Trends in Neurosciences

CellPress

Opinion

Useful misrepresentation: perception as embodied proactive inference

Joshua M. Martin,[1,2] Mark Solms,[3] and Philipp Sterzer[2,*]

According to the predictive processing framework, perception is geared to represent the environment in terms of embodied action opportunities as opposed to objective truth. Here, we argue that such an optimisation is reflected by biases in expectations (i.e., prior predictive information) that facilitate 'useful' inferences of external sensory causes. To support this, we highlight a body of literature suggesting that perception is systematically biased away from accurate estimates under conditions where utility and accuracy conflict with one another. We interpret this to reflect the brain's attempt to adjudicate between conflicting sources of prediction error, as external accuracy is sacrificed to facilitate actions that proactively avoid physiologically surprising outcomes. This carries important theoretical implications and offers new insights into psychopathology.

Highlights

Converging lines of evidence place allostatic physiological regulation as a core imperative of perception. From the perspective of predictive processing, we evaluate how such an optimisation is reflected by biases in expectations underlying perception.

Within perceptual inference, expectations and attention are often assigned functionally orthogonal roles, whereby attention alters the gain on sensory signals based upon their estimated utility, while expectations bias perception according to regularities in the environment.

We argue that expectations are utility driven, biasing perceptual inference towards useful outcomes that may be inaccurate but, importantly, facilitate actions that preemptively avoid biologically unfavourable regulatory conditions. Internal models are thus structured, tuned, and updated in ways that usefully misrepresent the world.

Better safe than sorry: adaptive biases in perception

Imagine walking in a forest and coming across a seemingly ambiguous object lying on the ground, for instance a thin, elongated, and curved dark shape. Upon encountering this object, the task of your brain is to infer what it is so that it can formulate an appropriate behavioural response (Figure 1). According to **predictive processing (PP)** (see Glossary), perception of this object will entail a process of approximate Bayesian inference [1–3], which combines sensory evidence ('likelihood') and background knowledge ('prior probability'). What determines perceptual experience then is the brain's current 'best explanation' of the available sensory evidence, that is, the hypothesis that 'makes the best predictions and that, taking priors into consideration, is consequently assigned the highest posterior probability' [4]. For our purposes, let us imagine that your brain formulates two plausible hypotheses regarding the identity of the ambiguous object: (i) it is a stick or (ii) it is a snake [5]. Let us also assume that the sensory evidence fits these hypotheses equally well (i.e., their likelihood is equal). Which one should you perceive?

Given that you have probably encountered more sticks than snakes over the course of your life (and the same may apply to our species over evolutionary timescales), one interpretation would be that you should perceive the object as a stick since it represents the available hypothesis that, based on your prior knowledge, is the most likely to be correct. But what if you are wrong? That is, what if you perceive (infer) that the object is a stick and it actually turns out to be a snake or vice versa? The very nature of probabilistic inference entails uncertainty and thus inevitable inferential errors. These misperceptions are of practical importance since their outcomes determine not only how we represent our environment but also how we interact with it. While all types of error are undesirable, some are costlier than others; mistaking a stick for a snake may cost you unnecessary evasive action, while mistaking a snake for a stick could cost one their life [5]. Considerations of this sort reflect the **utility** of a perceptual decision determined by the expected benefits or costs of possible outcomes. Once this additional information is taken into account, one may claim that even though the stick is the more accurate prediction, you

[1]Berlin School of Mind and Brain, Faculty of Philosophy, Humboldt-Universität zu Berlin, Berlin, Germany
[2]Department of Psychiatry and Psychotherapy, Charité–Universitätsmedizin Berlin, Campus Charité Mitte, Berlin, Germany
[3]Department of Psychology, University of Cape Town, Cape Town, South Africa

*Correspondence:
philipp.sterzer@charite.de (P. Sterzer).

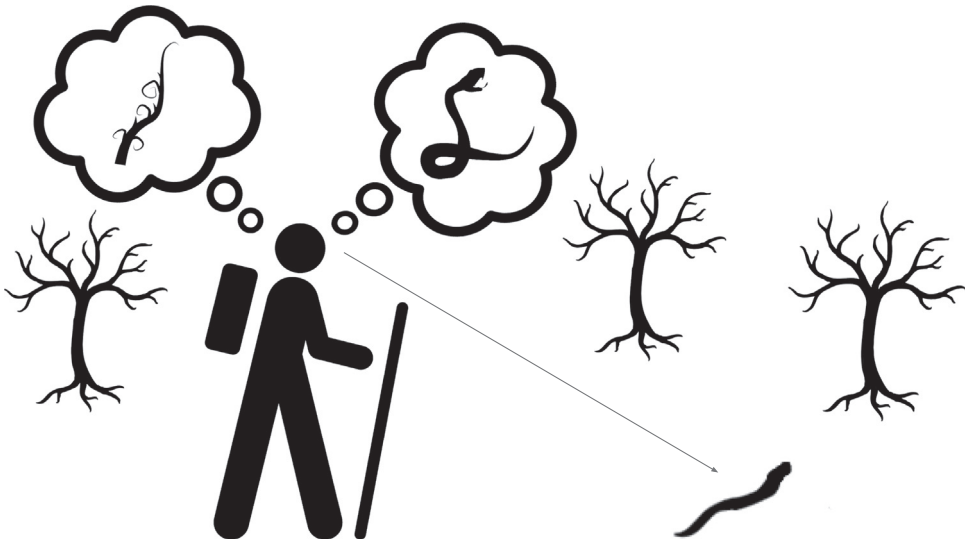

Figure 1. Stick or snake? Perceptual experience is thought to reflect a process of inference, or the brain's 'best guess' of what is out there, which combines sensory evidence and prior knowledge. Under conditions in which sensory evidence is ambiguous or fits competing perceptual predictions equally well, prior knowledge may act as a disambiguating factor in determining perception. For example, a hiker may perceive an ambiguous elongated object to be a stick or a snake depending upon their prior knowledge. This article explores how prior knowledge should be optimised to predict the world in a way that efficiently regulates physiological needs. A key question in this regard is how the brain should choose between *a priori* probable predictions (those that are more likely to be correct) as opposed to biologically useful ones (those that promote favourable consequences). Applied to the earlier scenario, should the hiker be biased to perceive this ambiguous object as a stick (more probable) or a snake (more useful)?

should 'play it safe' and rather perceive the snake since, on balance, it represents the more useful prediction that minimizes the chance of a costly outcome [6,7].

At first glance, PP would seem to favour the first interpretation (sticks over snakes), since it more closely resembles the way prior probability is used in Bayesian statistics (e.g., 0.5 prior probability for a tails on any given coin toss). However, PP is also consistent with the second interpretation (snakes over sticks) if we consider the notion of probability from the perspective of **active inference**, where probability and utility are taken together (see Box 1 for an overview). In this paper, we explore the extent to which prior predictive information is utility driven. To do this, we first outline how perception emerges within a PP framework, emphasising its embodied and proactive nature. While expectations are often assumed to be probability driven, we then draw upon insights and empirical evidence from error management theory [6,7] and utilized signal detection theory [8] to argue that they are instead utility driven, inferentially biased towards useful outcomes. Finally, we accommodate such biases within the wider imperative of prediction error minimisation, highlighting their ability to minimise uncertainty in relation to key physiological variables.

Active inference, embodied prediction, and pragmatic perceivers

Biological agents are characterised by the ability to maintain themselves within a limited range of states that are compatible with continued survival [9]. This can be described in information-theoretical terms, where organisms preferentially occupy a set of high-probability states through minimising 'surprising' sensory exchanges with the environment. According to PP, this is achieved via the deployment of a hierarchical generative model, which minimises the prediction error (or average **surprisal**) in its sensory input over time. Such a model includes predictions concerning sensory input deriving both from the outside world and its internal milieu [10].

Glossary

Active inference: a framework for incorporating the role of action into the PP framework. While perception functions to change one's predictions of the world to accommodate the sensory evidence, action functions to change the sensory input to fulfil these predictions. Action can thus be cast as a type of 'self-fulfilling prophecy' where we actively engage with and sample the world to bring about the very predictions that our internal model deploys.

Allostasis: the process of preserving physiological stability through regulating bodily needs via autonomic processes and adaptive behaviour. Allostasis ('stability through change') is proposed to be different from homeostasis ('stability through constancy'), as the desirable parameter values are not fixed but rather are proactively determined and vary according to changing demands. For example, the optimal set-point for heart rate may vary according to the state of the organism (e.g., whether she is asleep, resting, or in danger).

Dark Room Problem: a philosophical thought experiment that questions whether the unifying principle underpinning PP can accommodate curiosity and desire-driven behaviour. If creatures like ourselves are mandated to fulfil predictions (to minimise prediction error), then why do we not seek out and remain in dark rooms where the incoming sensory input is highly predictable? This remains a source of controversy in current literature.

Precision: in the context of PP, precision refers to estimates of second-order statistics reflecting the reliability of a given information source (statistically the inverse of variance). Information that is expected to be precise will be weighted more highly in the formulation of inferences. These precision estimates may be reflected at both higher ('prior precision') and lower ('sensory precision') levels of the cortical hierarchy, allowing for the brain to flexibly modulate the relative influence of sensory input in relation to prior beliefs in a context-dependent manner.

Predictive processing (PP): an influential computational framework in neuroscience that provides a unifying principle for perception, action, and cognition. Importantly, PP is agnostic to underlying neurophysiological implementation. Thus, it is distinguished

Box 1. Bayesian decision theory and active inference

Perhaps the most widely used framework for modelling the influence of hypothesis utility in optimal decision-making is that of Bayesian decision theory (BDT) [79]. BDT assumes that for every decision available to an agent, there are differences in the expected consequences that are contingent upon the true state of the world. These consequences can be suitably modelled through the inclusion of a utility function (alternatively a gain or loss function), which explicitly models the expected benefits and costs associated with various outcomes. Under such conditions, the optimal decision derives from the outcome that maximizes expected utility (or minimizes risk) based upon a convolution (in the mathematical sense) of probability and utility estimates. From this perspective, accurate decisions are not always useful ones. While BDT provides a possible solution for modelling utility influences upon perception, PP will require a different approach. This is because unlike BDT, which proposes distinct representations of probability and utility (each of which must be inferred separately by the agent), PP is 'an account that posits no pair of fundamentally distinct inner cognitive kinds' ([80], p. 4). In PP, this may be achieved by a move to active inference, where 'reward or cost functions are absorbed into prior beliefs about state transitions and terminal states' ([81], p. 523). Under this formulation, the brain is proposed to perform 'Bayesian decision theory in disguise' ([82], p. 7), as high-utility states are simply recast as high-probability states from the perspective of the generative model. The relative probability of an inference will thus depend upon the organism in question and their respective econiche. Being surrounded by water may be a high-probability state for a fish, but not for a camel. A high-probability state for mammals may involve not being poisoned. From this perspective, snakes may be inferred as probable if their perception is predicted to efficiently maintain an organism within its expected bounds. This is despite the fact that they may be, in purely objective terms, relatively rare.

Discrepancies between predictions and sensory input result in prediction errors, which are weighed by their expected **precision** (mathematically, their inverse variance), providing a measure of the reliability of a given signal. According to active inference, such prediction error may be minimised in two ways [9]: (i) perception, changing one's predictions to better fit the incoming sensory input; and (ii) action, changing the sensory input to better fit one's predictions. Perception and action are thus proposed to serve the same underlying imperative, albeit via distinct functional roles according to their respective direction of fit (mind to world versus world to mind [2]).

Embodied approaches emphasise that the ultimate imperative of the predictive machinery is to ensure survival through preserving physiological integrity (e.g., [11–14]). From this perspective, prediction error minimisation is fundamentally grounded in **allostasis** (cf. homeostasis), the process of regulating the internal physiological parameters of the body (e.g., temperature and blood pressure) within a viable range of values. Much like stimuli in the outside world, the values of these bodily variables are not directly available to the brain, and so their hidden causes must be inferred from the available interoceptive signals ('interoceptive inference' [15]). When such values deviate from their predicted set points, prediction error ensues, and action is engaged (via adaptive autonomic responses), minimising interoceptive prediction error through achieving allostatic equilibrium. Such physiological regulation is not simply reactive, since waiting for certain bodily variables to move outside an acceptable range before responding could prove fatal [16]. Thus, organisms must proactively preempt the trajectories of such key variables and their possible perturbations over time [17].

While some physiological variables are regulated through largely autonomic reflex processes (e.g., heart-rate/blood pressure/glucose levels), others (e.g., hunger/thirst/sex) can only be regulated through actively engaging with the environment. This forms the basis for the emergence of perceptual models [17], since successfully acting upon the outside world requires an agent to have an adequate model of its surroundings. Perceptual inference, from this perspective, can be cast as a form of affordance competition [18], where perceptual hypotheses represent opportunities for action (i.e., affordances [19]) in the environment. Once again, the proactive nature of this model should be emphasised; organisms will not only reactively infer worldly states in relation to current bodily concerns, but also proactively in relation to future ones. Such anticipatory control is crucial for organisms to survive over extended periods; they enable them to capitalise on opportunities for improving future allostatic conditions while avoiding potential threats to their physiological

from more specific proposals, such as predictive coding, which stipulate a mechanistic basis for how PP is implemented.

Surprisal: an information-theoretical measure denoting the relative implausibility of a sensory state occurring under a given model of the world (formally the negative log probability of an outcome). This should be distinguished from surprise in the more general use of the term, which refers to a feeling of unexpectedness experienced at the level of the agent. For example, the brain may infer that an object is a snake because it represents the least surprising perceptual prediction given a model of the world and the available sensory evidence. Nonetheless, this inference may still be experienced as highly surprising to the agent due to its relative unexpectedness.

Utility: a measure of worth or value. In perceptual decision-making terms, this reflects the expected pay-offs or consequences of a given perceptual outcome. In the present account, utility is closely aligned with allostasis, whereby benefits and costs reflect improving and deteriorating conditions of physiological regulation.

Utility bias: a tendency to perceive stimuli in a manner that maximises benefits and minimises costs. This is proposed to be driven by changes in prior predictive information (i.e., expectations and constraints) as opposed to attention. It leads to an increased occurrence of 'useful' false alarms. Utility biases are strongest under conditions of sensory uncertainty and where the potential benefits or costs are high.

integrity before these threats arise. In cognitively more complex organisms, this prospective ability extends to forms of cognition that operate over longer timescales so that 'the brain's ability to predict the consequences of actions enables it to link across levels of abstraction, and to bias immediate actions by the predicted long-term opportunities they make possible' ([20], p. 415). This leads to a conceptualisation of perception as a form of 'hierarchical affordance competition' [20], where representations of the environment compete for perceptual awareness across varying timescales, reflecting not only current but also future allostatic concerns. This enables organisms to predict future states or goals and select an appropriate sequence of actions ('policies') to attain them [21,22].

Perceptual inference is thus a pragmatic process; the brain's 'best guess' is not optimised to reflect the external world in some veridical sense, but rather is tailored to track the environment in terms of current and future allostatic action possibilities. While this view is not new, current accounts in the literature appear to be lacking in terms of their evaluation of how such an optimisation maps onto the distinct roles played by attention and expectations in perceptual decision-making. This is where we turn to next.

Attention and expectations: are they functionally orthogonal?

Perceptual decision-making is optimised by distinct underlying computational mechanisms. Signal detection theory [23] models and distinguishes these influences in terms of sensitivity, the discriminability of a stimulus (d'), and bias, the probability of a given response (C). While both may improve perceptual accuracy through improving the detection of a target stimulus, they are dissociable in terms of their effects upon detecting a target's absence, where increased sensitivity results in a higher proportion of correct rejections, while increased bias results in a higher proportion of false positives ('false alarms'). Using reaction times, drift diffusion modelling [24] may further decompose perceptual bias into underlying changes in evidence accumulation ('drift rate') or starting point ('response expectancy bias'). Changes in response expectancy are of particular relevance since they suggest that perception may be biased towards a particular response before the appearance of the stimulus itself [25].

Within PP, the optimisation of sensitivity and bias can be mapped onto the functional roles played by attention [26] and expectations [27,28], respectively. In this regard, attention acts to optimise expected precision and improves discrimination sensitivity through suppressing internal noise, while expectations flexibly influence discrimination bias through changing the baseline activity of signal-selective units [29]. These are also thought to lead to divergent changes in neural response [30]. Accordingly, expectations suppress neural activity due to either the prior sharpening of sensory representations [31] or a reduction in the redundancy of information [32]; meanwhile, attention increases neural activity [33] and may counteract the effect of expectations on neural suppression [34] (see also [35]).

Expectations are inherently context dependent and hierarchical ('top-down') in nature [36]. Their predictive influence on perception is thus flexible and may vary according to information that is independent of the input itself, such as the surrounding context or the expected reliability of the input. For example, expectations may facilitate the speed and accuracy of perceiving objects when they are placed within their typical contexts [37]. Similarly, expectations are most likely to modulate perceptual contents when sensory signals are expected to be noisy, imprecise, or ambiguous [30]. Importantly, perception may also be biased by context-independent mechanisms embedded in the bottom-up stream ('constraints' [36]). We explore the relationship between these distinct forms of predictive information in relation to the present framework in Box 2.

Box 2. Context dependency and mechanistic insight: constraints versus expectations

Recently, it has been proposed that the implementation of predictive information in the nervous system should structurally reflect the type of environmental regularities it captures [36]. This results in two main ways by which prior information can be mechanistically implemented: (i) expectations, hierarchical top-down predictive information that reflects context-dependent regularities (i.e., those that are relevant to specific encounters between an agent and her environment); and (ii) constraints, non-hierarchical predictive information embedded in the bottom-up stream that captures context-independent regularities (i.e., those that are relevant for every encounter between an agent and her environment). While this notion was originally framed in terms of capturing probability biases based upon environmental statistics, such reasoning may plausibly be applied in terms of an optimisation towards utility. Thus, we predict that context-sensitive affordances (i.e., those whose utility is specific to certain encounters) will be largely driven by hierarchical and flexible expectations, such as context-dependent modulations in utility biases that derive from external factors indexing risk [50–52] or state changes in the body [54–57]. By contrast, context-insensitive ones (i.e., those that bear high utility regardless of context) will be largely driven by automatic mechanisms suggestive of constraints. This may shed light on a number of candidate cases for constraints that were not included in the original proposal. For example, snakes are perceived in an automatic fashion according to their prototypical features, even in the absence of prior experience [83]. Similarly, the preferential perception of fearful faces has been shown to be driven automatically by low-level stimulus information, such as effective contrast and high spatial frequencies [84,85]. These cases are likely to be implemented neurally by inflexible mechanisms due to their state-independent utility. An important question for future research concerns whether constraint-based utility biases reflect prior predictive information that is specific to the utility bias in question or whether they derive from more general priors. For example, while some studies suggest top-down factors are involved in the facing-the-viewer bias [40], others suggest that it merely reflects the light-from-above prior [86], where surfaces are more likely to be perceived as convex [87].

Within the theoretical literature, attention and expectations are often assigned functionally orthogonal roles where 'the motivational considerations that guide attention are in principle orthogonal to the perceptual regularities that guide visual expectations' ([27], p. 404). Within this view, 'probability-driven' expectations are proposed to be utility neutral, insofar as they are optimised to recapitulate statistical regularities in the environment. Such a proposal is at least implicitly present in the literature where expectations are exclusively discussed in terms of perceptual biases that conform to probabilistic aspects of the environment [30,36]. Similarly, biases leading to false alarms are discussed in the context of 'optimal percepts', where 'given the structure and statistics of the world we inhabit, the optimal estimate of the worldly state…will be the one that, on some occasions, gets things wrong' ([1], p. 51). We interpret this to reflect a tendency in the literature to assume that expectations are ultimately geared towards 'getting the world right'. Attention then carries the functional burden of prioritising allostasis-relevant sensory inputs in relation to regulatory needs.

We argue against such functional orthogonality and defend the claim that expectations are utility driven, functionally biased towards inferring useful (i.e., allostasis-regulating) outcomes. Why should expectations be recruited in such a manner? There are limits to the extent that attentional processes can minimize internal and external sources of noise due to factors such as metabolic costs and inherent limitations of the sensory organs. There will thus always exist a level of irreducible uncertainty, providing a 'ceiling effect' for attention beyond which an adjustment of a decision criterion via biased expectations may further promote adaptive perceptual decision-making. Veridical representations are not always 'useful' ones [38,39,97], and thus perception may benefit from utility-driven biases as long as they lead to perceptual predictions that minimise prediction error in the long run. While such a **utility bias** would result in an increase in the net amount of perceptual errors (due to increased 'false alarms' for snakes and other useful affordances when they are not there), it will, more importantly, reduce the number of costly errors (type 2 errors associated with missed detections). A related line of argument was put forth by Seth and Tsakiris [17] in linking allostatic imperatives to useful but inaccurate percepts: 'we do not perceive the world (and self) as it is, but as it is useful to do so. This may involve systematically "misperceiving" the environment, by criteria of veridicality' (p. 9). Given that the probability-driven versus utility-driven models lead to divergent predictions, their relative validity can be scrutinised empirically.

Expectations are utility driven: support from the empirical literature

Defending the claim that expectations are utility driven is complicated by the fact that accuracy and utility often coincide in perceptual inference. However, the probability-driven versus utility-driven models should lead to divergent predictions under conditions where accuracy and utility conflict with one another [8]. Empirically, this has been achieved by evaluating the effects of expectations in relation to ambiguous stimuli that give rise to percepts of comparable probability yet of varying utility. For example, ambiguous point-light walkers are more likely to be perceived as facing the viewer than facing away, which is thought to reflect the utility of approaching agents [40]. Similarly, in binocular rivalry, people are more likely to perceive stimuli conditioned by threats or rewards [41–44]. Such biases may also be evident in perceptual estimates of a continuous nature, where useful estimates may be biased away from accurate ones. For example, 'looming biases' show that the distance and arrival time of approaching stimuli are underestimated relative to those of receding ones, which is thought to provide a temporal margin of safety against approaching threats [45]. Similarly, heights that present a falling risk are consistently overestimated when viewed from above, which is assumed to prevent life-threatening navigational decisions [46]. Further direct evidence for utility biases can be found in signal detection paradigms. For example, differences in pay-off structures lead to increased response biases in the perception of temporal intervals, with no related changes in discrimination sensitivity [47]. Interestingly, such biases are stronger in relation to imbalanced punishments as opposed to imbalanced rewards [47], suggesting that utility biases may be stronger in relation to negative outcomes. Finally, anxious individuals show greater response bias towards threat-related stimuli, such as spiders [48], fearful faces [49], and threatening words [25]. Despite this, in terms of sensitivity, they are not any better [25,48] and sometimes even worse than controls [49]. Drift diffusion analysis further indicates that this threat bias can be decomposed into changes reflecting both stimulus and response expectancy biases, suggesting that highly anxious individuals are biased to perceive threats even before the onset of the stimulus [25].

Different optimisations in expectations also lead to divergent predictions in context-dependent modulations. Specifically, utility-driven expectations should be modulated by factors informative of risk independent of their probabilistic basis. This is supported by studies showing that the strength of utility biases varies in relation to the severity of expected consequences. For example, arrival times in looming biases are underestimated more for threatening objects (snakes and spiders) than for neutral ones (butterflies and rabbits) [50]. Similarly, point-light walkers are more likely to be perceived as facing the viewer at narrower camera angles, which is thought to reflect an increased chance of the agent interacting with the viewer [51]. Finally, overestimations in height are extinguished when realistic falling costs are removed [52]. This may also explain modulations deriving from interoceptive influences (see [53] for a review) and, specifically, their modulating influence on utility biases. For example, looming biases are modulated by the response capacity of the motor system, where arrival times are further underestimated in weaker and more unfit individuals [54] (see [55–57] for related examples). Furthermore, signal detection analyses demonstrate that interoceptive manipulations lead to increased bias towards auditory [58] and social threats [59]. Such influences may also apply to the influence of motivational states, such as hunger, thirst, and drug craving, where perceptual representations that lead to satiety (food, water, and drugs) are strengthened [60,61]. This is despite the fact that, probabilistically speaking, agents are more likely to be surrounded by food and drink during periods of satiation as opposed to periods of hunger and thirst. From this perspective, interoceptive changes may function to increase the precision of prior beliefs (via expectations), thus biasing perception towards inferences that carry allostasis-regulating outcomes.

A related claim concerns how perceptual estimates are modulated in relation to sensory precision. Utility-driven expectations should increasingly bias perception away from accurate estimates and

towards useful ones under conditions of mounting sensory uncertainty. This is because expectations are more likely to influence perceptual outcomes when sensory precision is low [30]. This modulation may efficiently minimise costly errors, since perception is biased towards an increasingly inaccurate but nonetheless 'safer' percept to account for the increased chance of a perceptual error ('the less certainty, the more caution'). Initial support for this idea comes from studies comparing looming biases in the visual and auditory modalities, where biases are stronger for hearing than they are for vision [62]. This may reflect differences in spatial sensory precision, where hearing is expected to be less reliable than vision. This may also explain why improbable (yet frightening) agents, such as the bogeyman, tend to be perceived at night (cf. [63]) when sensory precision is expected to be low.

Finally, this perspective may have implications for our understanding of learning. Specifically, if expectations are utility driven, then the updating of priors should also reflect such an optimisation, as predictions are updated in relation to the severity of their outcomes. This may explain the role of dopamine in learning, where better than expected outcomes ('positive reward prediction errors') lead to a surge in dopamine, thus reinforcing effective regulatory behaviour [14]. Similarly, it may explain biases deriving from 'one-shot learning', where an isolated but potentially fatal event (e.g., food poisoning) leads to a 'sudden surge in highly precise interoceptive prediction errors...and affords for an equally precise alteration of internal models' ([12], see p. 33). This reweighting of priors leads to lasting perceptual and behavioural changes (e.g., distaste for and avoidance of similar foods), effectively protecting against costly errors in the future. In extreme cases, however, this can lead to pathological internal models of the world, such as those seen in post-traumatic stress disorder, where traumatic experiences lead to overly precise priors that drive pathological inferences such as hallucinations [64–66]. We consider this and other forms of useful misrepresentation with respect to psychopathology in Box 3.

At what cost? Utility biases efficiently minimize interoceptive prediction error

On the surface, utility biases may seem to contradict the principle of prediction error minimisation, since they systematically lead organisms towards predictions that, on average, deviate from actual states in the world. However, they may be suitably accommodated once we revisit the notion that prediction error can arise in both the external and internal milieu. With this in mind, we view these respective models to be minimizing two conflicting sources of prediction error.

Box 3. Utility biases gone too far: hallucinations may reflect 'useful' false alarms

The framework discussed in this paper has implications for how certain aspects of psychopathology may be understood. Specifically, we propose that certain cases of disordered inference may reflect an extended form of 'useful misrepresentation', where dysfunctional precision weighting leads the brain to be overly tuned towards inferring allostasis-relevant worldly states under conditions of little or no supporting sensory evidence. Within this context, hallucinations (perceptual inferences that occur in the absence of external stimuli) can be conceptualised as false alarms generated by overly precise perceptual priors [3,88]. This view is supported by studies that demonstrate that hallucination-prone individuals exhibit a stronger bias in relation to social and threat-related stimuli [58,89] and are more susceptible to experimentally induced hallucinations [90,91]. Moreover, patients with auditory hallucinations show both higher sensitivity and false-positive bias in perceiving voices embedded in noise [92]. This perspective may also help elucidate links between traumatic experiences and the formation of hallucinations with direct or indirect thematic links to trauma-related content [64–66] as well as hallucinatory content that can be directly linked back to themes of evolutionary concern, such as tactile and visual hallucinations involving bugs under the skin [93,94]. Although acquired over different timescales, these phenomena reflect 'learned' perceptual priors that, although not being typical events that an individual commonly encounters, are nonetheless strongly weighted due to their predicted relevance for maintaining physiological integrity. We speculate that the same principle may extend to the notion of abnormal source monitoring in patients who experience hallucinations [95]. This is because perceptual inference involves inferring not only the hidden cause of a stimulus in relation to its identity but also its respective source (i.e., whether it is internally or externally generated), a decision that also presents asymmetric costs to the agent. 'Voice-hearing' in schizophrenia may thus originate from incorrectly perceiving internally generated speech as being externally caused, as has been suggested previously [96]. While the cases discussed earlier are illustrative of perceptual disturbance, these ideas are also likely to apply to the other forms of disordered inference, such as delusions and paranoid beliefs.

On the one hand, expectations optimised towards probability ('sticks over snakes') minimise uncertainty in relation to exteroceptive sensory states, since they, on average, more accurately predict states of the world and their unfolding sensory trajectories. On the other hand, expectations that are biased towards utility ('snakes over sticks') minimise uncertainty in relation to interoceptive sensory states, since they more accurately predict internal states through proactively avoiding 'high surprisal' states of the body, such as being poisoned.

But will not such utility biases lead to other problems? Utility-driven expectations may help agents avoid getting poisoned, but this may inadvertently lead to new difficulties as the small costs of misperceptions add up in the long term. We believe this worry can be addressed by considering the specific conditions under which utility biases manifest. As mentioned earlier, utility biases are modulated in relation to sensory uncertainty and the severity of consequences. Against this background, the costs of misperceptions are diminished as such biases are tempered when sensory stimuli are unambiguous, risks associated with misperception are low, or, alternatively, when expectations are highly inaccurate [30]. Potential costs may be further minimised through the disambiguating role of action. According to active inference, action may be not only motivational but also epistemic in nature [67]. For example, upon inferring the snake, your brain may both prepare the body for evasive action while simultaneously sampling this object in such a manner so as to reduce the uncertainty in its prediction. Should mounting prediction error occur (as the world does not unfold according to one's prediction), efficient prediction error minimisation may then involve reverting back to perceptual inference as the current best hypothesis is now revised to accommodate the new sensory information ('it's a stick'). Collectively, these factors may help minimise costs associated with actions and policies deriving from misguided inferences.

Concluding remarks

The core function of cognition is to successfully regulate the internal milieu [14], and perceptual models are thought to emerge as a consequence of this imperative [68]. We suggest that this optimisation is mediated by biases in the structure and tuning of predictive information in the brain, leading to internal models that usefully misrepresent the environment. Allostasis thus plays an essential role in constraining perceptual outcomes, consistent with the notion that the 'brain recovers the distal environment through its generative model only insofar as it bears on its regulatory function' ([69], p. 157). In this sense, we do not experience the world usefully in so far as it is accurate; rather, we experience the world accurately in so far as it is useful [17].

While the focus of this paper has been on perception, the same principles may reasonably be applied to other forms of cognitive inference. Indeed, the utility bias literature includes findings ranging from evolutionarily ancient functions, such as immune responses, to more recently evolved cognitive functions, such as false and superstitious beliefs [70]. Beliefs evolve rather naturally out of this account, since they are thought to result from the same inferential processes as percepts and differ mainly with respect to timescale [2]. Utility biases may also be present in pseudo-perceptual states, such as imagination [71] and dreaming [72], which commonly involve improbable yet biologically relevant themes. Furthermore, action is thought to be enabled by a systematic underrepresentation of peripheral sensory precision estimates [38]. Useful misrepresentation may thus not be limited to inferential processes in the perceptual domain; it is more likely to reflect a ubiquitous principle of prediction at multiple levels of the brain and the body.

The present perspective may inform controversies surrounding the explanatory scope of PP in cases where utility and accuracy conflict with one another, such as reward-biased perception [73], affect-biased attention [74], and more general issues linked to the **Dark Room Problem** [75] (cf. [39,76,77]). Future open questions include whether utility-biased expectations differ

Outstanding questions

Are there differences between utility-driven expectations and those that coincide with perceptual accuracy? Specifically, do these types of expectations lead to dissociable changes in neural response? Do utility biases in expectation sharpen sensory representations similar to those driven by expectations based on environmental regularities?

If both attention and expectations are utility driven, what are the differences in their neural implementation? How do the neural effects of utility-driven expectations interact with those of attention?

In the framework of signal detection theory, what are the contributions of bias and sensitivity in the perceptual prioritisation of allostasis-relevant events?

How are utility-driven expectations reflected in standard models of perceptual decision-making, such as drift-diffusion models? Will such biases prove to be explainable in terms of evidence accumulation or response expectancy or both?

Do effects of utility in perception depend on sensory and/or environmental uncertainty? Will utility-driven expectations bias perception away from accurate estimates and towards useful ones under conditions of uncertainty?

Are perceptual biases, which usefully misrepresent the world, associated with changes in affect? If so, do these changes occur before or subsequent to the formation of the percept? Do utility biases deriving from negative affect differ to those from positive affect?

Can the framework of utility biases inform our understanding of perceptual disturbances in affective disorders such as major depression and bipolar disorder?

Can the framework of utility biases help us understand the phenomenology of perceptual states that arise in the absence of sensory stimuli, such as imagination, dreaming, and hallucinations?

from those that coincide with perceptual accuracy as well as the contributing role of affect, which has recently been postulated to underwrite perceptual consciousness [78] (see Outstanding questions).

How do utility biases affect decision-making in domains other than perception? Are the effects of utility-driven expectations in cognitive inference governed by similar principles to those that govern perception?

Acknowledgements

We acknowledge the Deutscher Akademischer Austauschdienst and the Berlin School of Mind and Brain for scholarship funding to J.M.M. and the Berlin Institute of Health for a Clinical Fellowship to P.S. in relation to this research. P.S. is further supported by grants from Deutsche Forschungsgemeinschaft (STE 1430/8-1 and STE 1430/9-1). Special thanks to Chiara Caporuscio, Marika Constant, and the members from the Visual Perception Laboratory for the fruitful discussions and motivational support.

Declaration of interests

The authors declare no competing interests in relation to this work.

References

1. Clark, A. (2015) *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*, Oxford University Press
2. Hohwy, J. (2013) *The Predictive Mind*, Oxford University Press
3. Sterzer, P. *et al.* (2018) The predictive coding account of psychosis. *Biol. Psychiatry* 84, 634–643
4. Hohwy, J. *et al.* (2008) Predictive coding explains binocular rivalry: an epistemological review. *Cognition* 108, 687–701
5. Johnson, D.D. *et al.* (2013) The evolution of error: error management, cognitive constraints, and adaptive decision-making biases. *Trends Ecol. Evol.* 28, 474–481
6. Haselton, M.G. *et al.* (2015) The evolution of cognitive bias. In *The Handbook of Evolutionary Psychology* (Buss, D.M., ed.), pp. 968–987, John Wiley & Sons
7. Haselton, M.G. and Nettle, D. (2006) The paranoid optimist: an integrative evolutionary model of cognitive biases. *Personal. Soc. Psychol. Rev.* 10, 47–66
8. Lynn, S.K. and Barrett, L.F. (2014) "Utilizing" signal detection theory. *Psychol. Sci.* 25, 1663–1673
9. Parr, T. and Friston, K.J. (2019) Generalised free energy and active inference. *Biol. Cybern.* 113, 495–513
10. Barrett, L.F. (2017) The theory of constructed emotion: an active inference account of interoception and categorization. *Soc. Cogn. Affect. Neurosci.* 12, 1–23
11. Allen, M. and Friston, K.J. (2018) From cognitivism to autopoiesis: towards a computational framework for the embodied mind. *Synthese* 195, 2459–2482
12. Allen, M. and Tsakiris, M. (2018) The body as first prior: interoceptive predictive processing and the primacy. In *The Interoceptive Mind: From Homeostasis to Awareness* (Tsakiris, M. and De Preester, H., eds), pp. 27–45, Oxford University Press
13. Kirchhoff, M. (2018) The body in action: predictive processing and the embodiment thesis. In *Oxford Handbook of Cognition: Embodied, Extended and Enactive* (Newen, A. *et al.*, eds), pp. 243–260, Oxford University Press
14. Schulkin, J. and Sterling, P. (2019) Allostasis: a brain-centred, predictive mode of physiological regulation. *Trends Neurosci.* 42, 740–752
15. Seth, A.K. (2013) Interoceptive inference, emotion, and the embodied self. *Trends Cogn. Sci.* 17, 565–573
16. Sterling, P. (2014) Homeostasis vs allostasis: implications for brain function and mental disorders. *JAMA Psychiatry* 71, 1192–1193
17. Seth, A.K. and Tsakiris, M. (2018) Being a beast machine: the somatic basis of selfhood. *Trends Cogn. Sci.* 22, 969–981
18. Cisek, P. (2007) Cortical mechanisms of action selection: the affordance competition hypothesis. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 1585–1599
19. Gibson, J.J. (1979) *The Ecological Approach to Visual Perception*, Houghton Mifflin
20. Pezzulo, G. and Cisek, P. (2016) Navigating the affordance landscape: feedback control as a process model of behavior and cognition. *Trends Cogn. Sci.* 20, 414–424
21. Pezzulo, G. *et al.* (2018) Hierarchical active inference: a theory of motivated control. *Trends Cogn. Sci.* 22, 294–306
22. Corcoran, A.W. *et al.* (2020) From allostatic agents to counterfactual cognisers: active inference, biological regulation, and the origins of cognition. *Biol. Philos.* 35, 32
23. Green, D.M. and Swets, J.A. (1966) *Signal Detection Theory and Psychophysics*, Vol. 1. Wiley
24. Ratcliff, R. and McKoon, G. (2008) The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922
25. White, C.N. *et al.* (2016) Using decision models to decompose anxiety-related bias in threat classification. *Emotion* 16, 196
26. Feldman, H. and Friston, K. (2010) Attention, uncertainty, and free-energy. *Front. Hum. Neurosci.* 4, 215
27. Summerfield, C. and Egner, T. (2009) Expectation (and attention) in visual cognition. *Trends Cogn. Sci.* 13, 403–409
28. Summerfield, C. and De Lange, F.P. (2014) Expectation in perceptual decision making: neural and computational mechanisms. *Nat. Rev. Neurosci.* 15, 745–756
29. Wyart, V. *et al.* (2012) Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proc. Natl. Acad. Sci. U. S. A.* 109, 3593–3598
30. De Lange, F.P. *et al.* (2018) How do expectations shape perception? *Trends Cogn. Sci.* 22, 764–779
31. Kok, P. *et al.* (2012) Less is more: expectation sharpens representations in the primary visual cortex. *Neuron* 75, 265–270
32. Kumar, S. *et al.* (2017) Encoding of predictable and unpredictable stimuli by inferior temporal cortical neurons. *J. Cogn. Neurosci.* 29, 1445–1454
33. Hesselmann, G. *et al.* (2010) Predictive coding or evidence accumulation? False inference and neuronal fluctuations. *PLoS One* 5, e9926
34. Garrido, M.I. *et al.* (2018) Bayesian mapping reveals that attention boosts neural responses to predicted and unpredicted stimuli. *Cereb. Cortex* 28, 1771–1782
35. Alink, A. and Blank, H. (2021) Can expectation suppression be explained by reduced attention to predictable stimuli? *NeuroImage* 231, 117824
36. Teufel, C. and Fletcher, P.C. (2020) Forms of prediction in the nervous system. *Nat. Rev. Neurosci.* 21, 231–242
37. Bar, M. (2021) From objects to unified minds. *Curr. Dir. Psychol. Sci.* 30, 129–137
38. Wiese, W. (2017) Action is enabled by systematic misrepresentations. *Erkenntnis* 82, 1233–1252
39. Baltieri, M. and Buckley, C.L. (2019) The dark room problem in predictive processing and active inference, a legacy of cognitivism? In *Artificial Life Conference Proceedings* (Fellerman, H. *et al.*, eds), pp. 40–47, MIT Press
40. Vanrie, J. *et al.* (2004) Bistability and biasing effects in the perception of ambiguous point-light walkers. *Perception* 33, 547–560
41. Alpers, G.W. *et al.* (2005) Binocular rivalry between emotional and neutral stimuli: a validation using fear conditioning and EEG. *Intern. J. Psychophysiol.* 57, 25–32
42. Marx, S. and Einhäuser, W. (2015) Reward modulates perception in binocular rivalry. *J. Vis.* 15, 11
43. Wilbertz, G. *et al.* (2014) Reinforcement of perceptual inference: reward and punishment alter conscious visual perception during binocular rivalry. *Front. Psychol.* 5, 1377
44. Wilbertz, G. *et al.* (2017) fMRI-based decoding of reward effects in binocular rivalry. *Neurosci. Conscious.* 3, nix013

45. Neuhoff, J.G. (2018) Adaptive biases in visual and auditory looming perception. In *Spatial Biases in Perception and Cognition* (Hubbard, T.L., ed.), pp. 366–472, Cambridge University Press
46. Jackson, R.E. and Cormack, L.K. (2007) Evolved navigation theory and the descent illusion. *Percept. Psychophys.* 69, 353–362
47. Akdoğan, B. and Balcı, F. (2016) The effects of payoff manipulations on temporal bisection performance. *Acta Psychol.* 170, 74–83
48. Becker, E. and Rinck, M. (2004) Sensitivity and response bias in fear of spiders. *Cogn. Emot.* 18, 961–976
49. Frenkel, T.I. *et al.* (2009) Individual differences in perceptual sensitivity and response bias in anxiety: evidence from emotional faces. *Cogn. Emot.* 23, 688–700
50. Vagnoni, E. *et al.* (2012) Threat modulates perception of looming visual stimuli. *Curr. Biol.* 22, R826–R827
51. Zhang, X. *et al.* (2017) The interaction of perceptual biases in bistable perception. *Sci. Rep.* 7, 42018
52. Jackson, R.E. and Cormack, L.K. (2010) Reducing the presence of navigation risk eliminates strong environmental illusions. *J. Vis.* 10, 9
53. Zadra, J.R. and Clore, G.L. (2011) Emotion and perception: the role of affective information. *Wiley Interdiscip. Rev. Cogn. Sci.* 2, 676–685
54. Neuhoff, J.G. *et al.* (2012) Strength and physical fitness predict the perception of looming sounds. *Evol. Hum. Behav.* 33, 318–322
55. Heenan, A. and Troje, N.F. (2014) Both physical exercise and progressive muscle relaxation reduce the facing-the-viewer bias in biological motion perception. *PLoS One* 9, e99902
56. Shen, M. *et al.* (2018) Guilt leads to enhanced facing-the-viewer bias. *PLoS One* 13, e0195590
57. Stefanucci, J.K. and Storbeck, J. (2009) Don't look down: emotional arousal elevates height perception. *J. Exp. Psych. Gen.* 138, 131
58. Dudley, R. *et al.* (2014) The effect of arousal on auditory threat detection and the relationship to auditory hallucinations. *J Behav. Ther. Exp. Psychiat.* 45, 311–318
59. Lynn, S.K. *et al.* (2012) Affective state influences perception by affecting decision parameters underlying bias and sensitivity. *Emotion* 12, 726
60. Weng, X. *et al.* (2019) Effects of hunger on visual perception in binocular rivalry. *Front. Psych.* 10, 418
61. Changizi, M.A. and Hall, W.G. (2001) Thirst modulates a perception. *Perception* 30, 1489–1497
62. DeLucia, P.R. (2015) Perception of collision. In *Cambridge Handbooks in Psychology. The Cambridge Handbook of Applied Perception Research* (Hoffman, R.R. *et al.*, eds), pp. 568–591, Cambridge University Press
63. Pezzulo, G. (2014) Why do you fear the bogeyman? An embodied predictive coding model of perceptual inference. *Cogn. Affect. Behav. Neurosci.* 14, 902–911
64. Kube, T. *et al.* (2020) Rethinking post-traumatic stress disorder—a predictive processing perspective. *Neurosci. Biobehav. Rev.* 113, 448–460
65. Linson, A. *et al.* (2020) Active inference, stressors, and psychological trauma: a neuroethological model of (mal) adaptive explore-exploit dynamics in ecological context. *Behav. Brain Res.* 380, 112421
66. Lyndon, S. and Corlett, P.R. (2020) Hallucinations in posttraumatic stress disorder: insights from predictive coding. *J. Abnorm. Psychol.* 129, 534
67. Friston, K. *et al.* (2015) Active inference and epistemic value. *Cogn. Neurosci.* 6, 187–214
68. Seth, A.K. (2015) The cybernetic Bayesian brain—from interoceptive inference to sensorimotor contingencies. In *Open MIND: 35(T)* (Metzinger, T. and Windt, J.M., eds), pp. 1–24, Frankfurt am Main
69. Williams, D. (2018) Predictive processing and the representation wars. *Minds Mach.* 28, 141–172
70. McKay, R. and Efferson, C. (2010) The subtleties of error management. *Evol. Hum. Behav.* 31, 309–319
71. Hesp, C. *et al.* (2020) Sophisticated affective inference: simulating anticipatory affective dynamics of imagining future events. In *International Workshop on Active Inference* (Verbelen, T. *et al.*, eds), pp. 179–186, Springer
72. Hobson, J.A. *et al.* (2014) Virtual reality and consciousness inference in dreaming. *Front. Psychol.* 5, 1133
73. Haas, J. (2021) Can hierarchical predictive coding explain binocular rivalry? *Philos. Psychol.* 34, 424–444
74. Ransom, M. *et al.* (2020) Affect-biased attention and predictive processing. *Cognition* 203, 104370
75. Sun, Z. and Firestone, C. (2020) The dark room problem. *Trends Cogn. Sci.* 24, 346–348
76. Seth, A.K. *et al.* (2020) Curious inferences: reply to Sun and Firestone on the Dark Room Problem. *Trends Cogn. Sci.* 24, 681–683
77. Van de Cruys, S. *et al.* (2020) Controlled optimism: reply to Sun and Firestone on the Dark Room Problem. *Trends Cogn. Sci.* 24, 680–681
78. Solms, M. (2021) *The Hidden Spring: A Journey to the Source of Consciousness*, WW Norton & Company
79. Berger, J. (2013) *Statistical Decision Theory: Foundations, Concepts, and Methods*, Springer Science & Business Media
80. Clark, A. (2020) Beyond desire? Agency, choice, and the predictive mind. *Australas. J. Philos.* 98, 1–15
81. Friston, K. *et al.* (2012) Active inference and agency: optimal control without cost functions. *Biol. Cyber.* 106, 523–541
82. Gershman, S.J. (2019) What does the free energy principle tell us about the brain? *arXiv* Published online January 23, 2019. https://arxiv.org/abs/1901.07945
83. Bertels, J. *et al.* (2020) Snakes elicit specific neural responses in the human infant brain. *Sci. Rep.* 10, 7443
84. Stein, T. *et al.* (2014) Rapid fear detection relies on high spatial frequencies. *Psychol. Sci.* 25, 566–574
85. Hedger, N. *et al.* (2015) Fearful faces have a sensory advantage in the competition for awareness. *J. Exp. Psychol. Hum. Percept. Perform.* 41, 1748–1757
86. Sun, J. and Perona, P. (1998) Where is the sun? *Nat. Neurosci.* 1, 183–184
87. Weech, S. *et al.* (2014) What causes the facing-the-viewer bias in biological motion? *J. Vis.* 14, 10
88. Corlett, P.R. *et al.* (2019) Hallucinations and strong priors. *Trends Cogn. Sci.* 23, 114–127
89. Stuke, H. *et al.* (2021) Overly strong priors for socially meaningful visual signals are linked to psychosis proneness in healthy individuals. *Front. Psychol.* 12, 1083
90. Powers, A.R. *et al.* (2017) Pavlovian conditioning-induced hallucinations result from overweighting of perceptual priors. *Science* 357, 596–600
91. Schmack, K. *et al.* (2021) Striatal dopamine mediates hallucination-like perception in mice. *Science* 372, eabf4740
92. Vercammen, A. *et al.* (2008) Hearing a voice in the noise: auditory hallucinations and speech perception. *Psychol. Med.* 38, 1177–1184
93. Fénelon, G. *et al.* (2002) Tactile hallucinations in Parkinson's disease. *J. Neurol.* 249, 1699–1703
94. Hinkle, N.C. (2011) Ekbom Syndrome: a delusional condition of "Bugs in the skin". *Curr. Psych. Rep.* 13, 178–186
95. Brunelin, J. *et al.* (2006) Source monitoring deficits in hallucinating compared to non-hallucinating patients with schizophrenia. *Europ. Psych.* 21, 259–261
96. Bentall, R.P. (1990) The illusion of reality: a review and integration of psychological research on hallucinations. *Psychol. Bull.* 107, 82–95
97. Tschantz, A. *et al.* (2020) Learning action-oriented models through active inference. *PLoS Comput. Biol.* 16, e1007805