Received: 1 April 2024 | Accepted: 20 December 2024

DOI: 10.1111/1365-2435.14750

Functional Ecology

REVIEW

Cognitive ecology of surprise in predator–prey interactions

Olivier Penacchio[1,2] | Liisa Hämäläinen[3,4] | Bibiana Rojas[4,5] | Kyle Summers[6] | Justin Yeager[7] | Thomas N. Sherratt[8] | Alice Exnerová[9]

[1]Computer Science Department, Computer Vision Center, Universitat Autònoma de Barcelona, Barcelona, Spain

[2]School of Psychology and Neuroscience, University of St Andrews, St Andrews, Fife, UK

[3]School of Science, Western Sydney University, Sydney, New South Wales, Australia

[4]Department of Biological and Environmental Science, University of Jyväskylä, Jyväskylä, Finland

[5]Department of Interdisciplinary Life Sciences, Konrad Lorenz Institute of Ethology, University of Veterinary Medicine, Vienna, Austria

[6]Department of Biology, East Carolina University, Greenville, North Carolina, USA

[7]Grupo de Biodiversidad Medio Ambiente y Salud, Facultad de Ingenierías y Ciencias Aplicadas, Universidad de Las Américas, Quito, Ecuador

[8]Department of Biology, Carleton University, Ottawa, Ontario, Canada

[9]Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic

Correspondence
Olivier Penacchio
Email: openacchio@gmail.com

Alice Exnerová
Email: alice.exnerova@natur.cuni.cz

Funding information
Natural Sciences and Engineering Research Council of Canada; Universidad de Las Américas Ecuador, Grant/Award Number: 483.A.XIV.24; University of Veterinary Medicine Vienna; Maria Zambrano Fellowship—NextGeneration EU (ALRC); Research Council of Finland, Grant/Award Number: 355869

Handling Editor: Raul Costa-Pereira

Abstract

1. In this review, we relate theoretical work on the importance of surprise in cognition to empirical research relevant to surprise in predator–prey interactions.
2. There have been multiple proposals as to how surprise should be defined and quantified in the context of animal cognition, including contributions from associative learning, information theory, Bayesian inference and the recent framework of active inference.
3. We argue that active inference provides a novel and powerful approach to quantifying surprise and advances the field by revealing how proactive behaviour on the part of predators relates to reducing surprise.
4. The active inference framework encompasses both proximate (e.g. neurobiological) and ultimate (evolutionary) aspects of surprise and brings new insights into key aspects of prey defences that exploit predator surprise.
5. We focus on surprise in defences that involve a sudden change in prey appearance (such as deimatic displays), and in defences that increase prey unpredictability (such as variation in chemical defences). We review literature that have investigated these phenomena and connect them to active inference. We also consider how multiple prey defences impact surprise in predators.
6. Finally, we consider the implications of active inference for future studies of predator–prey interactions, illustrate how this approach can be used to quantify surprise in prey defences and predict predator behaviour, and outline key questions that can be addressed within this framework.

Thomas N. Sherratt and Alice Exnerová are joint senior authors.

KEYWORDS
active inference, antipredator defence, associative learning, Bayesian inference, information theory, predation sequence, predator cognition, surprise

1 | INTRODUCTION

The evolution of camouflage, aposematism, mimicry and other forms of antipredator defence is a core research interest of evolutionary biologists (Ruxton et al., 2018). As argued in a recent review focused on interactions between prey defences and predator cognitive mechanisms, understanding the selective forces driving the evolution of prey defensive strategies requires an integrative approach that considers their coevolution with predator perception, cognition and behaviour (Kikuchi et al., 2023).

Many antipredator defences exploit the effects on predator cognition generated by surprise, which can influence the predator's behaviour. Here, we define surprise as a mismatch between an animal's expectation and what it observes or experiences (Barto et al., 2013). From the perspective of antipredator defence, surprise therefore refers to the extent to which the current observation or outcome of the encounter with a prey differs from predator's expectations. These expectations are based on previous experiences, both throughout the lifespan of the individual and those shaped over evolutionary time-scales (McNamara et al., 2006). Surprise has been at the centre of the theory of learning for several decades, from the Rescorla–Wagner model of associative learning (Rescorla & Wagner, 1972) and subsequent related concepts (Pearce & Hall, 1980), to theoretical frameworks based on Bayesian inference (Courville et al., 2006; Itti & Baldi, 2009). Recently, the concept of surprise has been extended to the active inference framework, which considers action alongside perception to explain how organisms interact with their environment (Clark, 2013; Friston, 2010; Parr et al., 2022).

As defined above, surprise relates to the unexpectedness of an outcome or observation. Almost by definition, unexpected prey defences may result in rapid change in a predator's beliefs. Surprise may therefore be a common, yet overlooked, aspect of antipredator defences. Surprise can be brought about by a sudden change in prey appearance, as in deimatic displays (Drinkwater et al., 2022), or it can be generated by variation in prey profitability (toxicity or other defence) even though the prey share the same appearance (Balogh et al., 2008). Despite the importance of surprise in theoretical concepts of cognition, its potential role in prey defences, and subsequent effects on predator cognition and behaviour, have never been systematically addressed. We propose that incorporating surprise-related theory into predator–prey interactions offers insight into how prey defences influence predator cognition and drive decisions and behaviours. This allows for the formulation of novel predictions and hypotheses.

Here, we compare approaches to the formal quantification of surprise used in cognitive sciences, evaluate the evidence for the potential role of surprise across several types of antipredator defence and elucidate the effects that surprising defences might have on predator cognition and behaviour. We propose that the principle of Bayesian learning and the theory of active inference may contribute to our understanding of the surprise-related aspects of predator–prey interactions.

2 | SURPRISE IN COGNITION

Animals use their senses adaptively, to reduce *uncertainty* (italicised names are defined in the glossary, Appendix S4) about their environment. Uncertainty corresponds to incomplete knowledge of the true value of a relevant aspect of the environment, for example of the probability that a potential prey item is profitable to eat or not (e.g. because it contains toxins). *Surprise* relates to uncertainty: there is no surprise in a world in which all is known with certainty. The concept of surprise has a long intellectual history that can be traced back to Aristotle, Hume and Darwin (see Reisenzein et al., 2019). Darwin was a devoted parent to his many children, but also observed them closely from a scientific perspective. He was particularly interested in their emotional expressions, how they compared to those of other animals, and whether they were innate or learned. At times he would try to surprise his children (by making sudden, loud noises for example), to observe their subsequent reactions and expressions. His observations contributed to his third major book (Darwin, 1872).

From studies on human cognition (Reisenzein et al., 2019), through visual science (Itti & Baldi, 2009), and studies on learning and animal cognition (Rescorla & Wagner, 1972), surprise has been consistently defined as a departure from expectation. Surprise is therefore underpinned by a mechanism for comparing expectation, or prediction, with the organism's current *observation* or experience (Barto et al., 2013). Here, we consider solely the information aspects of surprise, without any reference to emotional states (Ekman & Davidson, 1994). Despite qualitative agreement on the basic definition, surprise has not always been defined formally in a quantifiable way. Here we focus on three examples of how surprise has been formally quantified.

2.1 | Surprise in associative learning

In their seminal work, Rescorla and Wagner (1972) considered a classical conditioning paradigm. They proposed that the greater the uncertainty in an association between conditioned stimulus (CS; a biologically neutral stimulus, e.g. sound) and unconditioned stimulus (US; a biologically significant stimulus, e.g. food) the more can be learned from an observation/experience. Rescorla and Wagner (1972) chose to model the change in associative

strength V (i.e. the degree to which the animal expects the US in the presence of CS) between an unconditioned stimulus and a compound-conditioned stimulus AX after the exposition of an unconditioned stimulus as:

$$\Delta V_A = \alpha_A \beta (\lambda - V_{AX}) \quad (1.1)$$

and

$$\Delta V_X = \alpha_X \beta (\lambda - V_{AX}), \quad (1.2)$$

where $V_{AX} = V_A + V_X$, λ is the asymptotic level of associative strength the US can support, and α_A, α_X and β are coefficients between 0 and 1 that, respectively, represent the 'salience' of the CS and the 'associative value' of the US. The changes in associative strength between CS and US are therefore proportional to the 'surprisingness' of this association, defined as the difference between the asymptotic level of associative strength λ and the current associative strength V_{AX} (Pearce & Hall, 1980; Rescorla & Wagner, 1972). This formalisation of associative learning has been used to better understand the role of predator learning in mimicry (Balogh et al., 2008; Speed & Turner, 1999), where prey appearance is treated as the CS, prey unpalatability as the US and the strength of attack inhibition towards the prey appearance corresponds to the associative strength (V).

2.2 | Surprise in information theory

A natural way to express expectation for a set of alternative events is to associate a probability distribution (see Appendix S1) to this set of outcomes. Surprise, called 'surprisal' (or 'self-information'), is a synonym of information in the sense of Shannon's information theory (Shannon, 1948). It is expressed as the log of the inverse probability of an event y, which can, for example, correspond to an observation:

$$h(y) = \log\left(\frac{1}{p(y)}\right). \quad (2)$$

Accordingly, the occurrence of an event with low probability is more surprising (i.e. more informative) than the occurrence of an event with higher probability. For example, if a predator encountering a prey expects the prey to display a conspicuous red colour with probability 0.09, its surprisal when seeing the red colour will be $\log\left(\frac{1}{0.09}\right) = 2.42$ nats, where a nat is the unit of information based on natural logarithms (or, equivalently, measured in bits when considering the logarithm in base 2, $\log_2\left(\frac{1}{0.09}\right) = 3.47$ bits). Conversely, it will be $\log\left(\frac{1}{1-0.09}\right) = 0.09$ nats if the predator does not see the red colour. In the extreme, a certain event, $p(y) = 1$, brings no surprise (see Figure 1).

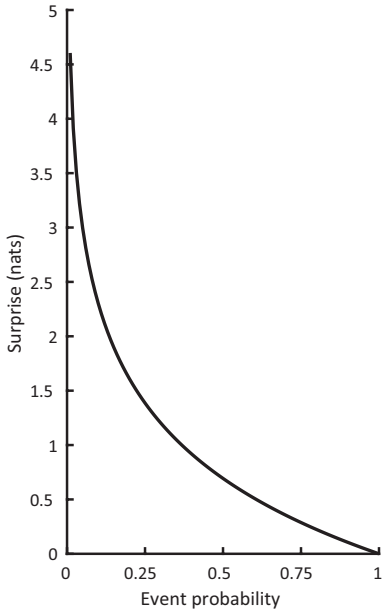

FIGURE 1 Surprisal of an event in function of its probability.

2.3 | Bayesian surprise

In Bayesian inference, the distribution of probability $P(x)$, called the *prior*, corresponds to an organism's prior beliefs about the state x of an aspect of the world it does not have direct knowledge of. Such states are called *hidden states*. Hidden states can be represented by continuous variables, for example when they represent the estimated size of an object, or discrete variables, when only a finite set of values is considered. For example, a predator approaching a prey may associate a probability p (a number between 0 and 1) to the belief that the newly encountered prey is profitable to eat (where x refers to its hidden state, i.e. its profitability) such that $p = P(x = \text{profitable})$. The predator's belief that the prey is unprofitable is therefore $1 - p = P(x = \text{unprofitable})$, as the probabilities of all states ('profitable' and 'unprofitable') must sum (or integrate) to one (Appendix S1). This probability p reflects the uncertainty of the predator about the environment, here about the true profitability of the prey. To perform Bayesian inference and update its prior belief in the light of new observations y gathered through its senses, the organism needs a *generative model*, that is a way to generate predictions about the hidden states of the world x given observations y. Such a generative model is implemented through a *likelihood* $P(y|x)$, which provides a (probabilistic) mapping between hidden states (x) and observations (y). For example, a new observation y may arise from seeing the red colour displayed by a potential prey, which may decrease the predator's belief that the prey is profitable if the predator's generative model associates red coloration with unprofitability (Figure 2).

The crucial point in Bayesian inference is to capture a change of belief using a general *updating rule* provided by Bayes' theorem (Appendix S1). For each event x in the space of hidden states,

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)}, \quad (3)$$

where $P(x)$ is the probability distribution of beliefs before any new data (prior), $P(y|x)$ is the *likelihood*, which quantifies how likely any incoming

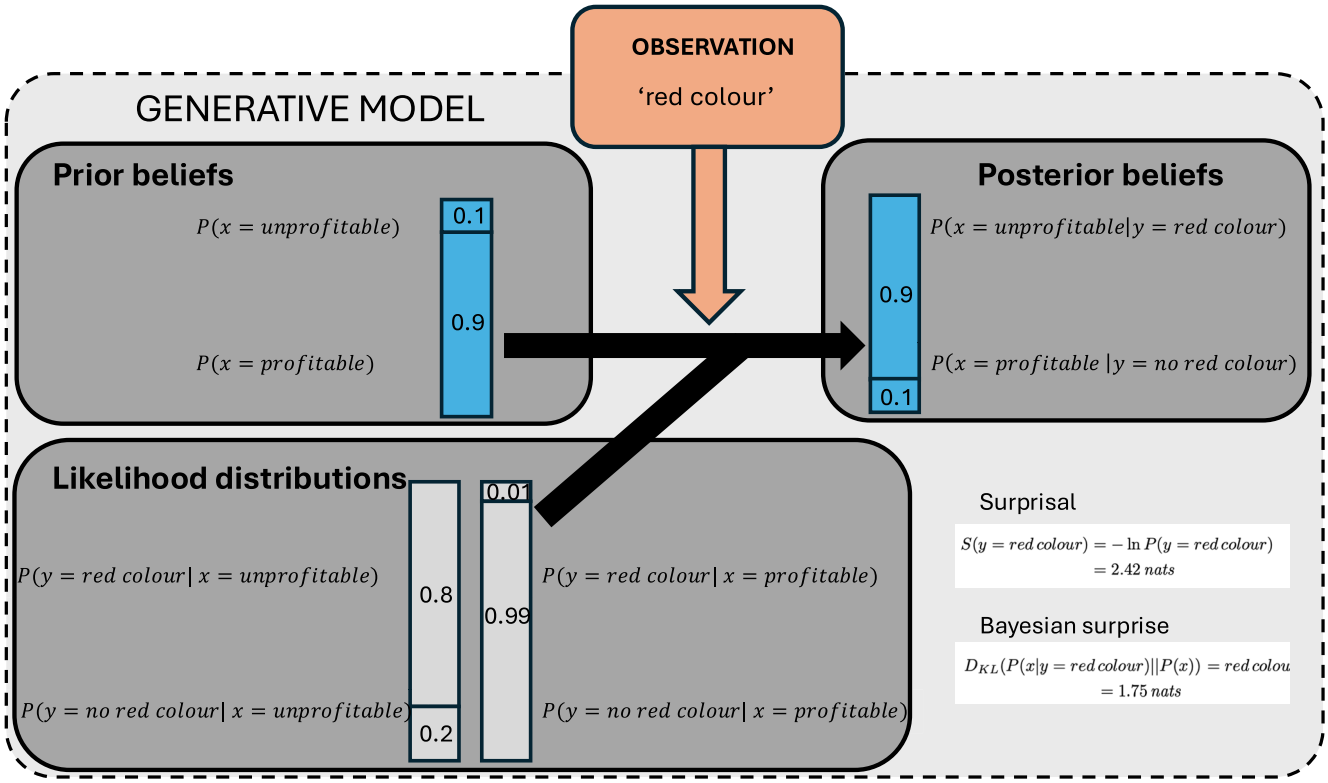

FIGURE 2 A simple generative model for Bayesian inference and active inference. Here we consider a space with two possible hidden states, $x \in \{$'unprofitable', 'profitable'$\}$ and two possible observations $y \in \{$'red colour', 'no red colour'$\}$. The prior distribution gives the predator's prior belief about the possible state of an incoming prey, while the likelihood expresses how the predator probabilistically predicts what observation should be linked to what state (bar lengths not to scale). In Bayesian inference, once the predator sees the red colour ($y =$ 'red colour') it can update its prior belief into its posterior belief. In active inference, the prior has a deeper meaning as it provides the organism's preferred states, which include its conditions for survival. In addition, active inference allows the organism to act so that its percepts match its preferences. Accordingly, when seeing the red colour, the predator may change its belief, by modifying its prior, or change its percepts, by moving away until encountering a prey that does not display the red colour (see Appendices S2 and S3 for the computation of surprisal and Bayesian surprise).

data are under the hypothesis associated to the value x, and $P(y)$ is the *marginal likelihood*, the probability of the observation. Bayes' theorem therefore provides $P(x|\,y)$, which represents the organism's beliefs after having observed data y and is called the *posterior*.

Itti and Baldi (2009) defined Bayesian surprise 'as a measure of how an organism's belief is modified by new observations. They reasoned that new observations do not bring any surprise if the belief of the observer is unaltered, that is its prior belief equals its posterior belief $P(x|\,y) = P(x)$.

By contrast, new observations are surprising if the prior and posterior differ. To measure the mismatch between prior and posterior they used a measure of divergence between probability distributions from information theory, the *Kullback–Leibler divergence* (also called *relative entropy*, Appendix S1):

$$D_{\mathrm{KL}}(p(x|\,y)\ \|\ p(x)) = \sum_x p(x|\,y)\log_2\left(\frac{p(x|\,y)}{p(x)}\right), \tag{4}$$

where the sum is taken over all hidden states. In this formalism, the Bayesian surprise brought by new observations is therefore quantified by a principled measure of disparity between two probability distributions, the prior and the posterior. Accordingly, Bayesian surprise (Equation 4) measures the overall amount of belief updating after making a specific observation whereas surprisal (Equation 2) measures how unlikely that observation was given a distribution of probabilities over all observations. Bayesian surprise is measured in 'wows' (Itti & Baldi, 2009), where a 'wow' corresponds to a factor 2 between $P(x|\,y)$ and $P(x)$, that is to $\log_2\left(\frac{p(x|y)}{p(x)}\right) = 1$ with log taken in base 2. We provide an illustration of the computation of Bayesian surprise in predator's response to a prey defensive display in Section 3.

In the example provided in Figure 2, a predator has a prior belief that a newly encountered prey is unprofitable of $P(x = \text{unprofitable}) = 0.1$. After having seen the red colour displayed by the prey, its belief is updated to $P(x = \text{unprofitable}) = 0.9$, and this change between prior and posterior corresponds to a Bayesian surprise $D_{\mathrm{KL}}(p(x|\,y = \text{red colour})\ \|\ p(x)) = 1.5$ nats (see Appendices S2 and S3 for a derivation).

It is worth noting that the predator's change in belief is possible because its generative model strongly associates red coloration with prey unprofitability (the association created through learned experience or over evolutionary time-scales). If instead the predator had no specific belief about profitability when seeing the red colour ($P(y = \text{red colour}|\,x = \text{unprofitable}) = P(y = \text{red colour}|\,x = \text{profitable}) = 0.5$),

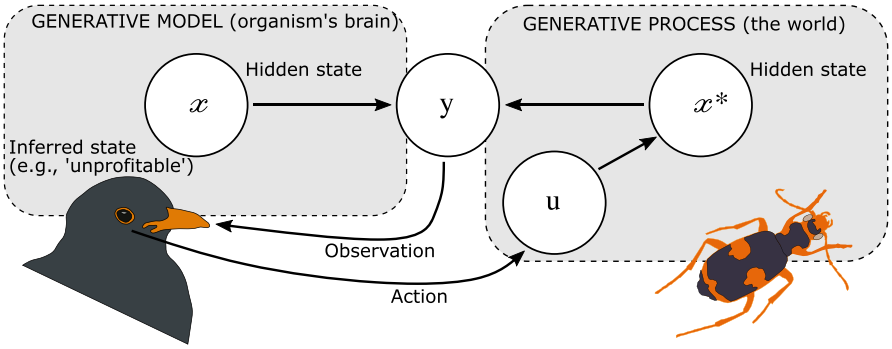

FIGURE 3 Generative model and generative process. Schematics of the framework used in active inference. An organism's generative model describes the way sensory data (observations) *y* could be caused by a physical process in the world, called a generative process. The organism has no direct access to the real cause x^* of its observations. It uses its observations to draw inferences about the possible causes among a range of hypotheses *x* in its generative model. Ideally, the generative model captures the aspects of the generative process relevant to the survival of the organism. Inference then leads the organism to act upon the world through actions *u*, which causes changes in the world and, subsequently, novel observations. Perception and action are tied together in an inferential loop driven by the reduction of surprise (technically, free energy).

it would not have changed its belief, and the same observation would have led to no surprise, $D_{KL}(p(x|\, y = \text{red colour}) \parallel p(x)) = 0$. This shows the importance of the generative model to convey the relevant part of the true process at play in the world, called the *generative process*. Here, the generative process, to which the organism has no access, corresponds to the real mechanisms x^* underlying the occurrence of red coloration in the potential prey (see Figure 3). The information provided by Bayesian surprise is meaningful if the organism's generative model reflects knowledge which is important for survival, for example the general association between red coloration and unprofitability. The theory that brain functions implement inference about the state of the environment through a constant updating of its internal model is called predictive coding and has received empirical support (Friston & Kiebel, 2009; Rao & Ballard, 1999).

In Itti and Baldi (2009), an eye-tracking experimental paradigm was used to explore the links between Bayesian surprise and visual attention. The locations more likely to be associated with a greater surprise in videos—for example where strong changes in colour content or motion happen—were derived using a computational neural model of image processing. The authors found a good correlation between these locations and the locations where human observers directed their gaze, concluding that uncertainty, as captured by Bayesian surprise, attracts attention.

2.4 | Surprise in active inference

Active inference proposes a unifying principle to understand behaviour in which surprise plays the central role. This principle states that an organism's behaviour results from a close coordination of perception and action with the imperative of minimising surprise (Friston, 2010; Parr et al., 2022). In this framework, the concept of surprise is expanded beyond that of a mismatch with expectations. Surprise is defined in reference to preferred sensory inputs (referred to as preferred states) expressed as a probability distribution over a range of possible sensory inputs, of which the preferred states are associated with a higher probability. This distribution encapsulates the essence of priors in adaptive behaviour: a prior corresponds to the states favoured by the organism, which are adaptive and therefore include its conditions for survival. Surprising states, which differ too much from the prior, are less likely to be pursued because they pose a threat on the organism's integrity. For example, the distribution of preferences over temperatures for endotherms will peak at an optimal temperature. Temperatures below or above that reference temperature will be associated with a lower 'preference' (probability) and hence a higher surprisal (Figure 1). Of central importance for understanding behaviour, active inference expands the Bayesian perspective of perception as inference by also including action to minimise the surprise generated from inference. For example, when sensing threatening temperatures, endotherms will act to return to less surprising states (e.g. by physiological or behavioural changes). By giving perception and action the aim to fulfil the same objective of avoiding surprise instead of working in isolation (Friston, 2010), active inference has the potential to offer a novel, more dynamic perspective on predator–prey interactions. Active inference predicts that, when facing surprising states, organisms can change their belief about the environment (perception) or/and act to sample the environment until their perceptual input matches their expectations, for example by orienting their gaze towards the surprising event to gather more information on its nature or by simply moving away to avoid a potential threat (action). In the above example of a predator seeing the red colour displayed by the prey, the predator faces two options to minimise surprise: (1) changing its beliefs (preference) about red coloration, which will decrease its surprise when next seeing the red colour to $S(y = \text{red colour}) = 0.33$ nats; (2) 'acting on the world' by going away until sampling a prey that does not display the red colour (action } $u = \text{go away}$' in Figure 3), with associated surprise $S(y = \text{no red colour}) = 0.05$ nats. This action will correspond to the imperative of minimising surprise, which will maintain the predator

in its preferred state of not having ingested a possibly harmful prey. In any case, prey defences that elicit surprise have the potential to change the behaviour of the predator.

When the probability distributions in the generative model are more complex than in the example shown in Figure 2, the computations required to minimise surprise may become intractable. To address this, active inference uses a more tractable proxy called free energy, which is always greater than and approximates surprise (see Appendix S6). Active inference thus involves minimising free energy through action and belief updating. Free energy can be decomposed into key components relevant to cognition, such as pragmatic value (exploitation) and information gain (exploration) (Friston et al., 2015; Schwartenbeck et al., 2019).

2.5 | Adaptation and learning in active inference

The overarching principle guiding action and perception in active inference, specifically the imperative to minimise surprise, extends directly to learning. More explicitly, all the parameters of the generative models involved in active inference can be modified through incremental changes as an organism acquires experience of its environment. These changes enable free energy to decrease after repeated experiences with the environment (Friston et al., 2016). Accordingly, the two central constructs of active inference, the prior and the likelihood, can change through evolution and development for inference through perception and behaviour to be adaptive (Constant et al., 2018). The prior or distribution of preferred states, $P(x)$, should adapt to reflect the survival conditions of an organism within its niche. On an ecological time-scale, some aspects should not change upon encountering new experiences, for example the narrow preference for a precise body temperature for endotherms. However, other aspects might change to facilitate adaptive decisions. For example, $P(x)$ may include a broader acceptance of toxic prey by predators when their energetic reserves are low, or alternative prey is scarce (Sherratt, 2003; Skelhorn, Halpin, & Rowe, 2016). Similarly, the likelihood $P(y|x)$ which maps the states of the world to their predicted percepts should inform the organisms about how percepts are likely to be linked to preferred or nonpreferred states of the world. In the case of warning signals, for example, selection should favour likelihood functions that clearly express a link between prey phenotypic traits (e.g. conspicuous coloration) and unprofitability.

2.6 | Active inference, ultimate and proximate explanations for surprise

Tinbergen's framework, and the dichotomy between ultimate and proximate explanations, have been central for understanding animal behaviour (Bateson & Laland, 2013; Tinbergen, 1963). Because it allows modelling at different time-scales, from evolutionary to individual real-time, active inference proposes a biologically plausible theory that applies to both ultimate (the imperative of avoiding surprise for an organism to stay within its living requirements—optimal body temperature, energy level, avoiding toxins, etc.) and proximate (e.g. neural activity to encode surprise and implement changes in beliefs, updates of synaptic connections underlying learning) processes (Ramstead et al., 2018). The prior expectations and likelihoods involved in inference are optimised at different time-scales, the shorter time-scale for somatic changes (learning), and the longer, evolutionary time-scale. Heritable priors are under selection pressure to engender adaptive behaviour (Friston, 2010). We refer to Ramstead et al. (2018) for a detailed correspondence between Tinbergen's four levels of explanation (mechanism, ontogeny, adaptation and phylogeny) and the constructs in active inference. In the next section, we will argue that both ultimate and proximate causes have the potential to further our understanding of the role of surprise in predator–prey interactions.

3 | ROLE OF SURPRISE IN ANTIPREDATOR DEFENCE

In this section, we identify known types of antipredator defences that match definitions of surprise and map those defences on the predation sequence (Endler, 1991). Surprise-related aspects may be found across the predation sequence including: (1) encounter, (2) detection, (3) identification, (4) approach (attack), (5) subjugation and (6) consumption (Endler, 1991). Defences that have the potential to benefit from predator surprise take place mostly at the approach and subjugation stages (Figure 4). Here, we focus on two categories of such defences, which differ in the mechanism generating surprise: (a) defences that involve a sudden change in prey appearance and affect immediate predator responses and (b) defences that increase prey unpredictability over repeated encounters and affect predator learning and decision-making strategies. We explore the ways they can affect predator behaviour and increase prey survival, specifically illustrating how Bayesian surprise can be estimated in a prey defensive display, and discuss how surprise can be understood within the framework of active inference, with an illustrative simulation of avoidance learning. See Table S1 for a detailed overview of antipredator defences that have the potential to elicit predator surprise, and examples of how these defences are used by prey species.

3.1 | Defences that involve a sudden change in prey appearance

Many surprise-related antipredator defences involve a sudden change in prey appearance or behaviour, which often elicits immediate predator responses. Deimatic displays are a classic example where prey respond to the approaching or attacking predator with specific displays, which trigger an unlearned predator response causing it to slow or stop the attack (Drinkwater et al., 2022). Prey using deimatic displays are commonly cryptic

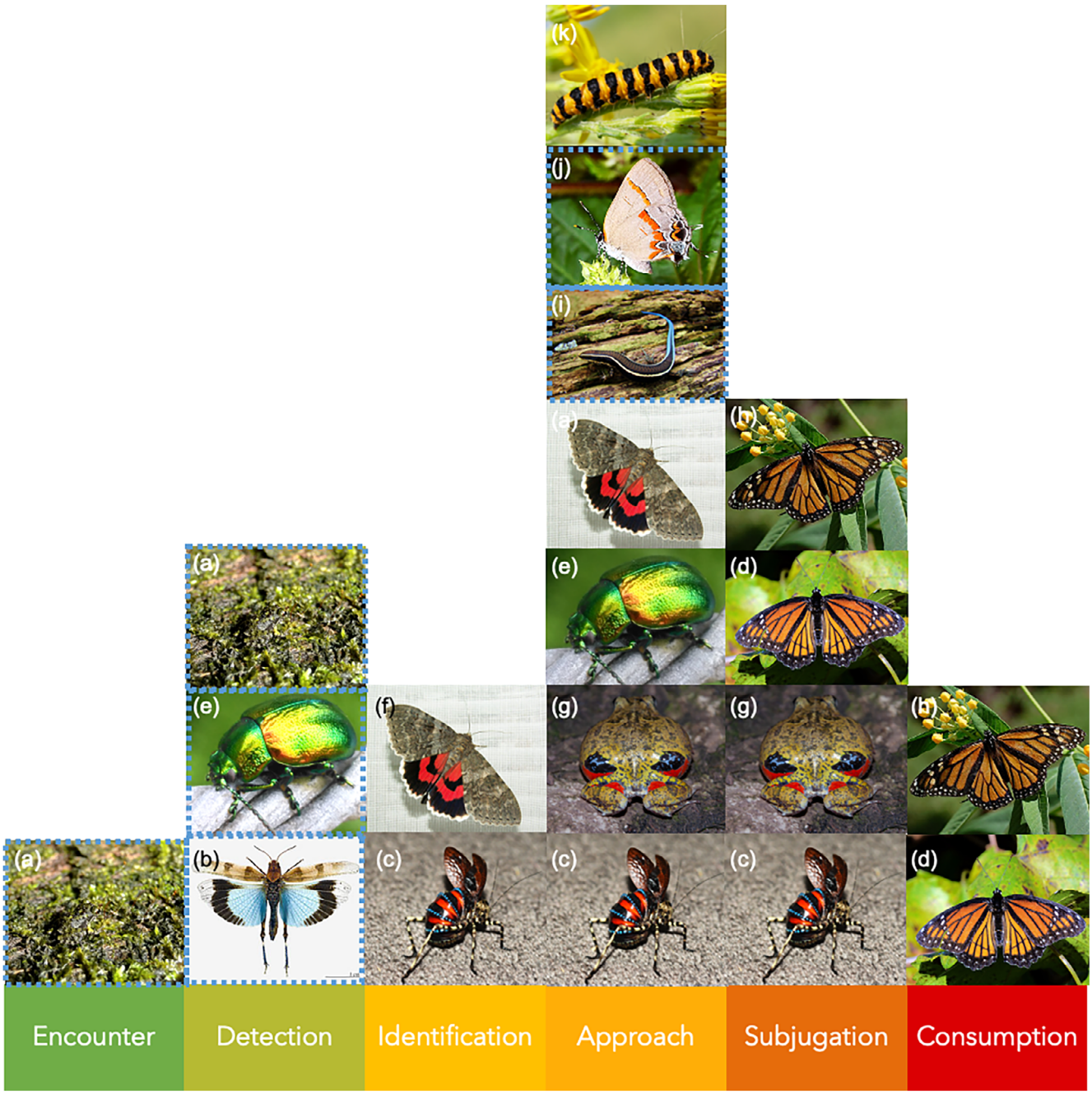

FIGURE 4 Examples of surprise-related prey defences along the predation sequence. (a) camouflage (background matching) in the frog *Pristimantis zeuctotylus* (©Bibiana Rojas); (b) flash coloration in the blue-winged grasshopper *Oedipoda caerulescens*; (c) deimatic display and/or hidden warning signal showing conspicuous coloration under an otherwise cryptic phenotype in the mountain katydid *Acripeza reticulata* (©Kate Umbers); Müllerian mimicry involving (d) viceroy *Limenitis archippus* and (h) monarch *Danaus plexippus* butterflies, which possess different types of chemical defences; (e) iridescent coloration in the tansy beetle *Chrysolina graminis*; (f) deimatic display in the underwing moth *Catocala nupta*; (g) putative deimatic display in the Colombian four-eyed frog *Pleurodema brachyops* (©Giovanni Chávez-Portilla), involving eyespots and body inflation; (i) deflection markings in the Magdalena River tegu *Tretioscincus bifasciatus* (©Luis Alberto Rueda); (j) false head in the red-banded hairstreak butterfly *Calycopis cecrops*; (k) distance-dependent pattern blending in the cinnabar moth *Tyria jacobaeae* caterpillar: Cryptic when seen from afar, conspicuous at close range. Photographs with the blue frame denote cases of defences whose relationship with surprise at a particular stage of the predation sequence is ambiguous (see Section 3.3 for further clarification). All photographs obtained from Wikimedia Commons unless otherwise stated.

at rest and the displays include revealing previously hidden bright colour markings, accompanied by specific movements and postures (Drinkwater et al., 2022; Figure 4c,f,g).

Surprise could also play a role in hidden warning signals where conspicuous colours are combined with camouflage and revealed only upon predator approach (Loeffler-Henry et al., 2023;

Figure 4c), in iridescent coloration where the hue and intensity of colours change depending on the viewing angle of the approaching predator (Kjernsmo et al., 2022; Figure 4e), and in distance-dependent pattern blending where the prey appear cryptic from a distance and conspicuous at close range (Barnett et al., 2017, 2018; Figure 4k). All these defences include sudden changes in sensory input and conspicuous components, which are atypical in natural environments and evoke surprise (Penacchio et al., 2024).

The unexpected change in prey appearance may trigger a predator's startle response. The startle response is defined as an animal's reaction to the sudden appearance of a salient stimulus and characterised by rapid onset and by causing an immediate interruption of any ongoing activity (Drinkwater et al., 2022; Forrester & Broom, 1980; Koch, 1999). Behavioural patterns associated with startle response include muscle contractions, limb flexion and crouching, followed by a short period of immobility, which makes it possible to distinguish the startle response from the escape response, even though the two are often linked (Forrester & Broom, 1980; Koch, 1999). Behaviours consistent with the description of the startle response have been observed in various predators in reaction to defences involving a sudden change in prey appearance, though evidence is mostly limited to avian species (Holmes et al., 2018; Ingalls, 1993; Kang et al., 2017; Kim et al., 2020; Umbers et al., 2019). Investigating startle responses in other predator taxa therefore provides a promising area for future research.

The adaptive significance of the predator's startle response is to protect it from a potential threat, as the response is likely to prevent injury and facilitate an escape (Koch, 1999). Startle response may force predators to focus on a surprising stimulus because it interrupts predator activity and is often followed by the orienting response, a behavioural and cognitive response used to gather information (Sokolov et al., 2002). The strength of the orienting response is contingent on the amount of cognitive processing dedicated to a stimulus (Kaye & Pearce, 1984). Since the response strength can be measured as a frequency or duration of specific behaviours (Kaye & Pearce, 1984), it could serve as a proxy for quantifying surprise.

Although the startle response is innate and cannot be eliminated (Koch, 1999), its intensity usually decreases through habituation (Davis, 1970; Shettleworth, 2010), which can make the display less effective across repeated encounters (Ingalls, 1993; Schlenoff, 1985). How quickly the predators habituate to surprising displays, how well and long displays are remembered, and how broadly they are generalised may depend on several factors including prey abundance, display components, and the age, personality and previous experience of the predators (Umbers et al., 2019).

Defences that involve a sudden change in prey appearance may also trigger predators' escape responses (Drinkwater et al., 2022). This is particularly the case for deimatic displays, which may cause the predator to misclassify the prey as a potential threat (Skelhorn, Holmes, & Rowe, 2016). For example, eyespots may resemble the eyes of a larger predator (Cott, 1940; De Bona et al., 2015). A sudden display of eyespots by cryptic prey frequently results in predators' recoiling, jumping back and moving away (Drinkwater et al., 2022). This behavioural response has been observed in various predators, including birds (De Bona et al., 2015; Olofsson et al., 2013) and mammals (Olofsson et al., 2013). Predator escape responses can be also triggered by prey displays in nonvisual modalities, such as sounds, which can be quantified in terms of Bayesian surprise (Itti & Baldi, 2009).

The stimulus misclassification by a predator is particularly likely when the display is highly surprising and requires a fast response even though potentially incorrect given the signaller's true nature (Trimmer et al., 2008). This might explain why some deimatic displays are only deployed upon close contact with the predator (Umbers & Mappes, 2015; Vallin et al., 2005; Figure 4) when there is maximal surprise, and the perceived imminent threat is likely to elicit predator escape response. Furthermore, some components of deimatic displays (such as sounds) are stronger at a closer distance (Drinkwater et al., 2022). The initial escape response may be followed by subsequent prey inspection (Kang et al., 2017; Vallin et al., 2007). The longer the predator observes a potentially harmful prey without being attacked, the less likely the display corresponds to a real danger (Sherratt et al., 2023). The optimal latency to approach a potentially dangerous prey decreases over repeated encounters in which the prey proves to be harmless, as the predator's expectation changes according to new observations (Sherratt et al., 2023), making the display less effective. If deimatic displays involve a rapid increase in apparent prey size, they may also trigger predator looming reflex, that is evasive response to rapidly approaching large and potentially dangerous objects (Drinkwater et al., 2022; Yamawaki, 2011). In this case, the displays may also be most effective at a closer distance, where the perceived increase in the stimulus size and the corresponding predator surprise is maximised.

These defences may confer benefits to prey by reducing the speed, or likelihood of an attack, giving the prey time to escape. Most of our knowledge about the effects of surprise in prey defences on predator behaviour, however, comes from experiments under laboratory settings with artificial prey (Drinkwater et al., 2022). Only a few studies have shown that surprising predators through defence strategies such as deimatic displays can increase the survival of real prey in encounters with avian (Umbers et al., 2019; Vallin et al., 2005) and mammalian predators (Olofsson et al., 2012).

Figure 5 illustrates how the mathematical tools of Section 2 can be used for the empirical investigation of deimatism. It shows how to estimate Bayesian surprise from a video of an artificial moth that simulates a deimatic display. The computation is based on how changes in luminance in the visual stimulus elicit changes in neuronal activity in the visual system of an avian predator. This computation can be generalised in different ways, for example, by considering the colour content of the stimulus in addition to its luminance content. As acknowledged in Itti and Baldi (2009), it is also possible to compute Bayesian surprise for other modalities than vision, for example acoustic stimuli. These measures of surprise are of direct interest as they can be contrasted to and regressed against typical measures

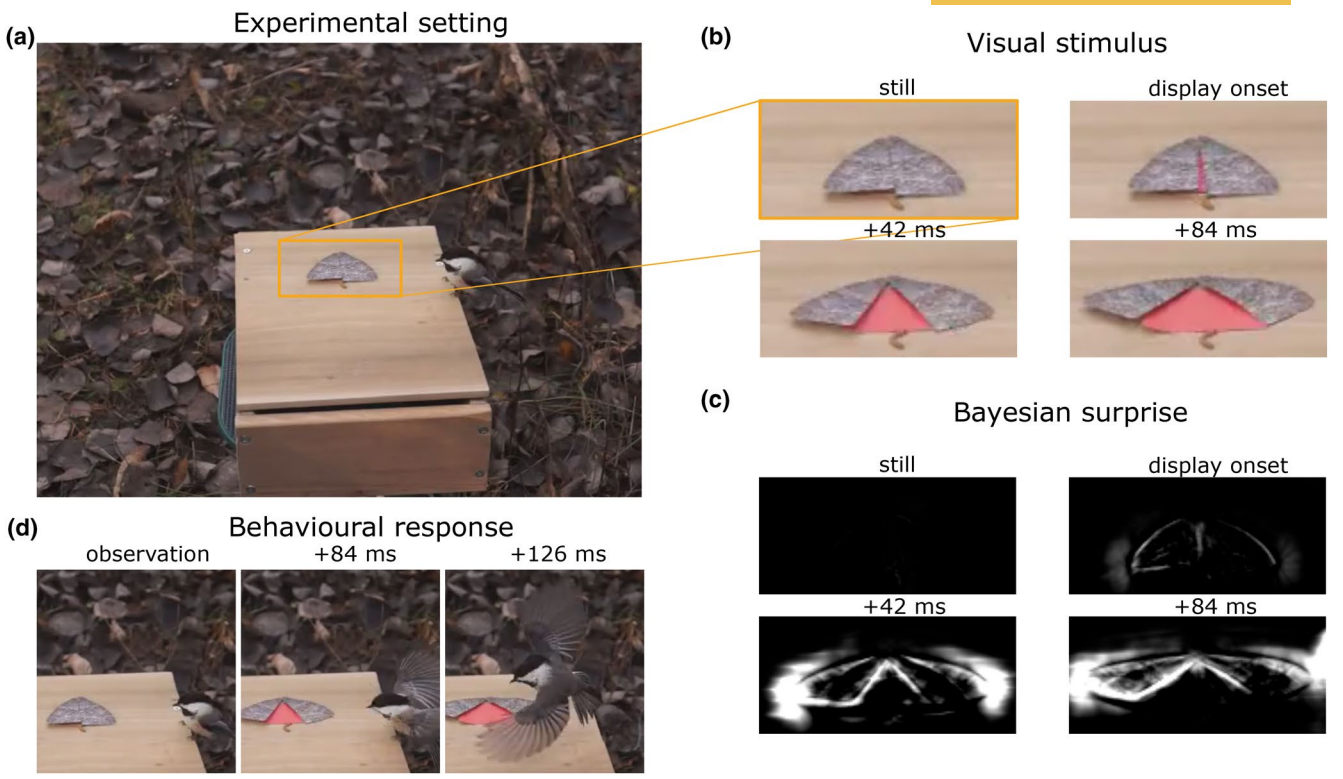

FIGURE 5 Quantification of (Bayesian) surprise in an empirical experiment on deimatism. (a) Setting of the behavioural experiment using a remote-controlled robotic moth model to test the responses of black-capped chickadees *Poecile atricapillus* towards the deimatic display of *Catocala* moths (see Kang et al., 2017 for details). (b) Close-up of the artificial prey for the video frames before, at, and 42, 84 ms after display onset. (c) Corresponding Bayesian surprise for the frames in (b) (after Itti & Baldi, 2009). Here Bayesian surprise is computed by considering abstracted neural units that encode the likelihood of incoming data at each pixel location (reminiscent of retinotopic locations in the visual system of the predator). The lighter the colour of a pixel, the greater surprise, as measured in 'wows' (see Appendix S5 for details). While surprise is computed locally (at each retinotopic location), its effect on a predator (e.g. attracting its attention) is likely to reflect the average surprise within specific areas of the scene. (d) Predator's behavioural response. Original video (see Video S1) courtesy of Changku Kang.

of predators' behaviour when facing prey defence such as rejection rate, reaction time and latency to attack.

3.2 | Defences that increase prey unpredictability across encounters

Surprise can also play a role in defences that may not involve sudden changes in prey appearance, but where prey profitability is difficult to predict across encounters. In this case, surprise is generated by variation in prey defences at the subjugation and consumption stages, which leads to changing contingencies between prey appearance (e.g. warning signals) and unprofitability. In automimicry, the intraspecific variation in unprofitability occurs, with individuals varying in the presence, strength and/or type of their chemical defence (Guilford, 1994; Speed et al., 2012; Svennungsen & Holen, 2007). Variation in chemical defences is also observed across species that share similar warning signals. This includes parasitic relations, such as Batesian mimicry, where an undefended mimic resembles a defended model species and increases predation on the model (Bates, 1862; Lindström et al., 1997), and quasi-Batesian mimicry, where both species are chemically defended but the less defended species acts in a Batesian manner (Rowland et al., 2010; Speed, 1993). Variation in defences also occurs in mutualistic Müllerian mimicry if the mimics that share similar warning signals are equally defended against predators but possess different defensive chemicals (Chouteau et al., 2019; Figure 4d,h). Surprise may affect predator behaviour in a way that could potentially change the mimetic relations of unequally defended species from parasitic to mutualistic, leading to quasi-Müllerian mimicry, where the presence of an undefended mimic is beneficial to the model, or super-Müllerian mimicry, where the model benefits more from a less defended mimic than from equally defended one (Balogh et al., 2008; Speed & Turner, 1999).

Surprise generated by the mismatch between predator's prior expectations and actual prey defences can increase predator uncertainty about prey profitability and make the previous experience less relevant, enhancing the rate of subsequent learning (Balogh et al., 2008; Courville et al., 2006). Therefore, predators may learn to avoid aposematic prey faster if prey individuals possess different rather than identical chemical defences (Skelhorn & Rowe, 2005). Using a modification of the Rescorla–Wagner model of associative learning, Balogh et al. (2008) confirmed that variation in prey unpalatability leads to a faster increase in predator attack inhibition compared to a constant level of unpalatability. Fast learning to avoid

prey with variable chemical defences may protect the predators from ingesting a high dose of toxins in a situation when it is difficult to predict the prey's toxin content (Skelhorn & Rowe, 2005). If the predators learn to avoid the prey with similar appearance but variable defence faster than the prey with the uniform defence, then the Müllerian mimics that differ in their defensive chemicals may be better protected than those that share the same chemical (Sherratt et al., 2004; Skelhorn & Rowe, 2005). Likewise, faster avoidance learning by predators may allow for the maintenance of intraspecific variation in defence when individuals differ in their chemical profiles (Speed et al., 2012).

Following the experience with defence unpredictability, predators may decide to avoid particular prey even if some individuals are undefended and prefer to consume moderately toxic prey rather than prey which could vary from non-toxic to highly toxic (Barnett et al., 2014). Again, this behaviour may help the predators to better control their toxin load, and minimising surprise over successive encounters may represent the cognitive mechanism that underpins this strategy. Active inference predicts that an increased randomness in reward leads to a more conservative sampling behaviour (Schwartenbeck et al., 2019). Predator avoidance of prey with variable defences would allow for the persistence of less defended and undefended individuals among defended prey (Svennungsen & Holen, 2007) and benefit unequally defended mimics (Balogh et al., 2008; Barnett et al., 2014).

Another way in which predators may respond to the unpredictability of prey defences is the go-slow strategy (Guilford, 1994; Holen, 2013). Instead of learning to avoid the prey with unpredictable defences, predators may learn to approach and handle such prey cautiously to avoid being harmed and to get more accurate information about prey defences (Brower & Fink, 1985; He et al., 2022; Skelhorn & Rowe, 2006). The proportion of prey sampled and rejected by avian predators was highest at the moderate frequencies of the defended prey, supporting the hypothesis that this behaviour is linked to predator uncertainty (He et al., 2022). Because the go-slow strategy allows predators to sample and reject defended prey individuals, it may lead to selection against less defended or undefended mimics (Guilford, 1994; Skelhorn & Rowe, 2006).

3.3 | Other defences that may have the potential to generate surprise

In camouflage, behaving in unexpected ways can enhance survivorship. For example, when a prey has a choice of two patches to hide in, it should not always hide in the patch where it is best concealed because otherwise a rational predator will only search there. Instead, the optimal solution for the prey is to reduce its predictability by occasionally choosing the patch where it is less well concealed (Gal & Casas, 2014; Nahin, 2007; Figure 4a).

Some of the defences that are typically deployed at early stages of the predation sequence and affect prey detection, such as iridescent coloration (Kjernsmo et al., 2020; Figure 4e) and flash displays (Sherratt & Loeffler-Henry, 2022; Figure 4b) involve a sudden change in predator's visual input. These defences therefore include a component of surprise (e.g. that can be quantified using Itti and Baldi's (2009) framework to compute Bayesian surprise). How this surprise-related component underwrites the effect of the defence is not clear, however. One possibility is that the change in prey appearance may lead to predator confusion (Drinkwater et al., 2022). Iridescent coloration makes prey detection more difficult, as it produces inconsistent shape cues and may thus work as dynamic disruptive camouflage (Kjernsmo et al., 2020). In flash displays, otherwise cryptic prey reveal conspicuous colour markings when fleeing from predators, which gives an impression of the prey being conspicuous (Sherratt & Loeffler-Henry, 2022) and makes the prey difficult to find once it has settled (Loeffler-Henry et al., 2018). Here, the effect of surprise may be linked to the prey seemingly disappearing when it resumes its cryptic coloration. In contrast to deimatic displays, which are usually deployed at a closer distance (Drinkwater et al., 2022), flash signals are more effective at longer distances, when the predator is unaware of the prey's cryptic resting appearance (Loeffler-Henry et al., 2021; Sherratt & Loeffler-Henry, 2022).

Some surprise-related defences may increase prey survival by confusing the predator such that it does not know exactly where to strike during an attack. For example, Murali (2018) found that human 'predators' were not able to catch the stimuli with dynamic flash coloration as frequently and accurately as stimuli whose colour did not change or matched the background, suggesting that such colour change may confuse the predator about the location of the prey. The finding that the magnitude of this benefit depends on the unpredictability of the prey movement (Murali & Kodandaramaiah, 2020) suggests that this defence may include surprise. In erratic, 'protean' escape behaviour, unexpected rapid changes in prey escape trajectory make it hard for a predator to predict prey movement and follow the escaping prey (Humphries & Driver, 1970; Richardson et al., 2018). The unpredictability of escape behaviour can be quantified using surprisal (Moore et al., 2017). In addition, false heads (Figure 4j) and deflective markings (Figure 4i) can make predators misdirect their attacks and lead to failure to capture the prey that is moving in an unexpected direction (Humphreys & Ruxton, 2018).

3.4 | Surprise and multiple defences

Finally, many prey species possess multiple defences that may be deployed either simultaneously or sequentially (Caro et al., 2016; Ruxton et al., 2018). One of the most intriguing questions about multiple prey defences is what determines their co-occurrence in particular prey and their timing relative to predator attack (Kikuchi et al., 2023; Wang et al., 2019). Multiple defences seem inherently surprising because in many cases predators cannot predict which combinations of defences prey are likely to use. Each defence component, for example a bright colour, leads to surprise because it elicits a percept that deviates from those typical in natural environments

(low probability leading to high surprisal), and because it suddenly modifies the distribution of probability over hidden states (Bayesian surprise) in a way that does not match the predator's preferences (higher free energy). Each defence component is therefore associated with neural activity signalling surprise, as prediction error and belief updating (Friston & Kiebel, 2009). Surprise is likely to be enhanced by the co-occurrence of the defensive components, as is often the case with multimodal percepts (Stein & Meredith, 1993; Stein & Stanford, 2008). This can increase the time required for information processing, and potentially result in sensory overload, where the predator's perception system is overwhelmed by receiving more information than can be processed at one time (Hebets & Papaj, 2004). The enhanced ability of multimodal or multicomponent signals to surprise predators may have been a key factor in the selection pressure leading to the evolution of this type of complex signalling.

3.5 | Understanding surprise in predator–prey interactions through active inference

Active inference offers a unifying perspective on all the aspects of surprise in antipredator defences introduced so far. To understand this, let us recall that in active inference: (1) surprise corresponds to a deviation from preferred (expected) observations, and (2) behaviour follows the imperative of minimising surprise (or free energy). In this framework, prey with traits that elicit unpreferred observations in predators are likely to be avoided. This concept is illustrated in Figure 6, which presents a simulation of avoidance learning in a bird encountering profitable and unprofitable prey with distinct phenotypes. Here, the bird acts as an active inference agent, meaning its behaviour follows the imperative of minimising free energy. This corresponds to the need to secure food while avoiding toxins. In this context, minimising free energy equates to minimising surprise, where surprise arises from encountering unpreferred outcomes, here hunger or sampling an unprofitable prey. See Appendix S6 for general methods, and Figures S4 and S5 for two variants of this simulation that illustrate modelling of dietary conservatism (Marples et al., 1998) and the effect of the predator's learning rate in learning the association between prey phenotype and profitability.

In terms of proximate mechanisms, active inference predicts accurately the interruption of ongoing activity observed in the startle response, followed by the orienting response (see Section 3.1). In active inference, surprise triggers an update of beliefs about the world, involving neural message passing between sensory-processing units ('bottom-up') and higher-order neural areas encoding beliefs about the world ('top-down') until a coherent state with no prediction error is achieved (Friston & Kiebel, 2009). The stronger the surprise, or mismatch with expectation, the more neural activity and time are required for this update (Schwartenbeck et al., 2015). The orienting response towards salient stimuli reflects the need to gather information about the source of surprise. Multimodal signalling further enhances surprise, increasing the cognitive load of the updating process (see Section 3.4). In active inference, all these phenomena lead to learning, where surprise drives belief updating, as in the Rescorla–Wagner model. Active inference incorporates the latter model by showing how surprise triggers neural updates, minimising prediction errors and facilitating learning (Anokhin et al., 2024).

Active inference also effectively predicts predator behaviour when facing prey unpredictability (see Section 3.2). For example, when an organism encounters greater uncertainty about the likelihood of receiving a reward from an action, it tends to prioritise sampling cues before directly pursuing the action (Schwartenbeck et al., 2019). Similarly, the theory suggests that after an initial period of sampling, called active learning, an organism will favour unambiguous states even if the average reward is similar (Schwartenbeck et al., 2019). More broadly, active inference provides a computational framework for exploring how environmental uncertainty and an organism's tendency to sample uncertain options influence the trade-off between exploration and exploitation. This framework is similar to those used in reinforcement learning (Sutton & Barto, 2018) but is linked to the neurophysiology of cognition. The organism's generative model, which shapes its representation of the world and guides behaviour, balances two components: one that drives information-seeking and uncertainty reduction (exploration), and another that favours the realisation of preferences (exploitation) (Schwartenbeck et al., 2019).

The key concept for understanding how active inference can address empirical questions in predator–prey interactions is *computational phenotyping*. The behaviour simulated in Figure 6 is determined by a set of parameters that define the generative model guiding the predator's actions. Computational phenotyping involves inverting the mapping from parameters to behaviour (see Parr et al., 2022). In practice, this means we can extract these parameters from observed behaviours in empirical experiments, both at individual and group levels (Schwartenbeck & Friston, 2016)—for instance, from the behaviour of each bird in an experiment like the one simulated in Figure 6. When applied to preferences, this method provides a numerical description of what each bird seeks or avoids (see Appendices S6 and S7 for details). Computational phenotyping can be applied to any component that defines active inference agents, including learning (Figure 6d) and the exploration–exploitation trade-off (Schwartenbeck et al., 2019). Although so far used to study human behaviour (see Parr et al., 2022), model-based data analysis in active inference holds significant potential for deepening our understanding of predator–prey interactions, particularly when surprise is involved.

4 | CONCLUSIONS AND FUTURE DIRECTIONS

The mismatch between predator expectations and the information presented by a prey can generate surprise, which can influence predator decisions and behaviours and ultimately affect prey survival. We have shown numerous ways in which prey defences can elicit

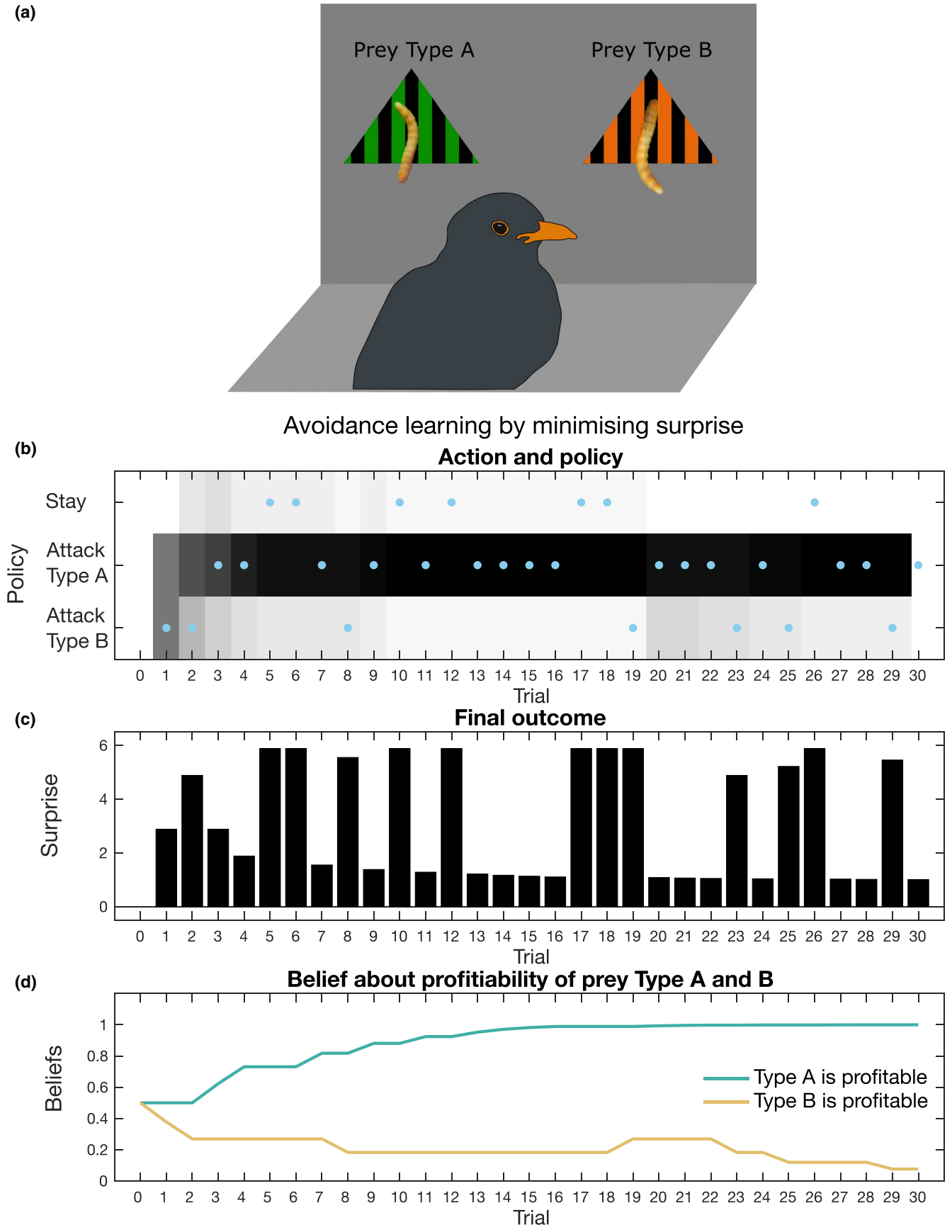

FIGURE 6 Active inference modelling of avoidance learning in a naïve predator. A bird is trained to attack artificial prey items (paper moths with pinned mealworms; Halpin et al., 2020). It then must repeatedly choose between two different prey types, 'Prey Type A' printed with a conspicuous but not typical warning coloration and with a non-modified (profitable) mealworm, and 'Prey Type B', showing a typical warning coloration and with a mealworm treated with quinine (unprofitable). We use the active inference framework to model avoidance learning (see Appendix S6 for details). (a) Schematics of the experimental setting. (b–d) Example simulation of an active inference bird over 30 trials with, in (b), the action chosen by the bird among the three possible actions ('Stay', 'Attack prey Type A', 'Attack prey Type B'; blue dots) and probability beliefs for actions (darker shades indicate higher probability); (c) surprise at the end of each trial; (d) Bird's belief about the profitability of Type A and B over time. The simulation is drawn from a generative model. Crucially, the process can be inverted to recover a generative model from empirical behaviour—computational phenotyping, see text—which allows assessment of the role of surprise in antipredator defence from the perspective of active inference (see Appendix S7 for details).

surprise, introduced metrics by which to measure surprise in predators, and explained why and how considering the concept of surprise is essential in interpreting predator responses to prey defences, although the field remains ripe for continued research.

Metrics to quantify surprise are not well resolved among different sensory modalities and across divergent predator species. The field would benefit greatly from a more thorough understanding of both the predator and prey perspectives related to surprise and its influence on decisions and behaviour. For example, how prey decide to deploy surprising defences, how predators process novel information and under what conditions predators take risks to continue a predation attempt even if their expectations are not met versus when they reject a potential prey. Similarly, we need a better understanding of whether surprising defences can elicit the same reaction in different taxa of predators, or if reactions and behaviours differ and are predator taxon specific. Another question is whether predator individual traits (such as personality) play a role in their reactions to surprise in prey defences.

Understanding the timing of surprise in the predation sequence would further clarify when and under what conditions prey choose to deploy defences which change predator expectations, and how this affects the strength of predator response to surprise. Likewise, we lack a clear understanding of the relationship between surprise and multiple defences. Further research is needed to understand which sequentially deployed defences precede and follow the surprise-generating defences, and which combinations of simultaneous defences are most effective in eliciting predator surprise.

Habituation may attenuate the effects of surprise in two ways. First, in terms of when surprising displays are triggered in prey over repeated interactions with predators. Second, in the effect of surprising prey displays on experienced predators whose expectations may have altered to include what were previously surprising prey displays. Unfortunately, we still know relatively little about predator habituation to surprising elements in prey defences, especially under field conditions, mainly because of the difficulties of observing surprise-involving interactions and subsequently following individual predators to collect additional observations.

Active inference offers a general, principled framework to understand cognition and behaviour, which links together perception and action with the objective for an animal to minimise surprise,

BOX 1 Outstanding questions on the role of surprise in predator-prey interactions

- Q1. Do the defences that involve a sudden change in prey appearance elicit stronger predator surprise if they are timed just before or when the predator makes contact with the prey compared to being deployed earlier in the encounter?
- Q2. Do predators habituate at a similar rate to surprising defences in different modalities (e.g. visual versus acoustic)? Can multimodality of display components protect surprising defences from habituation?
- Q3. How broadly do predators generalise over surprising defences? Can polymorphism in a salient defence component prevent habituation? Can imperfect mimetic species similarly take advantage of this to prevent habituation which could explain their persistence?
- Q4. What types of defence unpredictability (such as variation in quality versus quantity of defence chemicals) can enhance predator avoidance learning and/or lead to an 'unpredictability aversion'?
- Q5. Since the distribution of preferences used in active inference is shaped by a general preference for what is typical in natural environments, the stimuli that create surprise are likely to differ from the typical stimuli in the environment. Would it be possible to predict the characteristics of prey defences that would elicit maximum surprise relative to predator environment and perception? Can we find a common underlying principle for the design of surprising prey defences in terms of degree of difference from the most common stimuli in natural environments?
- Q6. Can the effects on predators of the different types of surprise elicited by prey defences be unified under the imperative of minimising surprise from observations in active inference?

We discuss in Appendix S7 how the frameworks proposed in this study, in particular active inference, may be applied to tackle these questions.

that is the mismatch between current observations and those corresponding to its preferred states, which reflect evolutionary adaptation. When confronted with surprising observations an animal faces two possibilities: altering its beliefs such that they agree with current observations, or modifying its behaviour until updated information is in line with prior beliefs or preferences. Both these pillars of active inference can benefit prey because they would entail an interruption of predator attack and/or a change in the predator sampling strategy. As such active inference provides a promising avenue for better understanding the role of surprise in predator–prey interactions. Because it applies to different time-scales—from real-time encounters throughout ontogeny, to evolutionary time-scales—and is associated with clearly defined plausible neural mechanisms, active inference can address each of Tinbergen's four levels of inquiry for understanding animal behaviour. Analysing data from empirical experiments using active inference models and computational phenotyping could provide a unified account for the cognitive ecology of surprise, from fundamental aspects in adaptive behaviour such as the trade-off between exploitation and exploration, to the timing of defence deployment in predator–prey interactions, to predictions about the neurophysiology of startle and escape responses. In Box 1, we outline what we see as key questions concerning the role of surprise in antipredator defences.

AUTHOR CONTRIBUTIONS

Alice Exnerová, Bibiana Rojas, Olivier Penacchio and Thomas N. Sherratt conceived the idea with input from all of the authors. All authors planned the overall layout of the paper together and contributed to the writing, figures and editing of the manuscript. Alice Exnerová and Olivier Penacchio led the writing of the manuscript. All authors gave final approval for manuscript submission.

ACKNOWLEDGEMENTS

A.E. and B.R. are thankful to Liza Holeski for the invitation to write this review. We thank John Skelhorn, Candy Rowe and Amanda Stefan for fruitful discussions, and Raul Costa-Pereira and three anonymous reviewers for helping us improve the manuscript. J.Y. was supported by UDLA grant 483.A.XIV.24. L.H. was supported by the Academy of Finland (#355869) and the Finnish Cultural Foundation. O.P. was funded by a Maria Zambrano Fellowship—NextGeneration EU (ALRC) for the attraction of international talent for the requalification of the Spanish university system—NextGeneration EU (ALRC). B.R. acknowledges start-up funds from the University of Veterinary Medicine Vienna. T.N.S. is funded by a Natural Sciences and Engineering Research Council of Canada Discovery Grant.

CONFLICT OF INTEREST STATEMENT

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY STATEMENT

This article does not use empirical data.

ORCID

Olivier Penacchio https://orcid.org/0000-0002-1544-2405
Liisa Hämäläinen https://orcid.org/0000-0002-3766-915X
Justin Yeager https://orcid.org/0000-0001-8692-6311
Alice Exnerová https://orcid.org/0000-0001-7937-1477

REFERENCES

Anokhin, P., Sorokin, A., Burtsev, M., & Friston, K. (2024). Associative learning and active inference. *Neural Computation*, *36*, 2602–2635. https://doi.org/10.1162/neco_a_01711

Balogh, A. C., Gamberale-Stille, G., & Leimar, O. (2008). Learning and the mimicry spectrum: From quasi-Bates to super-Müller. *Animal Behaviour*, *76*(5), 1591–1599. https://doi.org/10.1016/j.anbehav.2008.07.017

Barnett, C. A., Bateson, M., & Rowe, C. (2014). Better the devil you know: Avian predators find variation in prey toxicity aversive. *Biology Letters*, *10*, 20140533. https://doi.org/10.1098/rsbl.2014.0533

Barnett, J. B., Cuthill, I. C., & Scott-Samuel, N. E. (2017). Distance-dependent pattern blending can camouflage salient aposematic signals. *Proceedings of the Royal Society B: Biological Sciences*, *284*(1858), 20170128. https://doi.org/10.1098/rspb.2017.0128

Barnett, J. B., Cuthill, I. C., & Scott-Samuel, N. E. (2018). Distance-dependent aposematism and camouflage in the cinnabar moth caterpillar (*Tyria jacobaeae*, Erebidae). *Royal Society Open Science*, *5*(2), 171396. https://doi.org/10.1098/rsos.171396

Barto, A., Mirolli, M., & Baldassarre, G. (2013). Novelty or surprise? *Frontiers in Psychology*, *4*, 907. https://doi.org/10.3389/fpsyg.2013.00907

Bates, H. W. (1862). XXXII. Contributions to an insect fauna of the Amazon Valley. Lepidoptera: Heliconidæ. *Transactions of the Linnean Society of London*, *3*, 495–566.

Bateson, P., & Laland, K. N. (2013). Tinbergen's four questions: An appreciation and an update. *Trends in Ecology & Evolution*, *28*(12), 712–718. https://doi.org/10.1016/j.tree.2013.09.013

Brower, L. P., & Fink, L. S. (1985). A natural toxic defense system: Cardenolides in butterflies versus birds. *Annals of the New York Academy of Sciences*, *443*, 171–188. https://doi.org/10.1111/j.1749-6632.1985.tb27072.x

Caro, T., Sherratt, T. N., & Stevens, M. (2016). The ecology of multiple colour defences. *Evolutionary Ecology*, *30*, 797–809. https://doi.org/10.1007/s10682-016-9854-3

Chouteau, M., Dezeure, J., Sherratt, T. N., Llaurens, V., & Joron, M. (2019). Similar predator aversion for natural prey with diverse toxicity levels. *Animal Behaviour*, *153*, 49–59. https://doi.org/10.1016/j.anbehav.2019.04.017

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

Constant, A., Ramstead, M. J. D., Vessière, S. P. L., Campbell, J. O., & Friston, K. J. (2018). A variational approach to niche construction. *Journal of the Royal Society Interface*, *15*, 20170685. https://doi.org/10.1098/rsif.2017.0685

Cott, H. B. (1940). *Adaptive coloration in animals*. Methuen.

Courville, A. C., Daw, N. D., & Touretzky, D. S. (2006). Bayesian theories of conditioning in a changing world. *Trends in Cognitive Sciences*, *10*, 294–300. https://doi.org/10.1016/j.tics.2006.05.004

Darwin, C. R. (1872). *The expression of the emotions in man and animals*. John Murray.

Davis, M. (1970). Effects of inter-stimulus interval length and variability on startle response habituation in the rat. *Journal of Comparative and Physiological Psychology*, *72*, 177–192. https://doi.org/10.1037/h0029472

De Bona, S., Valkonen, J. K., López-Sepulcre, A., & Mappes, J. (2015). Predator mimicry, not conspicuousness, explains the efficacy of butterfly eyespots. *Proceedings of the Royal Society B: Biological Sciences*, *282*(1806), 20150202. https://doi.org/10.1098/rspb.2015.0202

Drinkwater, E., Allen, W. L., Endler, J. A., Hanlon, R. T., Holmes, G., Homziak, N. T., Kang, C., Leavell, B. C., Lehtonen, J., Loeffler-Henry, K., Ratcliffe, J. M., Rowe, C., Ruxton, G. D., Sherratt, T. N., Skelhorn, J., Skojec, C., Smart, H. R., White, T. E., Yack, J. E., … Umbers, K. D. (2022). A synthesis of deimatic behaviour. *Biological Reviews*, *97*(6), 2237–2267. https://doi.org/10.1111/brv.12891

Ekman, P. E., & Davidson, R. J. (1994). *The nature of emotion: Fundamental questions*. Oxford University Press.

Endler, J. A. (1991). Interactions between predator and prey. In J. R. Krebs & N. Davies (Eds.), *Behavioural ecology* (pp. 169–196). Blackwell Scientific Publications.

Forrester, R. C., & Broom, D. M. (1980). Ongoing behaviour and startle responses of chicks. *Behaviour*, *73*(1/2), 51–63. https://doi.org/10.1163/156853980X00159

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*, 127–183. https://doi.org/10.1038/nrn2787

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., O'Doherty, J., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews*, *68*, 862–879. https://doi.org/10.1016/j.neubiorev.2016.06.022

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

Friston, K., Rigoli, F., Ognibene, D., Mathys, C., Fitzgerald, T., & Pezzulo, G. (2015). Active inference and epistemic value. *Cognitive Neuroscience*, *6*(4), 187–214. https://doi.org/10.1080/17588928.2015.1020053

Gal, S., & Casas, J. (2014). Succession of hide–seek and pursuit–evasion at heterogeneous locations. *Journal of the Royal Society Interface*, *11*, 20140062. https://doi.org/10.1098/rsif.2014.0062

Guilford, T. (1994). "Go-slow" signalling and the problem of automimicry. *Journal of Theoretical Biology*, *170*(3), 311–316. https://doi.org/10.1006/jtbi.1994.1192

Halpin, C. G., Penacchio, O., Lovell, P. G., Cuthill, I. C., Harris, J. M., Skelhorn, J., & Rowe, C. (2020). Pattern contrast influences wariness in naïve predators towards aposematic patterns. *Scientific Reports*, *10*, 9246. https://doi.org/10.1038/s41598-020-65754-y

He, R., Pagani-Núñez, E., Goodale, E., & Barnett, C. R. A. (2022). Avian predators taste reject mimetic prey in relation to their signal reliability. *Scientific Reports*, *12*, 2334. https://doi.org/10.1038/s41598-022-05600-5

Hebets, E. A., & Papaj, D. R. (2004). Complex signal function: Developing a framework of testable hypotheses. *Behavioral Ecology and Sociobiology*, *57*, 197–214. https://doi.org/10.1007/s00265-004-0865-7

Holen, O. H. (2013). Disentangling taste and toxicity in aposematic prey. *Proceedings of the Royal Society B: Biological Sciences*, *280*, 20122588. https://doi.org/10.1098/rspb.2012.2588

Holmes, G. G., Delferriere, E., Rowe, C., Troscianko, J., & Skelhorn, J. (2018). Testing the feasibility of the startle-first route to deimatism. *Scientific Reports*, *8*, 10737. https://doi.org/10.1038/s41598-018-28565-w

Humphreys, R. K., & Ruxton, G. D. (2018). What is known and what is not yet known about deflection of the point of a predator's attack. *Biological Journal of the Linnean Society*, *123*(3), 483–495. https://doi.org/10.1093/biolinnean/blx164

Humphries, D. A., & Driver, P. M. (1970). Protean defence by prey animals. *Oecologia*, *5*, 285–302. https://doi.org/10.1007/BF00815496

Ingalls, V. (1993). Startle and habituation responses of blue jays (*Cyanocitta cristata*) in a laboratory simulation of anti-predator defenses of *Catocala moths* (Lepidoptera: Noctuidae). *Behaviour*, *126*, 77–96. https://doi.org/10.1163/156853993X00353

Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, *49*(10), 1295–1306. https://doi.org/10.1016/j.visres.2008.09.007

Kang, C., Zahiri, R., & Sherratt, T. N. (2017). Body size affects the evolution of hidden colour signals in moths. *Proceedings of the Royal Society B: Biological Sciences*, *284*, 20171287. https://doi.org/10.1098/rspb.2017.1287

Kaye, H., & Pearce, J. M. (1984). The strength of the orienting response during Pavlovian conditioning. *Journal of Experimental Psychology. Animal Behavior Processes*, *10*(1), 90–109. https://doi.org/10.1037/0097-7403.10.1.90

Kikuchi, D. W., Allen, W. L., Arbuckle, K., Aubier, T. G., Briolat, E. S., Burdfield-Steel, E. R., Cheney, K. L., Daňková, K., Elias, M., Hämäläinen, L., Herberstein, M. E., Hossie, T. J., Joron, M., Kunte, K., Leavell, B. C., Lindstedt, C., Lorioux-Chevalier, U., McClure, M., McLellan, C. F., … Exnerová, A. (2023). The evolution and ecology of multiple antipredator defences. *Journal of Evolutionary Biology*, *36*(7), 975–991. https://doi.org/10.1111/jeb.14192

Kim, Y., Hwang, Y., Bae, S., Sherratt, T. N., An, J., Choi, S.-W., Miller, J. C., & Kang, C. (2020). Prey with hidden colour defences benefit from their similarity to aposematic signals. *Proceedings of the Royal Society B: Biological Sciences*, *287*, 20201894. https://doi.org/10.1098/rspb.2020.1894

Kjernsmo, K., Lim, A. M., Middleton, R., Hall, J. R., Costello, L. M., Whitney, H. M., Scott-Samuel, N. E., & Cuthill, I. C. (2022). Beetle iridescence induces an avoidance response in naïve avian predators. *Animal Behaviour*, *188*, 45–50. https://doi.org/10.1016/j.anbehav.2022.04.005

Kjernsmo, K., Whitney, H. M., Scott-Samuel, N. E., Hall, J. R., Knowles, H., Talas, L., & Cuthill, I. C. (2020). Iridescence as camouflage. *Current Biology*, *30*(3), 551–555. https://doi.org/10.1016/j.cub.2019.12.013

Koch, M. (1999). The neurobiology of startle. *Progress in Neurobiology*, *59*(2), 107–128. https://doi.org/10.1016/S0301-0082(98)00098-7

Lindström, L., Alatalo, R. V., & Mappes, J. (1997). Imperfect Batesian mimicry—The effects of the frequency and the distastefulness of the model. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *264*(1379), 149–153. https://doi.org/10.1098/rspb.1997.0022

Loeffler-Henry, K., Kang, C., & Sherratt, T. N. (2021). The anti-predation benefit of flash displays is related to the distance at which the prey initiates its escape. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *288*, 0210866. https://doi.org/10.1098/rspb.2021.0866

Loeffler-Henry, K., Kang, C., & Sherratt, T. N. (2023). Evolutionary transitions from camouflage to aposematism: Hidden signals play a pivotal role. *Science*, *379*(6637), 1136–1140. https://doi.org/10.1126/science.ade5156

Loeffler-Henry, K., Kang, C., Yip, Y., Caro, T., & Sherratt, T. N. (2018). Flash behavior increases prey survival. *Behavioral Ecology*, *29*(3), 528–533. https://doi.org/10.1093/beheco/ary030

Marples, N. M., Roper, T. J., & Harper, D. G. C. (1998). Responses of wild birds to novel prey: Evidence of dietary conservatism. *Oikos*, *83*(1), 161–165. https://doi.org/10.2307/3546557

McNamara, J. M., Green, R. F., & Olsson, O. (2006). Bayes' theorem and its applications in animal behaviour. *Oikos*, *112*, 243–251. https://doi.org/10.1111/j.0030-1299.2006.14228.x

Moore, T. Y., Cooper, K. L., Biewener, A. A., & Vasudevan, R. (2017). Unpredictability of escape trajectory explains predator evasion ability and microhabitat preference of desert rodents. *Nature Communications*, *8*, 440. https://doi.org/10.1038/s41467-017-00373-2

Murali, G. (2018). Now you see me, now you don't: Dynamic flash coloration as an antipredator strategy in motion. *Animal Behaviour*, *142*, 207–220. https://doi.org/10.1016/j.anbehav.2018.06.017

Murali, G., & Kodandaramaiah, U. (2020). Size and unpredictable movement together affect the effectiveness of dynamic flash coloration. *Animal Behaviour*, *162*, 87–93. https://doi.org/10.1016/j.anbehav.2020.02.002

Nahin, P. J. (2007). *Chases and escapes: The mathematics of pursuit and evasion*. Princeton University Press.

Olofsson, M., Jakobsson, S., & Wiklund, C. (2012). Auditory defence in the peacock butterfly (*Inachis io*) against mice (*Apodemus flavicollis* and *A. sylvaticus*). *Behavioral Ecology and Sociobiology*, *66*, 209–215. https://doi.org/10.1007/s00265-011-1268-1

Olofsson, M., Løvlie, H., Tibblin, J., Jakobsson, S., & Wiklund, C. (2013). Eyespot display in the peacock butterfly triggers antipredator behaviors in naïve adult fowl. *Behavioral Ecology*, *24*, 305–310. https://doi.org/10.1093/beheco/ars167

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

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variation in the effectiveness of conditioned but not unconditioned stimuli. *Psychological Review*, *87*, 532–552. https://doi.org/10.1037/0033-295X.87.6.532

Penacchio, O., Halpin, C. G., Cuthill, I. C., Lovell, P. G., Wheelwright, M., Skelhorn, J., Rowe, C., & Harris, J. M. (2024). A computational neuroscience framework for quantifying warning signals. *Methods in Ecology and Evolution*, *15*, 103–116. https://doi.org/10.1111/2041-210X.14268

Ramstead, M. J. D., Badcock, P. B., & Friston, K. J. (2018). Answering Schrödinger's question: A free-energy formulation. *Physics of Life Reviews*, *24*, 1–16. https://doi.org/10.1016/j.plrev.2017.09.001

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*, 79–87. https://doi.org/10.1038/4580

Reisenzein, R., Horstmann, A., & Schutzwohl, A. (2019). The cognitive-evolutionary model of surprise: A review of the evidence. *Topics in Cognitive Science*, *11*(1), 50–74. https://doi.org/10.1111/tops.12292

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: The effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning, 2: Current research and theory* (pp. 64–69). Appleton-Century-Crofts.

Richardson, G., Dickinson, P., Burman, O. H., & Pike, T. W. (2018). Unpredictable movement as an anti-predator strategy. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1885), 20181112. https://doi.org/10.1098/rspb.2018.1112

Rowland, H. M., Mappes, J., Ruxton, G. D., & Speed, M. P. (2010). Mimicry between unequally defended prey can be parasitic: Evidence for quasi-Batesian mimicry. *Ecology Letters*, *13*(12), 1494–1502. https://doi.org/10.1111/j.1461-0248.2010.01539.x

Ruxton, G. D., Allen, W. L., Sherratt, T. N., & Speed, M. P. (2018). *Avoiding attack* (2nd ed.). Oxford University Press.

Schlenoff, D. H. (1985). The startle responses of blue jays to *Catocala* (Lepidoptera: Noctuidae) prey models. *Animal Behaviour*, *33*, 1057–1067. https://doi.org/10.1016/S0003-3472(85)80164-0

Schwartenbeck, P., FitzGerald, T. H., Mathys, C., Dolan, R., & Friston, K. (2015). The dopaminergic midbrain encodes the expected certainty about desired outcomes. *Cerebral Cortex*, *25*(10), 3434–3445. https://doi.org/10.1093/cercor/bhu159

Schwartenbeck, P., & Friston, K. (2016). Computational phenotyping in psychiatry: A worked example. *eNeuro*, *3*(4), ENEURO.0049-16.2016. https://doi.org/10.1523/ENEURO.0049-16.2016

Schwartenbeck, P., Passecker, J., Hauser, T. U., FitzGerald, T. H., Kronbichler, M., & Friston, K. J. (2019). Computational mechanisms of curiosity and goal-directed exploration. *eLife*, *10*(8), e41703. https://doi.org/10.7554/eLife.41703

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, *27*, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Sherratt, T. N. (2003). State-dependent risk-taking by predators in systems with defended prey. *Oikos*, *103*, 93–100. https://doi.org/10.1034/j.1600-0706.2003.12576.x

Sherratt, T. N., Dewan, I., & Skelhorn, J. (2023). The optimal time to approach an unfamiliar object: A Bayesian model. *Behavioral Ecology*, *34*(5), 840–849. https://doi.org/10.1093/beheco/arad032

Sherratt, T. N., & Loeffler-Henry, K. (2022). The adaptive significance of flash behavior: A Bayesian model. *Frontiers in Ecology and Evolution*, *10*, 903769. https://doi.org/10.3389/fevo.2022.903769

Sherratt, T. N., Speed, M. P., & Ruxton, G. D. (2004). Natural selection on unpalatable species imposed by state-dependent foraging behaviour. *Journal of Theoretical Biology*, *228*, 217–226. https://doi.org/10.1016/j.jtbi.2003.12.009

Shettleworth, S. J. (2010). *Cognition, evolution, and behaviour*. Oxford University Press.

Skelhorn, J., Halpin, C. G., & Rowe, C. (2016). Learning about aposematic prey. *Behavioral Ecology*, *27*(4), 955–964. https://doi.org/10.1093/beheco/arw009

Skelhorn, J., Holmes, G. G., & Rowe, C. (2016). Deimatic or aposematic? *Animal Behaviour*, *113*, e1–e3. https://doi.org/10.1016/j.anbehav.2015.07.021

Skelhorn, J., & Rowe, C. (2005). Tasting the difference: Do multiple defence chemicals interact in Müllerian mimicry? *Proceedings of the Royal Society B: Biological Sciences*, *272*, 339–345. https://doi.org/10.1098/rspb.2004.2953

Skelhorn, J., & Rowe, C. (2006). Avian predators taste–reject aposematic prey on the basis of their chemical defence. *Biology Letters*, *2*, 348–350. https://doi.org/10.1098/rsbl.2006.0483

Sokolov, E. N., Spinks, J., Naatanen, R., & Lyytinen, H. (2002). *The orienting response in information processing*. Psychology Press.

Speed, M. P. (1993). Muellerian mimicry and the psychology of predation. *Animal Behaviour*, *45*(3), 571–580. https://doi.org/10.1006/anbe.1993.1067

Speed, M. P., Ruxton, G. D., Mappes, J., & Sherratt, T. N. (2012). Why are defensive toxins so variable? An evolutionary perspective. *Biological Reviews*, *87*(4), 874–884. https://doi.org/10.1111/j.1469-185X.2012.00228.x

Speed, M. P., & Turner, J. R. (1999). Learning and memory in mimicry: II. Do we understand the mimicry spectrum? *Biological Journal of the Linnean Society*, *67*(3), 281–312. https://doi.org/10.1111/j.1095-8312.1999.tb01935.x

Stein, B. E., & Meredith, M. A. (1993). *The merging of the senses*. The MIT Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron. *Nature Reviews Neuroscience*, *9*, 255–266. https://doi.org/10.1038/nrn2331

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An Introduction* (2nd ed.). The MIT Press.

Svennungsen, T. O., & Holen, Ø. H. (2007). The evolutionary stability of automimicry. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1621), 2055–2063. https://doi.org/10.1098/rspb.2007.0456

Tinbergen, N. (1963). On aims and methods of ethology. *Zeitschrift für Tierpsychologie*, *20*, 410–433. https://doi.org/10.1111/j.1439-0310.1963.tb01161.x

Trimmer, P. C., Houston, A. I., Marshall, J. A. R., Bogacz, R., Paul, E. S., Mendl, M. T., & McNamara, J. M. (2008). Mammalian choices: Combining fast-but-inaccurate and slow-but-accurate decision-making systems. *Proceedings of the Royal Society B: Biological Sciences*, *275*, 2353–2361. https://doi.org/10.1098/rspb.2008.0417

Umbers, K. D., & Mappes, J. (2015). Postattack deimatic display in the mountain katydid, *Acripeza reticulata*. *Animal Behaviour*, *100*, 68–73. https://doi.org/10.1016/j.anbehav.2014.11.009

Umbers, K. D., White, T. E., De Bona, S., Haff, T., Ryeland, J., Drinkwater, E., & Mappes, J. (2019). The protective value of a defensive display varies with the experience of wild predators. *Scientific Reports*, *9*(1), 463. https://doi.org/10.1038/s41598-018-36995-9

Vallin, A., Jakobsson, S., Lind, J., & Wiklund, C. (2005). Prey survival by predator intimidation: An experimental study of peacock butterfly defence against blue tits. *Proceedings of the Royal Society*

B: Biological Sciences, *272*(1569), 1203–1207. https://doi.org/10.1098/rspb.2004.3034

Vallin, A., Jakobsson, S., & Wiklund, C. (2007). "An eye for an eye?"—On the generality of the intimidating quality of eyespots in a butterfly and a hawkmoth. *Behavioral Ecology and Sociobiology*, *61*, 1419–1424. https://doi.org/10.1007/s00265-007-0374-6

Wang, L., Ruxton, G. D., Cornell, S. J., Speed, M. P., & Broom, M. (2019). A theory for investment across defences triggered at different stages of a predator-prey encounter. *Journal of Theoretical Biology*, *473*, 9–19. https://doi.org/10.1016/j.jtbi.2019.04.016

Yamawaki, Y. (2011). Defence behaviours of the praying mantis *Tenodera aridifolia* in response to looming objects. *Journal of Insect Physiology*, *57*, 1510–1517. https://doi.org/10.1016/j.jinsphys.2011.08.003

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

Appendix S1. Basics on probabilities.

Appendix S2. Derivation of the computations involving probabilities in the main text.

Appendix S3. A simple Bayesian model of surprise.

Appendix S4. Glossary for surprise-related terms.

Appendix S5. Computation of Bayesian surprise in Figure 5, main text.

Appendix S6. Active inference modelling of avoidance learning: Methods for Figure 6, main text.

Appendix S7. Questions in Box 1, main text and computational quantification of surprise.

Table S1. Surprise-related antipredator defences: overview and examples.

Video S1. The behavioural response of a black-capped chickadee *Poecile atricapillus* to the remote-controlled robotic moth model simulating a deimatic display of the *Catocala* moth (courtesy of Changku Kang).

How to cite this article: Penacchio, O., Hämäläinen, L., Rojas, B., Summers, K., Yeager, J., Sherratt, T. N., & Exnerová, A. (2025). Cognitive ecology of surprise in predator–prey interactions. *Functional Ecology*, *00*, 1–17. https://doi.org/10.1111/1365-2435.14750