead of building and strengthening connection between each other which is the case in Hebbian learning) and sparse coding (Földiák, 1990). Anti-Hebbian learning, which in humans is important to store details, refers to neural units representing differences rather than similarities and produces de-correlated output. Joint activity of these units becomes less likely with time because the connection between neural units becomes progressively weaker. Sparse coding complements anti-Hebbian learning since it allows multiple objects to be represented with little overlap in the receptive field due to each neuron or feature detector responding selectively to a specific stimulus or pattern (i.e., the adequate stimuli to activate the neural unit are ‘sparse’), minimising redundancy in unit activation and interference between them (Olshausen and Field, 2004; Földiák and Young, 1995). This reduces representational overlap, enabling efficient storage and retrieval of distinct sensory sequences while preserving their individual characteristics.

Recently, LLMs make use of a “mixture of experts” design (Jacobs et al., 1991; Chen et al., 2022; Aoki et al., 2022), where a collection of specialised models collaborates to produce answers to users’ queries. This contrasts with earlier designs where one big, general model would do all the work by itself. In the case of such big models, the model averages over more information, allowing for greater generalisation. However, in this process particularities will inevitably be lost. In other words, by averaging over many knowledge domains it is possible to get valid abstractions and generalities that fit each domain to some approximation. However, the valuable details of those domains will be lost because they are not retrieved adequately from the model. More specifically, accuracy depends on a combination of several factors, like how effectively specific details are encoded in the model’s parameters alongside more general patterns; the ability of the model to recall the correct details for a particular input, rather than a more general pattern or a hallucination, and the model’s ability to correctly avoid overgeneralisation and recognise exceptional cases (outliers of a statistical distribution). This can be solved to some extent by increasing the number of parameters (Kaplan et al., 2020). Having lots of parameters affords having more specialised weights for particular cases, but it also increases the computational cost of running the model. The mixture of expert design is a different way of ameliorating the problem of lost details, but the need for this engineering solution indicates that generalisation comes at the cost of precision. Conversely, we hypothesise that deep accurate prediction, and with it adaptive control, becomes increasingly viable the more narrow the context, environment, or ecological niche is. This suggests that LLMs’ lack of accuracy despite having capabilities for abstraction, is due to a lack of grounding in concrete situations, and a lack of re-concretisation ability. Embodied experience, on the other hand, allows human-like biological systems to fill in details when making concrete predictions that complement abstract reasoning principles.

We can certainly develop a predictive system that uses an optimisation algorithm searching for the parameter values that give the least error. Nevertheless,

from the perspective we outlined here, there remains a crucial difference between this system and a biological counterpart, and this difference lies in the accuracy conditions on prediction. In particular, the minimisation of error in predictive AI systems remains unconstrained by thermodynamic factors or energetic cost. To illustrate, let's consider experiences with everyday physics. For example, a trained ballet dancer learning salsa for the first time could draw on their embodied experience with balance, timing, and movement coordination to allow them to pick up the new style quickly. An LLM, even if trained on vast amounts of text about dance theory, may describe the abstract principles of rhythm, foot placement, and partner connection. But it lacks the embodied sense of momentum, the tactile feedback of another dancer's lead, and the proprioceptive adjustments required to accurately predict how the body should move.

One may object that both natural selection and human design seem to have converged on the *behaviour* of 'next-token prediction', and the best explanation is that they have done so by means of recurrent networks (Whittington et al., 2021). But to understand the important differences between biology and technology it needs to be more strongly acknowledged that the accuracy of the predictions has, at least so far, very different teleological origins. In biology, it is ultimately the environment that determines what is accurate or not. That is, given a specific goal, the relevant environment will put pressure on the organism in focus to develop certain skills rather than others, for example, which abstractions are relevant and where to direct attention.

4.2 Embodied intelligence and accurate re-concretisation

The relevance of embodied capabilities for abstract reasoning tasks is more subtle in the contrast between human performance on standard tests like Raven's matrices (Figure 1) and the more recent Abstraction and Reasoning Corpus (ARC) (Chollet, 2019). The ARC consists of a number of grid-based tasks, where the grid ranges in size from 1x1 to 30x30 (Figure 2). The latest version at the time of writing is ARCV2 (Kamradt, 2025). This new version is particularly aimed at reasoning, or 'thinking' LLMs, collecting tasks that require visual manipulation and reasoning. The test now also places limitations on the amount of computational power participants can use since ARCV1 became less useful as a benchmark distinguishing AI from humans when OpenAI's resource-intensive AI services became strong enough to make the difference negligible. When attempting the ARCV1 or ARCV2 tests, the task is to identify the pattern in a set of demonstration instances and thereby complete a test instance correctly. Patterns consist of coloured grid squares, and success is determined by correctly predicting the colour of each square in the test instance (see Figure 2). The idea of the ARC tests is that the tasks are intuitively easy for humans but difficult for machines to solve. The contrast of human ease vs. computer difficulty is intended to reveal the details of what AI systems are currently missing in terms of human-like intelligence.

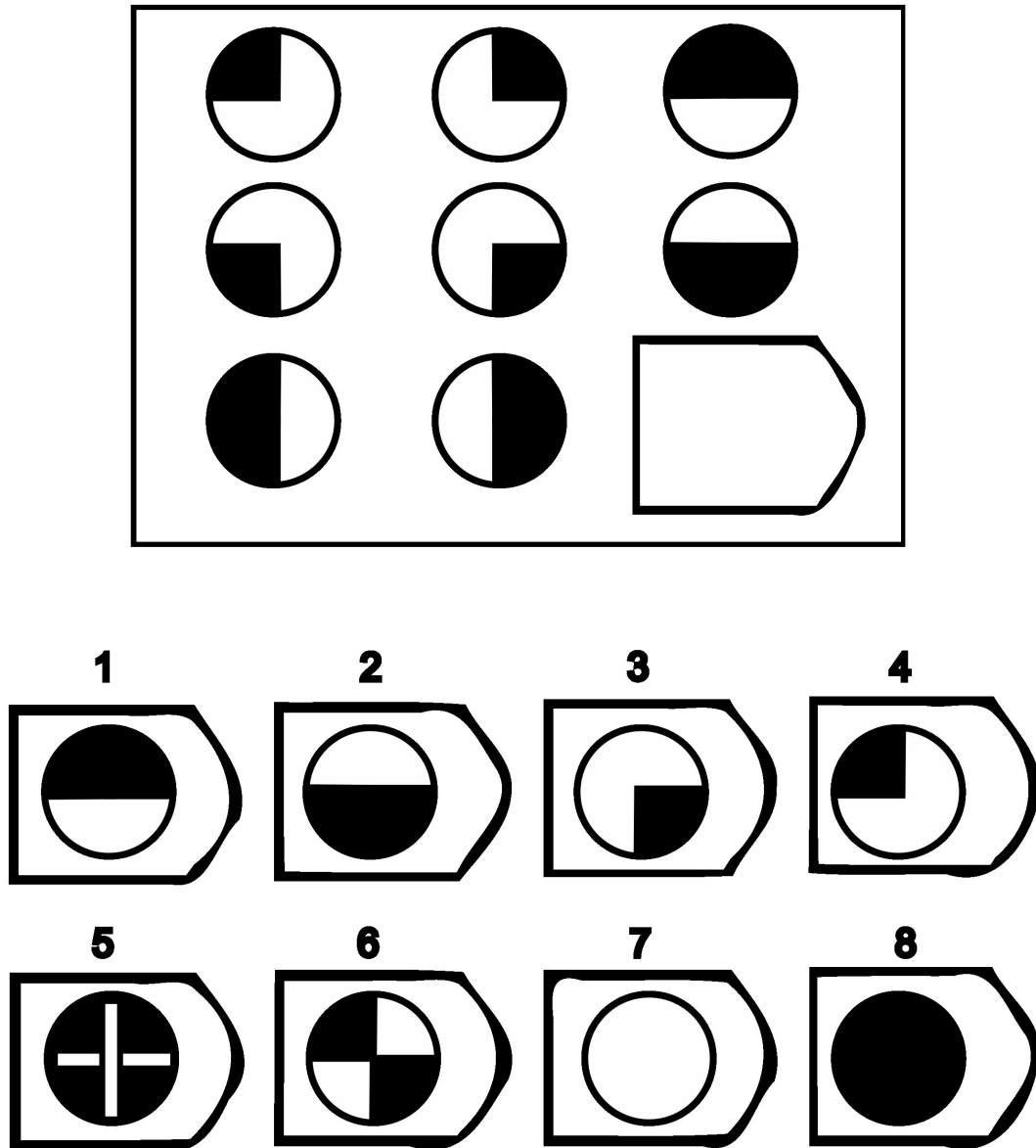


Figure 1: Example of Raven's matrices (adapted from Blair et al., 2005). The correct answer is no 8 in the lower right-hand corner, which is the combination of the two preceding circles.

Accurate prediction in this context pertains to first, identifying what the task is, second, identifying which actions are required to do the task, and third, identifying what the outcome should be given that those actions are successfully completed. For the ARC tests, the actions would be something like doing a comparison to identify the differences among instances of a given trial, and inferring an abstract pattern or rule from those differences to subsequently employ this rule to produce the prediction of the required outcome in the third instance of the task.

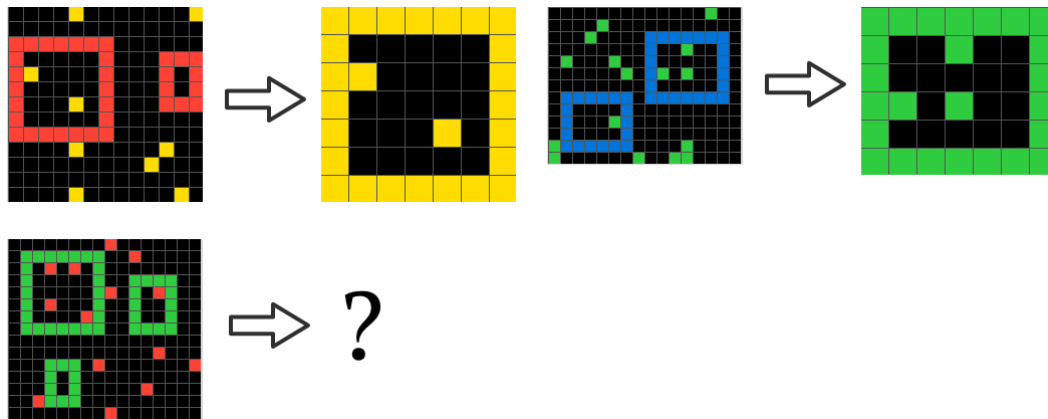


Figure 2: Example of the ARCv2 test, adapted from the ARCv2 training set. The correct answer involves counting the dots inside the frames, then copying the frame with the highest number, using the colour of the dots.

Crucially, that LLMs are ‘pretrained’ presupposes that there is a static aspect of the environment which stands in tension with the need for organisms to consider the dynamic changes of their environments. This presupposition reveals the subtleties alluded to above; one hypothesis of how humans are capable of performing the ARC tasks is that they rely on prior (subpersonal-level) predictions of what features are relevant for matching patterns in the visual array. For example, finding the overlap or commonality of abstract feature combinations of shapes per row requires *skilled expertise to effectively guide action*. By skill in this context we mean procedural knowledge, concerning both how to direct attention in terms of motor control of the eyes, but also cognitive skills involving making comparisons, and identifying promising differences. There is conceptual knowledge concerning what the problem is, for example specific factors such as knowing what is relevant foreground and irrelevant background, as well as more overarching knowledge concerning what comparison tasks such as these involve generally. There can also be factual knowledge about aspects of the problem such as that particular colours and shapes should be broken up and held apart. Evolved biological systems have gained the necessary competences for solving tasks like this through their adaptation to a visual world that requires the ability to direct attention, identify and work towards goals, and inhibit information that is irrelevant to those goals. In the human brain, these processes are mediated by areas including, coarsely, the occipital lobe for visual processing, the hippocampus for memory, and the pre-frontal areas for directing attention, inhibiting noise, as well as selecting arbitrary goals. Artificial systems, and LLMs in particular, draw their apparent ‘skills’ from patterns implicit in their training set. This means that to be able to do reliably well on tasks like those in the ARCv2 dataset, it is necessary for the system to have representative examples to identify statistical patterns. Further, LLMs require human feedback to build a probability distribution from those patterns that allows for

correct generalisation when encountering examples that are outside the training set.

One often overlooked aspect of abstract thinking and reasoning is that abstractions typically need to be made actionable to be of any practical use. That is, after having identified an abstract analogy, or pattern, actions and outcomes need to be *re-concretised* into the situation at hand. In other words, details need to be filled in to make predictions accurate for particular cases. Biological systems have a high capacity for filling in accurate detail in concrete situations, as evidenced in the example of New Caledonian crows above. To use the language of cybernetics again, a predictive controller with memory and a model of the world can use previous experience and knowledge to scaffold predictions in novel contexts. Sensory feedback would then serve to continuously refine predictions in the unfamiliar environment by means of free energy minimisation. For example, Northwestern crows have been observed to use different methods to hide food from other crows, including rock sides, grass clumps, and river banks (James and Verbeek, 1983); this could be viewed as re-concretising the pattern of ‘food occlusion’ into a variety of contexts. However, as explained in section 4.1, artificial systems tend to lose details and particulars in their training process. This makes re-concretisation arbitrary. That is, there are no constraints in the system to rule out inaccurate detail in the generation process. In practice, constraints are added during post-training through reinforcement learning with human feedback (RLHF), where humans work to weed out inaccuracies and promote correct answers by giving the system a ‘thumbs up’ or ‘thumbs down’. This is problematic if the AI system is supposed to be an autonomous system since they will not be able to translate their skill sets to novel contexts¹¹.

To summarise this section, we first (Section 4.1) looked at what the notion of accuracy means in the case of AI and in biology. For biological systems, the accuracy of predictions is afforded and constrained by their coupling with the environment and energy costs involved in minimising prediction errors. In contrast, the predictive accuracy of a deep neural network depends on the quality of its training set. To avoid inaccurate generalisation, care must be taken to include enough variety in the data to allow the optimisation algorithm to infer the correct categories. AI systems like LLMs with their huge training sets can generalise across different contexts, but the problem for them is to retrieve and correctly apply valuable detail from the training data. Such detail remains necessary to generate accurate predictions in particular cases. More specifically, accuracy depends on several factors, like how effectively specific details are encoded in the model’s parameters alongside more general patterns; and crucially, the ability of the model to recall the correct details for a particular input, rather than a more general pattern or a

¹¹ That is not to say that biological systems aren’t limited too; they are. In completely novel situations animals will get confused and come up short. Nevertheless, they can usually draw on previous experiences that at least share some similarities with the situation at hand; completely novel situations remain uncommon.

hallucination. Second (Section 4.2), we have suggested that ‘re-concretisation’ – the ability to translate abstract knowledge into concrete solutions – is necessary for accurate prediction when using abstractions to solve novel problems. To test the abstract reasoning powers of AI, the ARC tests work as kinds of IQ tests for machines. Though AIs so far have to train on examples of the ARC tests, humans use sparse coding and other mechanisms to conserve useful details from experience. These aspects of sparse coding might be key for allowing humans to do re-concretisation of abstract patterns into new situations with more accuracy than AI.

5 Concluding remarks

Taken together, we have argued for a PP-view of biological intelligence as a form of adaptive control achieved by performing accurate prediction. This view emphasises the importance of embodied prediction, and the need for organisms to rely on their interactions with the environment to optimise their actions and conserve energy. The key feature of accurate predictions is that they allow organisms with those needs to accurately transfer predictions across different contexts and thus effectively prepare for energy-optimising action and control the anticipated environmental effects. This view highlights an important contrast between biological and artificial intelligence. In particular, one way of achieving adaptive control is through a process of *re-concretisation*, where a learned abstraction is reified into a novel situation through the direct experience with the world that embodiment affords. Under the plausible assumption that AI models lack the relevant constraints from worldly interactions, the view provides a rationale for why AI models often generalise at the expense of detail, while biological systems can tailor predictions to specific environments and changes over time. Thus, while prediction is important for adaptive behaviour, control additionally relies on embodied models that align predictions with real-world constraints.

Acknowledgements

We would like to thank two anonymous reviewers and the editor of this journal for immensely helpful comments that improved the quality of this paper. All mistakes are our own.

References

- Aoki, R. Y., Tung, F., & Oliveira, G. L. (2022). Heterogeneous multi-task learning with expert diversity. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 19(6), 3093–3102. <https://doi.org/10.1109/TCBB.2022.3175456>
- Bansal, T., Pachocki, J., Sidor, S., Sutskever, I., & Mordatch, I. (2017). Emergent complexity via multi-agent competition. *arXiv preprint, arXiv:1710.03748*. <https://doi.org/10.48550/arXiv.1710.03748>

Poth, N. L., Tjøstheim, T. A., & Stephens, A. (2025). Rethinking intelligent behaviour through the lens of accurate prediction: Adaptive control in uncertain environments. *Philosophy and the Mind Sciences*, 6. <https://doi.org/10.33735/phimisci.2025.11780>



- Bechtel, W. (2022). Levels in biological organisms: Hierarchy of production mechanisms, heterarchy of control mechanisms. *The Monist*, 105(2), 156–174. <https://doi.org/10.1093/monist/onab029>
- Beni, M. D. (2018). The reward of unification: A realist reading of the predictive processing theory. *New Ideas in Psychology*, 48, 21–26. <https://doi.org/10.1016/j.newideapsych.2017.10.001>
- Blair, C., Gamson, D., Thorne, S., & Baker, D. (2005). Rising mean IQ: Cognitive demand of mathematics education for young children, population exposure to formal schooling, and the neurobiology of the prefrontal cortex. *Intelligence*, 33(1), 93–106. <https://doi.org/10.1016/j.intell.2004.07.008>
- Brohan, A., Brown, N., Carbajal, J., Chebotar, Y., Chen, X., Choromanski, K., Ding, T., Driess, D., Dubey, A., Finn, C., Florence, P., Fu, C., Arenas, M. G., Gopalakrishnan, K., Han, K., Hausman, K., Herzog, A., Hsu, J., Ichter, B., ... Zitkovich, B. (2023). Rt-2: Vision-language-action models transfer web knowledge to robotic control. *arXiv preprint*, arXiv:2307.15818. <https://doi.org/10.48550/arXiv.2307.15818>
- 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>
- Bruineberg, J., & Rietveld, E. (2014). Self-organization, free energy minimization, and optimal grip on a field of affordances. *Frontiers in Human Neuroscience*, 8(599), 1–14. <https://doi.org/10.3389/fnhum.2014.00599>
- Bubic, A., Von Cramon, D. Y., & Schubotz, R. I. (2010). Prediction, cognition and the brain. *Frontiers in Human Neuroscience*, 4(25). <https://doi.org/10.3389/fnhum.2010.00025>
- Buckner, C. J. (2024). *From deep learning to rational machines: What the history of philosophy can teach us about the future of artificial intelligence*. Oxford University Press.
- Camp, E. (2009). Putting thoughts to work: Concepts, systematicity, and stimulus-independence. *Philosophy and Phenomenological Research*, 78(2), 275–311. <https://doi.org/10.1111/j.1933-1592.2009.00245.x>
- Chen, Z., Shen, Y., Ding, M., Chen, Z., Zhao, H., Learned-Miller, E. G., & Gan, C. (2022). Mod-Squad: Designing mixture of experts as modular multi-task learners. *arXiv*, abs/2212.08066. <https://doi.org/10.48550/arXiv.2212.08066>
- Chollet, F. (2019). On the measure of intelligence. *arXiv preprint*, arXiv:1911.01547. <https://doi.org/10.48550/arXiv.1911.01547>
- Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford University 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. (2016). *Surfing uncertainty: Prediction, action, and the embodied mind*. Oxford University Press.
- Craik, K. J. W. (1944). The nature of explanation. *Philosophy*, 19(73), 173–174. <https://doi.org/10.1038/153605a0>
- Deary, I. J. (2020). *Intelligence: A very short introduction* (Vol. 39). Oxford University Press.
- Figdor, C. (2021). Shannon + Friston = content: Intentionality in predictive signaling systems. *Synthese*, 199(1), 2793–2816. <https://doi.org/10.1007/s11229-020-02912-9>
- Földiák, P. (1990). Forming sparse representations by local anti-Hebbian learning. *Biological Cybernetics*, 64, 165–170. <https://doi.org/10.1007/BF02331346>
- Földiák, P., & Young, M. P. (1995). Sparse coding in the primate cortex. In M. A. Arbib (Ed.), *The handbook of brain theory and neural networks* (pp. 895–898). The MIT Press.
- Friston, K. J. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences*, 360(1456), 815–836. <https://doi.org/10.1098/rstb.2005.1622>
- 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., Daunizeau, J., Kilner, J., & Kiebel, S. J. (2010). Action and behavior: A free-energy formulation. *Biological Cybernetics*, 102, 227–260. <https://doi.org/10.1007/s00422-010-0364-z>
- Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese*, 159, 417–458. <https://doi.org/10.1007/s11229-007-9237-y>
- Frith, C. (2007). *Making up the mind: How the brain creates our mental world*. John Wiley & Sons.
- Froese, T., & Ziemke, T. (2009). Enactive artificial intelligence: Investigating the systemic organization of life and mind. *Artificial Intelligence*, 173(3/4), 466–500. <https://doi.org/10.1016/j.artint.2008.12.001>
- Gamez, D. (2021). Measuring intelligence in natural and artificial systems. *Journal of Artificial Intelligence and Consciousness*, 8(2), 285–302. <https://doi.org/10.1142/S2705078521500090>
- Geary, D. C. (2009). The evolution of general fluid intelligence. In S. M. Platek & T. K. Shackelford (Eds.), *Foundations in evolutionary cognitive neuroscience* (pp. 22–56). Cambridge University Press.
- Gładziejewski, P. (2016). Predictive coding and representationalism. *Synthese*, 193, 559–582. <https://doi.org/10.1007/s11229-015-0762-9>

Poth, N. L., Tjøstheim, T. A., & Stephens, A. (2025). Rethinking intelligent behaviour through the lens of accurate prediction: Adaptive control in uncertain environments. *Philosophy and the Mind Sciences*, 6. <https://doi.org/10.33735/phimisci.2025.11780>



- Gładziejewski, P. (2019). Mechanistic unity of the predictive mind. *Theory & Psychology*, 29(5), 657–675. <https://doi.org/10.1177/0959354319866258>
- Halina, M. (2021). Insightful artificial intelligence. *Mind & Language*, 36(2), 315–329. <https://doi.org/10.1111/mila.12321>
- Hassabis, D., Kumaran, D., Summerfield, C., & Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2), 245–258. <https://doi.org/10.1016/j.neuron.2017.06.011>
- Hawkins, J. (2004). *On intelligence*. Henry Holt; Company.
- Hawkins, J. (2022). *A thousand brains: A new theory of intelligence*. Basic Books.
- Henaff, M., Weston, J., Szlam, A., Bordes, A., & LeCun, Y. (2016). Tracking the world state with recurrent entity networks. *ArXiv preprint arXiv:1612.03969*. <https://doi.org/10.48550/arXiv.1612.03969>
- Heylighen, F., & Joslyn, C. (2001). Cybernetics and second-order cybernetics. *Encyclopedia of Physical Science & Technology*, 4, 155–170. <https://doi.org/10.1016/B0-12-227410-5/00161-7>
- Hohwy, J. (2013). *The predictive mind*. Oxford University Press.
- Hutter, M. (2021). The Hutter Prize [<http://prize.hutter1.net/>].
- Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. *Neural Computation*, 3, 79–87. <https://doi.org/10.1162/neco.1991.3.1.79>
- James, P. C., & Verbeek, N. A. (1983). The food storage behaviour of the northwestern crow. *Behaviour*, 85, 276–290. <https://www.jstor.org/stable/4534267>
- Kamradt, G. (2025). ARC-AGI-2 + ARC Prize 2025 is live! [<https://arcprize.org/blog/announcing-arc-agi-2-and-arc-prize-2025>].
- Kaplan, J., McCandlish, S., Henighan, T., Brown, T. B., Chess, B., Child, R., Gray, S., Radford, A., Wu, J., & Amodei, D. (2020). Scaling laws for neural language models. *ArXiv preprint arXiv:2001.08361*. <https://doi.org/10.48550/arXiv.2001.08361>
- Kiefer, A., & Hohwy, J. (2018). Content and misrepresentation in hierarchical generative models. *Synthese*, 195, 2387–2415. <https://doi.org/10.1007/s11229-017-1435-7>
- Kim, M. J., Pertsch, K., Karamcheti, S., Xiao, T., Balakrishna, A., Nair, S., Rafailov, R., Foster, E., Lam, G., Sanketi, P., Vuong, Q., Kollar, T., Burchfiel, B., Tedrake, R., Sadigh, D., Levine, S., Liang, P., & Finn, C. (2024). Openvla: An open-source vision-language-action model. *ArXiv preprint arXiv:2406.09246*. <https://doi.org/10.48550/arXiv.2406.09246>
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, 27(12), 712–719. <https://doi.org/10.1016/j.tins.2004.10.007>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- Lefkowitz, I. (1966). Multilevel approach applied to control system design. *ASME Journal of Basic Engineering*, 88(2), 392–398. <https://doi.org/10.1115/1.3645868>
- Legg, S., & Hutter, M. (2007). A collection of definitions of intelligence. *Frontiers in Artificial Intelligence and Applications*, 157, 17–24. <https://doi.org/10.48550/arXiv.0706.3639>
- Mesarovic, M. D. (1970). Multilevel systems and concepts in process control. *Proceedings of the IEEE*, 58(1), 111–125. <https://doi.org/10.1109/PROC.1970.7545>
- Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Penguin Random House UK.
- Mobus, G. E., & Kalton, M. C. (2015). *Principles of systems science*. Springer.
- Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3), 353–383. [https://doi.org/10.1016/0010-0285\(77\)90012-3](https://doi.org/10.1016/0010-0285(77)90012-3)
- Olshausen, B. A., & Field, D. J. (2004). Sparse coding of sensory inputs. *Current Opinion in Neurobiology*, 14(4), 481–487. <https://doi.org/10.1016/j.conb.2004.07.007>
- Parr, T., Da Costa, L., & Friston, K. J. (2020). Markov blankets, information geometry and stochastic thermodynamics. *Philosophical Transactions of the Royal Society A*, 378(2164), 20190159. <https://doi.org/10.1098/rsta.2019.0159>
- Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Morgan Kaufmann.
- Perconti, P., & Plebe, A. (2020). Deep learning and cognitive science. *Cognition*, 203, 104365. <https://doi.org/10.1016/j.cognition.2020.104365>
- Pezzulo, G., & Sims, M. (2021). Modelling ourselves: What the free energy principle reveals about our implicit notions of representation. *Synthese*, 199(3–4), 7801–7833. <https://doi.org/10.1007/s11229-021-03140-5>
- PiekarSKI, M. (2023). Incorporating (variational) free energy models into mechanisms: The case of predictive processing under the free energy principle. *Synthese*, 202(58), 1–33. <https://doi.org/10.1007/s11229-023-04292-2>
- Poth, N. (2022). Schema-centred unity and process-centred pluralism of the predictive mind. *Minds and Machines*, 32(3), 433–459. <https://doi.org/10.1007/s11023-022-09595-w>

Poth, N. L., Tjøstheim, T. A., & Stephens, A. (2025). Rethinking intelligent behaviour through the lens of accurate prediction: Adaptive control in uncertain environments. *Philosophy and the Mind Sciences*, 6. <https://doi.org/10.33735/phimisci.2025.11780>



- Rajan, K., & Saffiotti, A. (2017). Towards a science of integrated AI and robotics. *Artificial Intelligence*, 247, 1–9. <https://doi.org/10.1016/j.artint.2017.03.003>
- Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1), 79–87. <https://doi.org/10.1038/4580>
- Redish, A. D. (2016). Vicarious trial and error. *Nature Reviews Neuroscience*, 17(3), 147–159. <https://doi.org/10.1038/nrn.2015.30>
- Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of Research and Development*, 3(3), 210–229. <https://doi.org/10.1147/rd.33.0210>
- Schulkin, J., & Sterling, P. (2019). Allostasis: A brain-centered, predictive mode of physiological regulation. *Trends in Neurosciences*, 42(10), 740–752. <https://doi.org/10.1016/j.tins.2019.07.010>
- Seth, A. K. (2014). A predictive processing theory of sensorimotor contingencies: Explaining the puzzle of perceptual presence and its absence in synesthesia. *Cognitive Neuroscience*, 5(2), 97–118. <https://doi.org/10.1080/17588928.2013.877880>
- Seth, A. K. (2015). The cybernetic Bayesian brain: From interoceptive inference to sensorimotor contingencies. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (Vol. 35(T)). MIND Group. <https://doi.org/10.15502/9783958570108>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Sprevak, M. (2020). Two kinds of information processing in cognition. *Review of Philosophy and Psychology*, 11(3), 591–611. <https://doi.org/10.1007/s13164-019-00438-9>
- Sprevak, M. (2024). Predictive coding I: Introduction. *Philosophy Compass*, 19(1), e12950. <https://doi.org/10.1111/phc3.12950>
- Sprevak, M., & Smith, R. (2023). An introduction to predictive processing models of perception and decision-making. *Topics in Cognitive Science*, 1–28. <https://doi.org/10.1111/tops.12704>
- Stephens, D. W., & Krebs, J. R. (1986). *Foraging theory* (Vol. 6). Princeton University Press.
- Tiehen, J. (2023). Perception as controlled hallucination. *Analytic Philosophy*, 64(4), 355–372. <https://doi.org/10.1111/phib.12268>
- Tjøstheim, T. A., & Stephens, A. (2022). Intelligence as accurate prediction. *Review of Philosophy and Psychology*, 13(2), 475–499. <https://doi.org/10.1007/s13164-021-00538-5>
- von Uexküll, J. (2001). The new concept of Umwelt: A link between science and the humanities. *Semiotica*, 134(1/4), 111–123. <https://doi.org/10.1515/semi.2001.018>
- Weir, A. A., Chappell, J., & Kacelnik, A. (2002). Shaping of hooks in New Caledonian crows. *Science*, 297(5583), 981–981. <https://doi.org/10.1126/science.1073433>
- Whittington, J. C., Warren, J., & Behrens, T. E. (2021). Relating transformers to models and neural representations of the hippocampal formation. *arXiv preprint arXiv:2112.04035*. <https://doi.org/10.48550/arXiv.2112.04035>
- Wiener, N. (1948). *Cybernetics: Control and communication in the animal and the machine*. Wiley.
- Wiese, W. (2015). Perceptual presence in the Kuhnian-Popperian Bayesian brain: A commentary on Anil K. Seth. In T. Metzinger & J. M. Windt (Eds.), *Open MIND* (Vol. 35(C)). MIND Group. <https://doi.org/10.15502/9783958570207>
- Wiese, W., & Metzinger, T. (2017). Vanilla PP for philosophers: A primer on predictive processing. In *Philosophy and predictive processing* (Vol. 1). MIND Group. <https://doi.org/10.15502/9783958573024>
- Zador, A., Escola, S., Richards, B., Ölveczky, B., Bengio, Y., Boahen, K., Botvinick, M., Chklovskii, D., Churchland, A., Clopath, C., DiCarlo, J., Ganguli, S., Hawkins, J., Körding, K., Koulakov, A., LeCun, Y., Lillicrap, T., Marblestone, A., Olshausen, B., ... Tsao, D. (2023). Catalyzing next-generation Artificial Intelligence through NeuroAI. *Nature Communications*, 14, 1597. <https://doi.org/10.1038/s41467-023-37180-x>

Open Access

This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Poth, N. L., Tjøstheim, T. A., & Stephens, A. (2025). Rethinking intelligent behaviour through the lens of accurate prediction: Adaptive control in uncertain environments. *Philosophy and the Mind Sciences*, 6. <https://doi.org/10.33735/phimisci.2025.11780>



©The author(s). <https://philosophymindscience.org> ISSN: 2699-0369