Why it hurts: with freedom comes the biological need for pain

Keith D. Farnsworth * [1] and Robert W. Elwood[2]

[1]School of Biological Sciences, Queen's University Belfast, 19 Chlorine Gardens, Belfast BT95DL, UK. orcid=0000-0001-8025-1114

[2]School of Biological Sciences, Queen's University Belfast, 19 Chlorine Gardens, Belfast BT95DL, UK. orcid=0000-0002-6932-8691

Abstract

We argue that pain is not needed to protect the body from damage unless the organism is able to make free choices in action selection. Then pain (including its affective and evaluative aspects) provides a necessary prioritising motivation to select actions expected to avoid it, whilst leaving the possibility of alternative actions to serve potentially higher priorities. Thus, on adaptive grounds, only organisms having free choice over action selection should experience pain. Free choice implies actions must be selected following appraisal of their effects, requiring a predictive model generating estimates of action outcomes. These features give organisms anticipatory behavioural autonomy (ABA) for which we propose a plausible system using an internal predictive model, integrated into a system able to produce the qualitative and affective aspects of pain. Our hypothesis can be tested using behavioural experiments designed to elicit trade-off responses to novel experiences for which algorithmic (automaton) responses might be inappropriate. We discuss the

---
*k.farnsworth@qub.ac.uk

empirical evidence for our hypothesis among taxonomic groups, showing how testing for ABA guides thinking on which groups might experience pain. It is likely that all vertebrates do and plausible that some invertebrates do (decapods, cephalopods and at least some insects).

Acknowledgements

1 Introduction - the three faces of pain

It is still common for pain to be explained as an evolved mechanism for protecting body parts from acute injury, even though it is accepted that reflex withdrawal is often sufficient for that purpose. Our question here is not about immediate responses to nociception; we seek a biological explanation for pain as defined by the International Association for the Study of Pain (IASP): "An unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage" – (Raja et al. 2020). For that we need to define the terms of emotional experience.

1.1 Working definitions

- *Sentience*: Crump et al. (2022) provide an excellent definition: "*Sentience is the capacity to feel. Understood broadly, sentience encompasses all felt experiences, including sensory experiences (e.g. visual, auditory, tactile, olfactory) as well as (for example) feelings of warmth, comfort, fatigue, hunger, thirst, boredom, excitement, distress, anxiety, pain, pleasure and joy. This capacity to feel should be distinguished from other, related capacities: a sentient being might not be able to reflect on its feelings or to understand others' feelings*". Sentience is one of several

2

dimensions of consciousness, though often the words sentience and consciousness are used interchangeably. Sentience is self-evidently needed for pain, but consciousness might not be: this is the crux of contention over whether animals of different kinds can feel pain, so we agree with Browning and Birch (2022), that a clear distinction is essential.

- *Feelings (qualia)*: are subjective (phenomenal) experiences, including pain. We cannot detect feelings by direct empirical study because they are definitively subjective and so bounded by the agent experiencing them: this fact has led to much philosophical debate over whether they even exist (Tye 2021). It is useful to consider them as emergent phenomena generated by (brain) information processing. Recently, Clark et al. (2019) made that idea concrete using the predictive processing theory of perception and consciousness, concluding that qualia are intermediate-level models generated by "Bayesian brains". More generally, we take qualia to be mental constructs that can be functional and, crucially here, motivational (Hall 2008, Fulkerson 2021).

- *Consciousness*: includes several dimensions additional to sentience (Birch et al. 2020b), though sentience is one of its requirements (see Nani et al. 2021). Consciousness is notoriously difficult to identify and study scientifically, partly because it is subjective, but also because we still have no consensus on its definition (Michel 2020). Of greatest significance here are the dimensions of a) self-awareness, derived from a 'meta-perception' system that perceives the perception of internal and external stimuli and b) the integration of perception from internal and external stimuli, along with memory and any available outputs from internal generative models, to form a coherent whole 'mental image'. Pain requires sentience because it is a feeling and it requires integration because it operates at the whole organism level, but pain might not require the other dimensions of consciousness, though

researchers differ over meta-perception: e.g. Key et al. (2021; 2022) consider it the primary requirement for pain experience.

- *Emotion (affect)*: has been implicated in appraisal (Scherer et al. 2001), for action selection (Mendl and Paul 2020) and also direct motivation (Barlassina and Hayward 2019). Helm (2002) defined emotions as "not mere phenomenal states but evaluative responses to one's situation", though recognising that "emotions are feelings" as well. Affect is usually regarded as a top-level (system) phenomenon that sets the internal context for information processing and action selection: an internal psychological milieu (via neurohormones) modulating the parameters of judgement. To that extent emotions are evaluative in *function*. Confusion arises because we know from introspection that emotions have associated feelings (some say they *are* feelings): there is definitely something it is like to be joyful or disgusted, etc.. It is useful here to consider affect as a summarising self-appraisal of an organism's situation as represented by an internal model; one that exists at the level of the integrated whole of the organism (including physiological responses and motor expressions) (Scherer 2022).

1.2 Approaches to pain

Following the pioneering model of Melzack and Casey (1968), pain is broadly recognised to have three dimensions: sensory-discriminative; affective-motivational and cognitive-evaluative (Corns 2014). Pains are feelings with perceptual specificity of location, intensity and quality that are generally noxious, draw attention and motivate those experiencing them to avoid them in future. A simple animal such as the protist *Stentor* can withdraw and guard itself following a noxious stimulus but we doubt it is capable of 'feeling' anything. Similarly, *Cnidarians* are are usually assumed non-sentient. Though capable of sensitisation (an escalating response to a stimulus (e.g. Cheng 2021)), evidence of sentience among them is lacking. By definition (Crump et al. 2022), feelings

4

require sentience to create a phenomenal experience, so only sentient organisms can feel pain.

From philosophy, the 'imperativist' account of pain (Hall 2008, Klein 2007, Martinez 2011; 2015) and the (related) realisation that pain could be interpreted as a part of a homeostatic regulation system for the body, provides valuable context. The imperativist account is that pain is not information about bodily damage or its potential, but rather is a command or motivation for taking action to protect the body from damage (actual or potential). This idea (explained further in the Appendix: *On the philosophical analysis of pain*), has been corroborated by animal studies showing lasting changes in motivation and behaviour following noxious experiences (Sneddon et al. 2014). At the heart of our present thesis is the realisation that this command may be functional only for organisms that have freedom to choose among a range of options for action. That is, only if action-selection mechanisms are not pre-programmed (algorithmic), but rather are the result of the *evaluation* of possible action. Importantly, pain is not required if action selection is strictly reactive with no anticipation of possible futures - in such cases, a rigid relation between perception and action is always sufficient.

For organisms able to anticipate future states, action selection is the result of an internally generated decision based on modelling and evaluating possible future states. The decision is based on the optimisation of some (hedonic) utility currency and is a free choice. We term this process *proactive autonomy* and organisms possessing the freedom it implies would benefit from a strong motivation to attend to injury when evaluating all the options. Conversely, an organism whose action selection is determined by a state-dependent information processing algorithm (if in state S do X), however complicated, gains no advantage from such motivation as pain could provide. The algorithm would instantiate the necessary and sufficient internal information for action to be taken, whether it is to protect against (further) injury, or to continue the current behaviour (e.g. fighting). The information constituting this algorithm could sufficiently

5

be obtained by inheritance and may include sensor and activation threshold shifts in response to repetition of stimulus, enabling habituation or conditioning (simple non-associative learning), without the need for modelling and evaluation. Such an algorithm would facilitate what we call *reactive autonomy* (Fig 1).

An organism that has an internal model predicting accessible future states could in principle initiate behaviours without an external stimulus (i.e. it does not depend on reaction) and could therefore be capable of *anticipatory behavioural autonomy* (ABA - defined in Section 2.2). It is this ability that most readily justifies a cognitive-evaluative dimension to pain, since that would describe the comparison of the expected outcomes of available actions. Pain's role would be to motivate the organism to prioritise attention towards the source of the pain (i.e. salience), but would leave open the possibility of attending to a more pressing matter, such as escape.

In this view, pain is part of the organism's behaviour control system. In general, control is constraint (see Montévil and Mossio (2015)) and all constraint is the result of organising information (Bich et al. 2020, Farnsworth et al. 2013, Farnsworth 2022, Montévil and Mossio 2015, Mossio et al. 2016). This information is not merely the signal[1] of nociception, but crucially includes the causal structure of the cybernetic system: either as an algorithm for reactive autonomy or an internal model and evaluation system for proactive autonomy. The operation of cybernetic systems that determine action selection is entirely one of information processing, i.e. computation, coupled to the physical world by actuators that physically perform the actions. Understanding this information basis for control is important in identifying the autonomy required for ABA (see Sections 2.1 - 2.2).

There are three broad approaches to establishing whether an organism can feel pain. Firstly behavioural responses, especially in highly constrained experimental arrange-

---

[1] We use 'signal' in the standard engineering sense of variation indicating data concerning its source, rather than the special sense of an organism generated sign conveying information, used in behaviour science.

ments, can match our expectation for an organism feeling pain, but since pain is necessarily subjective, this can never provide a definitive answer. Most taking this approach carefully limit their interpretation as: observed behaviours are consistent with expectations for pain (Elwood 2019; 2021). Secondly, we may seek the neural circuitry thought to be necessary for pain (as in Key 2015, Key and Brown 2018, Key et al. 2021). However, we do not yet know what circuits are necessary and rely on either broad categories of processing, e.g. that there must be a subsystem to monitor and create awareness of the internal state of the perception system, or specific hypotheses about parts of the necessary circuits, e.g. that they must include feed-forward and comparator elements (Key et al. 2021). The problem with the former is that it can be too broad, leaving answers unclear. The problem with the latter is that any system proposed as necessary for generating the subjective feeling of pain remains an untested hypothesis until we know what is necessary. The third approach, which has received remarkably little attention, asks which evolved system (or behaviour) needs the subjective feeling of pain in order to work. If we can identify a system that requires pain for its functioning, together with the organisms that possess that system, then we might reasonably presume they will feel pain. This is the (philosophically functionalist) approach we adopt here.

2 Understanding systems that might use pain

Pain requires sentience and it is widely thought that sentience requires an internal model of the self: "*subjective experience arises from [...] an integrated simulation of the state of the animal's own mobile body within the environment*" – (Barron and Klein 2016). This internal model is an essential component of computer representations of animals in welfare research, conceived with widely differing perspectives (e.g. Budaev et al. 2020, Key et al. 2022). Within philosophy, such models are intrinsically implied by representational accounts of pain and are necessary for evaluative accounts beyond the strictly reactive

(i.e. whenever options are to be evaluated for their future consequences). Philosophers rarely refer to models in the technical sense of inferential networks of causal relations, but do assume them as necessary for sentience: classically as the 'inner world', host to the 'Cartesian theatre', or the 'global workspace', or 'multiple drafts' of consciousness (Dennett 1991) and more explicitly in the case of the 'neuromatrix theory' of pain (Melzack 2001). A self-model was conceived by Farnsworth (2017) as part of a mechanistic explanation for free choice in general systems, including organisms and AI systems and a conceptually similar system was proposed by Ridderinkhof (2017). In every mentioned case, the self-model forms part of an allostatic (predictive homeostatic) system that justifies and makes concrete the motivational aspect of felt experiences.

We propose that pain provides for a universal currency (accessible throughout the organism's behaviour control system) for evaluation in anticipatory action selection, one that can command salience and encode information in its qualitative character (as Cabanac (1992) describes in relation to pleasure). This strongly suggests a felt experience, implying sentience, but not necessarily the self-awareness, derived from a 'meta-perception' system (Cunningham 2001), as thought essential by Key and Brown (2018), and argued for by Brown et al. (2021) in response to Birch et al. (2020a). Higher-order-thought theories of consciousness imply that for awareness of pain there must be a subsystem (module) that 'listens in' to the universal signals and reports to a hypothetical executive centre, supposed to be the 'theatre of consciousness'. This idea has been criticised (e.g. Dennett 1991) for falling into the 'homunculus fallacy' (Baltzer-Jaray 2018) and certainly strays from the principle of parsimony that we should adhere to in scientific explanation. We believe that the formation of an internal representation of the self can produce a phenomenal state with intrinsic evaluative character and that this is sufficient to explain the qualitative feeling of e.g. pain. What it is like to be in some degree of pain is the same as what it is like to have a particular self-model result. That is not a model output, since the result is a state of the internal model. In turn, since

8

the model is an integral, globally accessible part of the organism, the whole organism is in that state: a particular phenomenal state we term its Q-state. In this view, pain is a dispositional state of an organism having a predictive model of possible actions and using feelings as the arbiter of choice among them.

2.1 Autonomy and Action Selection

Autonomy is the property of a system undergoing state changes caused by internal events, so that it is at least partly controlled by internalised information rather than entirely by external causes. Action selection is the resolution of conflicts between competing behavioural options. We define *Proactive autonomy* as the ability of an agent to act in the physical world in a way that is determined by the free (non-random) choice of the system. Since there is a choice, there must be at least two viable options and some sort of action-selection system that implements fitness enhancing decision, which in turn implies a system-level utility function to be maximised by the choice. Proactive autonomy implies proximate agent causation: the agent is the causal source of the action. An explanation of how this is possible for living systems is provided in the Appendix: *On Autonomy.* Organisms possessing proactive autonomy display the ability to respond differently to the same external stimulus depending on their independent assessment, enabling appropriate responses to be made to novel circumstances and to take account of future possibilities such as deferred rewards. It is the freedom of choice, enabled by a-priori indeterminacy of outcome, that requires a normative (reward/punishment) evaluation of possible outcomes. The indeterminacy of outcome does not mean that it is random; rather, it is contingent upon some internal computation that is not preprogrammed. Proactive action selection solves an optimisation problem, for which it needs a common currency Y to represent the desirability of each competing behaviour. An arbitrary set of actions can be compared to find which maximises Y given the conditions. Y then acts as an objective function (in the optimality theory sense), the maximisation

9

of which will be the 'goal'. The idea of common currency in this context was pioneered by McFarland et al. (1975), interpreted as biological fitness in the ecological context by McNamara and Houston (1986) and as pleasure by Cabanac (1992) who extended it to an explanation for emotion (Cabanac 2002).

In homeostasis (the most basic form of goal-dependent control shown in Fig. 1.a) the goal is embodied as a set-point. Different perception signals (S1 and S2) can be 'hard wired' to modulate one another to achieve a rudimentary form of action selection Fig. 1.b). If only one action is possible (e.g. in the escape reaction of *Paramecium* (Brette 2021)), then action-selection does not arise, but a homeostatic system comparing multiple perceived signals with their corresponding goals may be used to switch the action ON or OFF using summation, or a winner takes all circuit (Tymoshchuk and Shatnyi 2015) (Fig. 1.c). Single-celled organisms are equipped with these sort of action-selection systems, e.g. for selecting between tumbling and swimming in the chemotaxis (actually kinesis) of *E. coli* (Berg 2004). When there are multiple perceptual signals and multiple possible actions, computation of the most appropriate response rapidly increases in complexity: a problem that could be alleviated using a global modulation signal that integrates the deviations on all the channels (Fig. 1.d). Organisms with small distributed neural networks may implement this sort of control architecture[2] .

_____

[2] *C. elegans* provides a clear example, where modulation and integration were found through molecular-level studies of individual neurons associated with specific behaviours such as chemokinesis, repulsion and aggregation. Cheung et al. (2005) showed the modulation of roaming bahaviour by the aerokinetic (oxygen seeking) motive in *C. elegans*. A suite of similar cross modulation systems and their integration was reviewed by Bargmann (2012), covering *C. elegans* and *Drosophila* neural circuits. In both cases, multiple behavioural motivation systems were found to be extensively cross-modulated by neurohormone control systems. At a higher level of behavioural integration - the 'threat-reward' decision system of *C. elegans* was found by Liu et al. (2020) to be cross-modulated by GABA secretion in reward biased motor neurons, with reception in cholinergic pre-motor neurons that control avoidance behaviour. The effect was that the D-AVA circuit integrates simultaneous attracting and repelling stimuli to produce an outcome that is "*dynamically regulated by the motor system*". This finding corroborates the theory presented by Kaplan et al. (2018) , in reviewing the evidence for inter-neuron integration and modulation of behaviour control (action selection) in *C. elegans*. Rather than segregated feed-forward sensory–to–motor control systems, they suggested that distributed integration of sensory and motor signals, in conjunction with neurohormones, performed computations to generate the observed behaviour (analogous to the computations of an artificial neural network). Further support for this comes from the entirely different approach of dynamic modelling of the complete neural network of *C. elegans* by

Antonopoulos et al. (2016), where the information theoretic measure ϕ, from Integrated Information Theory (Tononi 2008) , revealed significant computation creating new information within the network.
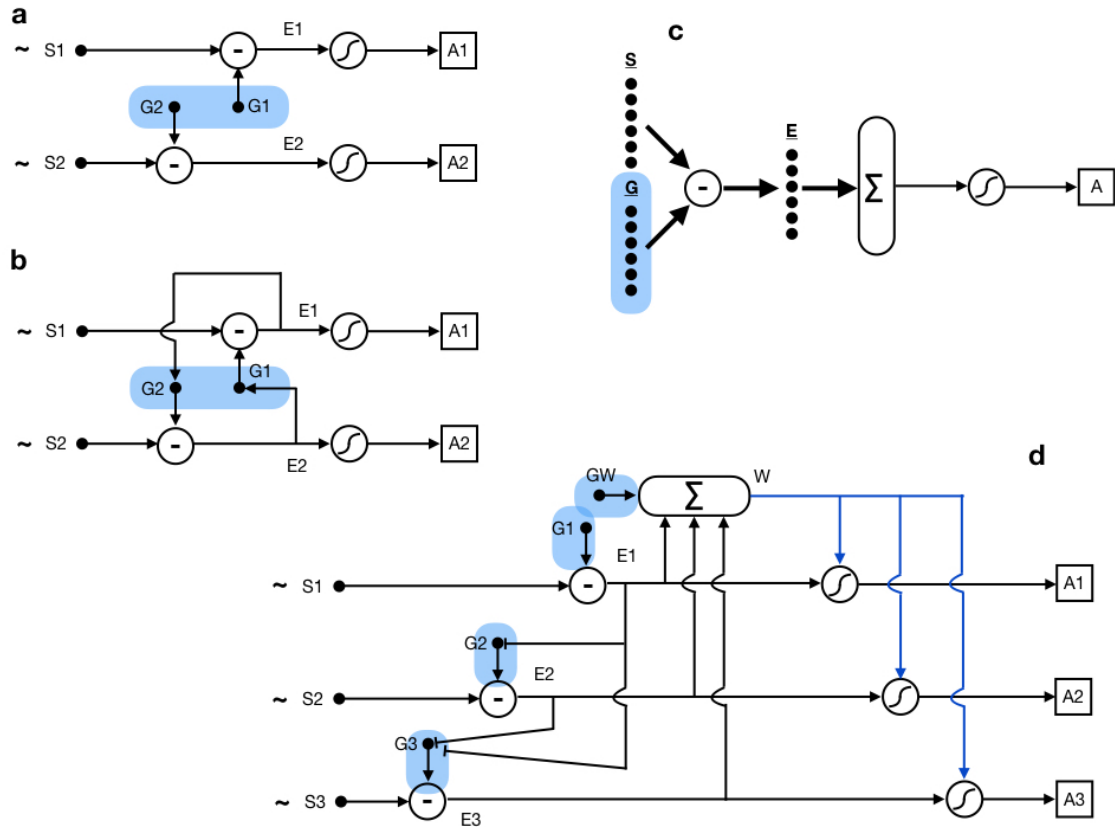
Figure 1: Reactive action-selection systems (autonomous control parts shaded). **a** is the simplest kind with two sensor - actuator channels acting independently. The actuators (A1 and A2) are triggered by exceeding a threshold in error signal (E1 and E2) which is the difference between the input signals (S1 and S2) and the set points (goal G1 and G2). **b** adds cross modulation for resolving conflict between A1 and A2 (e.g. E1 could inhibit A2 by increasing G2). **c** shows multiple sensors and their associated set points combined by summation into a general action (arousal) signal for a single action A (e.g. escape); this could also be implemented through a 'winner takes all' algorithm instead of the summation. In **d**, three sensors add complication, especially in conflict resolution among actions (A1, A2 and A3). In principle, a complicated algorithm could embody a solution to all possible states for this system. In practice it is simpler to solve it by modulating the response thresholds of action signals with a general (arousal) signal W generated by summing the errors (E1, E2 and E3) along with an overall set point for arousal GW. This solution is generalisable to any number of sensors and actuators. In this example, E1 has an inhibitory effect (reduces) G2 and G3 and E2 reduces G3, but several other cross-modulations are possible here. (Blue shading for internally generated (free) signals).

Optimisation of a single (global) currency does not require a set-point since the maximum or minimum are self-evident extrema. Thus in principle, action selection does not need independently stored information (as the set point), but solving the optimisation problem simply by reaction (i.e. in the absence of anticipation as illustrated in Fig. 1.a-d)) would entail repeated cycling through the behavioural options to measure the realised value of the objective function in search of its optimum. That would likely be very inefficient and perhaps also risky. The alternative is to *anticipate* the objective function value for each candidate behaviour and select based on these predictions. This enables the action-selection system to be more general as it can solve the optimality problem in any conditions for which the outcomes of each behaviour can be estimated. There is now convincing evidence that this sort of anticipatory action selection is available to *Drosophila* flies (Barajas-Azpeleta et al. 2021, Cheriyamkunnel et al. 2021, Jiang and Pan 2022), as well as cephalopods (Ponte et al. 2022) and vertebrates such as corvids (Clayton et al. 2003).

2.2 Anticipatory action selection

Anticipatory action selection is action selection in which future states enter the decision-making. It includes allostatic systems (Sterling 2012) and predictive processing schemes, especially active inference (Friston et al. 2013, Pezzulo et al. 2022), which uses a generative model. More mechanistic models representing multiple possible futures have been incorporated in computer simulations of animal behaviour (e.g. Butz and Hoffmann 2002, Matsumoto and Tani 2020, Budaev et al. 2020). Since anticipatory action selection uses unrealised future states in its determination, it necessarily implies proactive autonomy.

We have seen that in reactive autonomy, the response to stimuli is determined by a response-generating algorithm. That algorithm is information embodied as an internal mechanism of the system and is a consequence of evolution and development. In simple (allostatic) systems, anticipation is also built into the control system by an algorithm

13

with pre-programmed information, for example the anticipation of diurnal temperature variation in physiological control (Pezzulo et al. 2022). The algorithm in these cases creates a link between stimulus and response that may be mediated by internal signal processing (e.g. by servomechanisms and internal oscillators (Cheng 2022)), but is causally necessary, i.e. part of a continuous uninterrupted chain. By contrast, an agent capable of *proactive autonomy* responds to a stimulus with an action *chosen* through evaluating the predicted outcome for each available option, using an internally generated goal as a guide (Hoffmann 2003), breaking the causal chain and introducing branching and other options (Ellis and Kopel 2019). The key difference between causally necessary linkage and proactive autonomy is captured by the idea that the former could be analysed using the engineering 'black box' approach to characterising systems by their input-output relations, while the latter produces outputs that cannot be understood from a knowledge of the inputs alone.

Predictions could, in principle, be provided by matching to memories of possible outcomes for every anticipated situation (a sort of database), but that would likely be cumbersome and inflexible. A strong competitive advantage can be gained from the ability to predict a possible future and select the action that maximises an objective function in novel circumstances, especially in an information-rich environment (Butz and Hoffmann 2002). The information system that fulfils the purpose, even for previously unanticipated circumstances, is a model of the self within the environment. It is a transformation (in the mathematical sense) between an input set of stimulus signals and an output that represents the desirability of an outcome (hedonic valuation). The transformation depends on both the action under evaluation and the state of the agent following the action, given the sensory inputs. Feed-forward models (systems that predict afferent signals, given the current efferent signals) are typically used to perform the transformation in anticipatory control systems (Fig. 2.A). Artificial neural networks are often used for the computation in engineering (e.g. Matsumoto and Tani 2020) and

14

neural networks are known to implement it for the motor control of organisms (e.g. Jékely et al. 2021). This is extended to action selection by implementing a forward model for each potential action, predicting its outcome prior to realisation. Outcomes are generalised by a hedonic signal to be optimised for action selection. This signal may in practice be a neurohormone encoding valence information, which can then be used to select actions, e.g. by controlling the thresholds for actions to be realised Fig. 2.B). Since the information for appraisal results from training (by reinforcement learning) of the forward model, it is internal and inherent to the control system (i.e. the organism) and to that extent free from exogenous control. It could function as a distress signal, but does not fulfil all the requirements for pain itself.
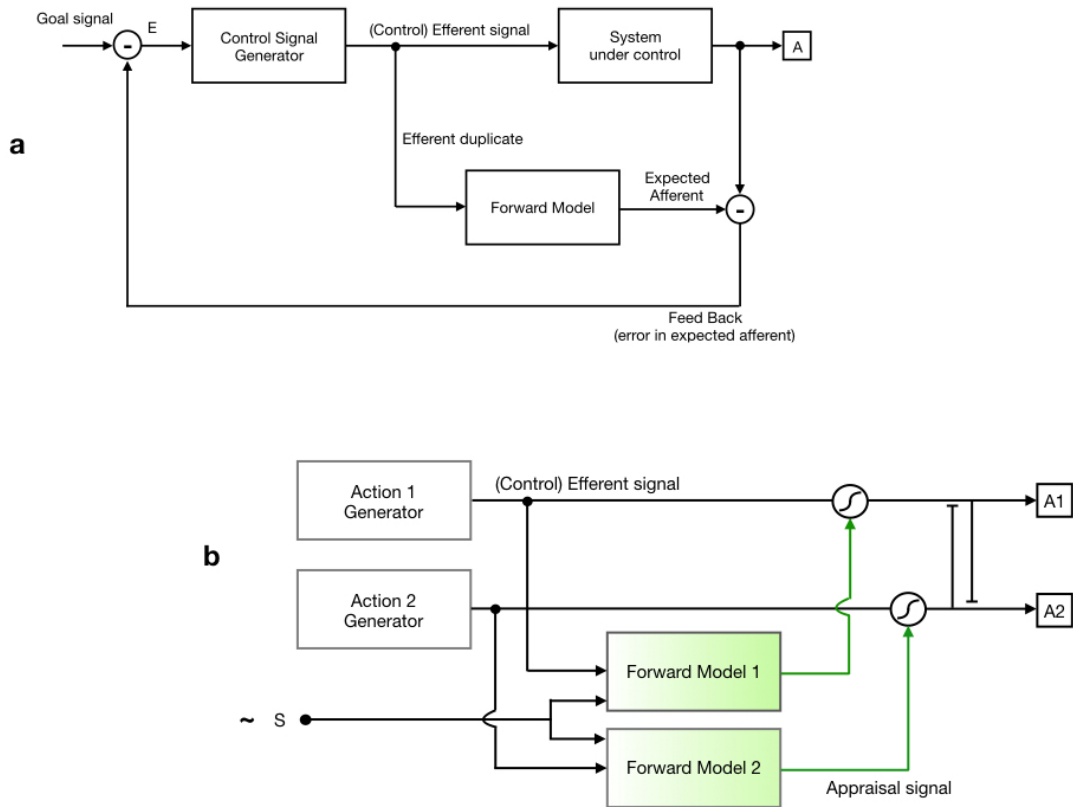
Figure 2: Control systems using forward models to predict the afferent result of actions. In **a**, the forward model continuously predicts the effect of the current control signals to enable feedback through which they are refined. This is a standard method for refining motor control. In **b**, this is adapted to predict a normative (hedonic) summary of the effect of each potential action (A1, A2) for use in action selection. Each forward model uses the corresponding efferent signal together with environmental perception, (S) for context, to generate an *appraisal signal* which may be implemented as a neurohormone level (indicated with green shading). This signal modulates the thresholds for enacting A1 and A2 (and the threshold gated action signals mutually inhibit to prevent indecision). The potential actions are realised as control signals generated by internal pre-programmed routines, but the forward models are trained by conditioning (reinforcement learning) to produce appropriate hedonic appraisal signals which therefore are internal (free) signals.

Forward models that can learn to generate a hedonic signal from potential actions, given a perceived context, can be implemented by recurrent neural networks with hor-

16

mone secreting output neurons. This system is limited to a finite set of 'rote' behaviours with fixed threshold control circuitry implementing action selection, but it can be generalised. Farnsworth (2017) proposed a more general system for which the internal model might have to be a universal Turing machine, though this is not a particularly onerous condition in practice (Graves et al. 2014). Alternatively, in the active inference approach, the idea of selecting an optimal action is replaced by finding optimal inferences (Bayesian beliefs) about likely future behaviours and their consequences (Friston et al. 2013). That uses an internal generative model, the states of which become probabilistic representations of external states (the physical world including the self). A hidden Markov model is a natural fit for it, readily implemented by a neural network. Active inference incorporates the goal as a minimisation of the divergence between the probability distribution of attainable states and states believed to confer high utility (Friston et al. 2013). It accounts for motivation (Clark 2020, Tate 2021), though not initiative (Klein 2018).

Initiative (the ability to change or initiate a behaviour independent of external stimulus or 'pre-programmed control') is the main emergent property of proactive autonomy derived from an internal model. We call it *anticipatory behavioural autonomy* (ABA) to emphasise the autonomy of decision-making and rational intention of the initiative – a point made by Hoffmann (2003) who termed the phenomenon "anticipatory behavioral control". Proactive autonomy, based on evaluation of outcomes predicted by an internal model, enables *behavioural autonomy* (Schneider 2018). If the evaluation uses a common currency to represent the desirability of outcomes (e.g. a valence-informing hormone signal), then an arbitrary set of behaviours and outcomes can be compared in that common currency. With hormones, accessible to the whole organism's behavioural control system, outcomes from different behaviours can be integrated to find the solution to trade-offs such as between feeding and threat or reproductive opportunities. Dopamine (generally involved in reward), serotonin (mediating anxiety) and cortisol or hyperglycemic hor-

17

mone (arousal) are plausible candidates for this. ABA, then, is behavioural autonomy in which the choice is based on the organism's prediction of a global hedonic value under each of the available options. Because the decision is based on the anticipated value, rather than following prescribed rules (an algorithm), the organism's response is not entirely predictable from knowledge of the stimulus alone. Low predictability of behaviour, especially in novel circumstances, could therefore be an empirical indicator of ABA and by consequence, of the usefulness of pain.

In parallel, recent conceptual work, backed by empirical evidence, shows that subjective experience also requires an internalised model of the self that is anticipatory and involves a comparator between expectations and incoming sensory signals (Key and Brown 2018, Key et al. 2021; 2022). Key and coworkers contend that conscious awareness is a necessary component of pain (also see Adamo 2019). But, using optimal control theory applied to both natural and artificial systems, Schneider (2018) demonstrated that proactive autonomy is effectively implemented by a combination of an internal model and monitoring of internal states, from which emotion-like phenomena spontaneously emerge. In his analysis, the internal states being monitored give rise to an integrating affect-like signal that provides a "fast and frugal" heuristic for appraisal in action selection. That is effectively equivalent to the process enabling ABA.

2.3 A hypothetical model implementation

The "free-will machine" from Farnsworth (2017), taken as a hypothetical ABA generating system, can be implemented by a neural-hormonal control system that is consistent with the concept of pain. In Fig. 3, **S** represents perception inputs (signals from transducers, including nociceptors). They are compared to a model of expected inputs (**M**) by the comparator (**-**), this model is updated by e.g. Bayesian inference, and the modelled signals are compared to internally set goals (**G**) for the signals (desired or expected states). The difference between **M** and **G** on each channel (**E**) informs self-modelling

18

about the current state. The result is the formation of a self-model that emerges in a particular Q-state. This Q-state may be interpreted as the informational embodiment of a quale. The model can generate as many different qualia as it has states: a number that increases rapidly with the number of neurons instantiating the model. Note the internal model does not have outputs *per se*, just its Q-states. The self-model has access to memories of Q-states, which it seeks to match. It is also multiply connected with the rest of the body (soma), crucially including hormonal releasers and receptors. The integration of the self-model with the somatic system raises the Q-state to a state of affect: an emotional feeling (short term) or a mood (long term). This emotionally charged state of the combined model (neurons) and hormone system then modulates the drive to perform a finite set of actions (just two illustrated: A1,A2), each generated from a pre-programmed routine (R1, R2). The neurons that produce the routines are connected with the self-model such that the self-model modulates their thresholds for action. For example, a particular Q-state may down-regulate the threshold for A1 and up-regulate the threshold for A2, with the result that A1 is performed. Note that attention (salience) emerges from the somatic-self-model system as the hormones create the strength of the feeling of being in Q (that feeling being the quale). Thus, for example, if **S** is carrying substantial nociception, **E** will be large and the self-model will emerge in a pain Q-state, which will strongly stimulate hormones associated with being in pain and a state of suffering (emotional pain) will ensue, which will strongly down-regulate the thresholds for escape, guarding and other pain-related behaviours (turning them on), while simultaneously up-regulating the thresholds for all other behaviours, effectively stopping them.

This may seem complicated, but it is not unduly demanding of number and inter-connectivity of neurons. Greve et al. (2016) showed an artificial neural Turing machine could learn to solve a double T maze using just 70 nodes (artificial neurons), providing more than 10^{21} possible states. By contrast, drosophila has $\sim$100k neurons (Scheffer

and Meinertzhagen 2019), each typically with $\sim$100 synapses (estimate total of $2\mathrm{x}10^7$ synapses (Scheffer et al. 2020)), so even if only 1% of neurons implemented Q-states, there could be 10^{300} of them. The figures are beyond 'astronomical' for vertebrates.
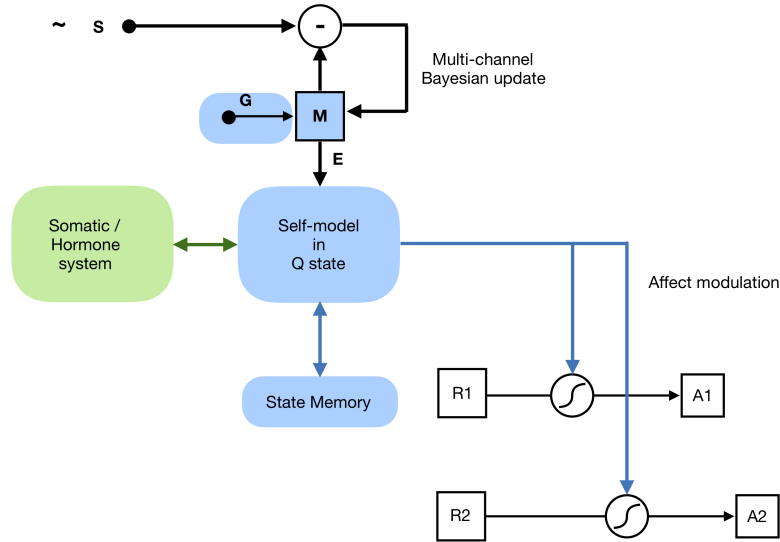
Figure 3: A hypothetical affect-driven action-selection system giving anticipatory behavioural autonomy. Bold symbols and lines represent vector (muti-channel) signals. **S** represents perception inputs (signals from transducers, including nociceptors). They are compared to a model of expected inputs (**M**) by the comparator (**-**), this model is updated by Bayesian inference, and the modelled signals are compared to internally set goals **G** for the signals (desired or expected states). The difference between **M** and **G** on each channel (**E**) informs self-modelling about the current state: the self-model emerges in a particular Q-state. It has access to memories (either experienced or pre-programmed) of Q-states, which it seeks to match. It is also multiply connected with somatic hormonal releasers and receptors. Integration of the self-model with the somatic system raises the Q-state to a state of affect resulting in an emotionally charged signal which modulates the drive to perform a finite set of actions (A1, A2), each generated from a pre-programmed routine (R1, R2). Modulation is achieved via action threshold modification. Further details in the text. (Blue shading for internally generated (free) signals on neurons, green for hormonal signals).

2.4 Anticipation, alone, does not require pain

Predictive processing is successful in explaining elementary cognition-response systems (e.g. Pezzulo et al. 2022). More generally, anticipatory action selection could, with relatively small systems, be implemented by a finite state automaton (FSA) leaving the organism absent of free choice and therefore not requiring pain. Examples of FSA-based anticipatory control systems typically depend on feed-forward models, as in Key et al. (2021), or internal models implementing active inference, as in Matsumoto and Tani (2020). Here we see the need for at least a memory, which may be elaborated into a model of the self, but having action selection still produced through the FSA architecture. Even if a global modulating signal (e.g. a neurochemical mediated state variable) were introduced to add nuance to the action selection, it could be implemented without recourse to agent freedom, still in principle leaving pain unnecessary.

This is roughly the conception of Key and Brown (2018), who developed a hierarchical predictive system in their search for the minimum system necessary to generate subjective experience. It consists of a nested pair of feed-forward predictive models, the inner model predicting the response to stimulus, the outer predicting the difference between this prediction and the realised response, given both signals together with 'global input' from other 'brain' areas. Key et al. (2021) argue that animals lacking a recognisably equivalent neural processing system would be incapable of the subjective experience that is pain. Since Key et al.'s (2021) two-level feed-forward model is only one of several plausible systems, that is a strong claim. One primary requirement, they claim, is that the higher-level prediction (or its error signal) is shared (broadcast) with the global system, for they say that the '3rd order awareness' generated by their system is only sufficient for 'pre-conscious' awareness and it is the global availability of its output that produces conscious awareness, though they do not explain how or why. The whole system they propose remains reactive, since it does not incorporate any goal or desire and also has no action selection component (it was not intended for that purpose). We therefore need to

22

add goal-seeking to obtain a model of experience-driven autonomy. Key et al. (2021) are keen to distinguish their model from other predictive processing schemes, principally on the grounds that their predictive models are not "*embedded within* the internal sensory processing stream", claiming that subjective experience cannot be supported without that separation of computational tasks. But computationally, it makes no difference whether the nested predictors are depicted as within the stimulus-reaction processing system or as a separate module sharing signals with it. What really makes the difference is the sharing of prediction signals with the global system. Though Key et al. (2021) recognise that necessity, they say nothing about what the global system has to do with these shared signals in order to generate subjective experience and we are in danger of entering an infinite regress in search of the elusive process that generates subjective experiences out of all these signals. Key et al. (2021) appear to imply that subjective experience is just 'what it is like' to have global availability of certain signals, or more generally what it is like to be in some particular states. We strongly agree to that – no mysterious, even metaphysical, experience-generating processor needs to be added to a brain in order to generate subjective experience; it is simply what it is like to be in a particular state (what we termed the Q-state). If we are to avoid the homunculus fallacy, we must concede this point at some stage.

In summary, subjective experience has an evaluative component – a normative character – arising from the difference between the current state and that sought by a goal-directed action selection system. It is therefore part of an anticipatory autonomy system. Pain feels bad because it is a state that is far from that desired and it motivates action in response. What it feels like to be in pain is the awareness of the gulf between a current state and the comfortable (homeostatic) state constantly sought. Thus anticipation is necessary for pain, but only jointly with goals and a global-level evaluation.

23

2.5 Autonomous evaluation necessitates pain

The key difference between an automaton system and a free autonomous agent is that in the latter actions are selected based on their evaluation in a common currency. It is the independent evaluation, a computational process isolated from perception-response, that provides the freedom of free autonomy. Evaluation is subjective and context-dependent; it cannot be replaced with a FSA algorithm, not only because it entails an indeterminate number of states, but because it is necessarily a faculty of the whole organism: the only level to which we may accord the status of freedom. An organism that is a free autonomous agent is able to make free choices because, as an integrated whole, it is the embodiment of the information entailed in making the choices (this is an important matter of attribution). These choices are not determined by exogenous causes, nor by immutable internal causal structure (an inbuilt algorithm) in any component part of the organism; instead they are determined by the goal-seeking intention of the unified whole of the organism. This optimisation can be termed the 'will' of the organism only because the goal is instantiated at the highest level of causal organisation (Farnsworth 2018; 2017). The goal is the maximum of a global utility function which, by natural selection, should coincide with Darwinian fitness, but for the individual organism it may be represented by an effect-like signal on the pleasure/ pain axis as described by Hoffmann (2003) and (Schneider 2018) (noting this may be a simplification since pain and pleasure are thought to be separate systems (Pietri et al. 2013)).

Evaluation requires a universal currency to compare the value of each option regardless of its nature, similar to the economists' notion of 'utility', which enables comparison of cinema tickets with cheese. This universal currency needs the properties of valence (good/badness) and intensity (activation or arousal level). These are provided by the 'emotional space' defined by Russell (1978) and elaborated in Russell and Barrett (1999). This idea of evaluation on valence and arrousal axes of a universal currency is compatible with the cognitive appraisal theories reviewed in Scherer et al. (2001). More recently, the

evaluation component of emotion has gained broad acceptance within (human) emotion theory, brought together under an inclusive definition by Scherer (2022):

... "emotions 1) consist of an episodic process in response to a perceived event or situation of major significance, 2) which is characterised by recursive causal effects (forward and backwards) between several components that include the evaluation of the event in terms of its significance for the goals and values of the individual, 3) creating physiological reactions, motor expressions, and action tendencies and 4) that this process is partially accessible to consciousness, resulting in feelings that 5) can be categorised and subsequently labelled by the individual in terms of its subjective conceptual structure".

The two-dimensional circumplex model of affect (Russell 1978, Posner et al. 2005) is the antithesis of the so-called 'basic emotion' model in which emotions are discrete separate sensations. The circumplex model has gained considerable empirical support and the idea that a wide range of emotions can be constructed from just two axes of latent variation (valence and arousal) is commonly invoked in human psychology, though contested (Ortony 2022). The practical (fitness) value of simple emotions has been shown using reinforcement learning in artificial intelligence systems, which can be enhanced by incorporating simulated emotions into action selection (Sequeira et al. 2015). This integrates current perception signals with memories and model expectations to produce an autonomous self-centred decision-making process. Affective signals are used as an overall hedonic objective function to be maximised, using both current and anticipated states for possible actions in the context of action selection. A typical arrangement involves a joy vs. distress axis, which may be further enhanced with a hope vs. fear axis, identified as the anticipation of joy or distress, respectively (Broekens et al. 2015).

At least for the present purpose, the single dimension of a valence is very suitable for the simplest motivational signal: the contrast between pleasure and pain, with their

associated general response of approach and withdrawal. In short, a single general signal of valence is enough to provide for the affect aspect of pain (and pleasure) and to function as a universal comparator of actions available for selection suggested by Hoffmann (2003). Further, we cannot ignore the obvious parallel between a universal signal of valence and the widely acting neurohormones, which are taken by many as an objective indicator of psychological stress or arousal, and in animal studies as a surrogate for pain: cortisol for vertebrates (Cerqueira et al. 2021, Stafford and Mellor 2005, Wagner 2010) and hyperglycaemic hormone for invertebrates, such as crustaceans (Chang 2005, Elwood and Adams 2015). Autonomous evaluation can be implemented using such hormones as a universal and integrating motivational quantity that is continuously variable and gives effect to the state of the internal model: in particular the feeling of pain.

2.6 Origins of pain and primary consciousness

Our proposal suggests that pain is only adaptive for those animals able to make autonomous anticipatory decisions. If pain is considered a part of primary consciousness, then it cannot precede the evolutionary origins of consciousness for which Feinberg and Mallatt (2013) and Ginsburg and Jablonka (2019) provide complementary accounts. But evidence for primary consciousness is also very hard to obtain; e.g. Suzuki (2021) found insufficient evidence to determine if the consciousness criteria provided by Feinberg and Mallatt (2013) and Ginsburg and Jablonka (2019) applied to the first vertebrates, based on studies of extant cyclostomes (jawless fish). According to our present hypothesis, pain is contingent upon free decision-making, the hallmarks of which, listed above, may be more accessible to empirical testing. Of these the most diagnostic may prove to be state-dependent trade-offs under noxious stimulus. As we noted earlier, such trade-offs could in principle be brought about by a very complicated automaton system, but more simply and flexibly produced by a proactive affect-driven autonomous decision system (consistent with ABA). The requirements for this coincide with those for "unlimited

associative learning" (UAL), which was identified as the transition marker for the evolution of consciousness in Ginsburg and Jablonka's (2019) account. As (Birch et al. 2020a) indicates: the requirements for UAL, in turn, coincide with those for sentience: 1) global accessibility and broadcast; 2) binding/unification; 3) selective attention and exclusion; 4) intentionality; 5) integration over time; 6) an evaluative system; 7) agency and embodiment and 8) registration of self/other distinction. The self-model required for ABA enables self/other distinction and provides for integration over time. The agency required for ABA is provided by an evaluation system that is not merely an automaton implementation (this point being emphasised by Ginsburg and Jablonka 2019). It also requires the integration of the model with evaluation system and our suggestion of hormonal mediation fulfils that role as well as fulfilling global accessibility, broadcast and binding/unification requirements and embodiment. Thus ABA also shares the requirements for sentience. This means that state-dependent trade-offs under noxious stimulus could provide a valuable surrogate in the search for pain, and more general sentience, in extant organisms.

3 Empirical support for ABA implying the need for pain

The function of pain, distinct from nociception, is identified by Sneddon (2009) as enabling an organism to "quickly learn to avoid the noxious stimulus and demonstrate sustained changes in behaviour that have a protective function to reduce further injury and pain, prevent the injury from recurring, and promote healing and recovery". In other words, pain should elicit persistent changes of behaviour through modulation of action selection. For example, we see conditioned place avoidance for areas associated with noxious stimuli in shore crabs (Magee and Elwood 2013) and octopuses (Crook 2021). Further, octopuses that could not avoid noxious stimuli (acetic acid injection) preferred areas associated with a local anaesthetic. Other long-term changes in behaviour observed

after noxious stimuli include alterations of shell preference in hermit crabs, which last at least 24 hours following electric shock (Appel and Elwood 2009, Elwood and Appel 2009) and the onset of anxiety-like states in crayfish after shock (Fossat et al. 2014), which are also seen in fish (de Abreu et al. 2020) and amphibians (Brown et al. 2013). Anxiety-like states are usually associated with serotonin (Best et al. 2020, Curran and Chalasani 2012), but do not alone imply pain; e.g. a simple (algorithmic) mechanism for their manifestation, requiring only two neurons, has been found in *C. elegans* (Eliezer et al. 2019), but without evidence of evaluation. Anxiety-like states do demonstrate anticipation, and generalised modulation of action selection, but could, in principle, be generated by an automaton. So though the behavioural observations above are consistent with expectations of pain in a wide range of species (Sneddon et al. 2014), they do not conclusively support our hypothesis concerning autonomous and anticipatory behaviour. That hypothesis broadly suggests that pain is only useful, and hence likely to be present, in animals that can make a free choice between available responses in the presence of a noxious stimulus. To test this, we would need evidence of 1) mental models of the self and the environment to support anticipation; 2) flexibility in behavioural responses to stimuli (showing that options are available); 3) proactive choice and forward planning (actions based on anticipated consequences, rather than just the current state) and 4) free choice of response to noxious stimuli (not algorithmically pre-programmed) that is rational rather than random (shown by e.g. state-dependent trade-offs).

3.1 Models of self and the environment

Models of self presumably developed early in evolution with examples emerging in a wide range of multicellular animals (Jékely et al. 2021). The most basic of these models involve reafference, which is the term given to the ability of an animal to discriminate between sensory changes due to self-movement and those due to environment change (Jékely et al. 2021). This ability is important because the two are likely to have very

28

different meanings for the animal. For example, an object taking up more space on the retina (or compound eye), i.e., a looming stimulus (Temizer et al. 2015), could be due to the animal approaching the object or the object approaching the animal. In the first case there is little risk whereas in the latter case, looming might indicate danger to the animal. But reafference also applies to a wide range of stimuli such as flow of water over the body surface due to own movement contrasted with that due to environmental flow, or to deformation of the body due to own movement or to some external force. That is, the animal has a model of self, and largely disregards inputs due to self-movement, whereas those due to external changes receive attention.

Animals also form models about the environment, demonstrated by the classic experiment in which chicks anticipated the timing of a light being switched on and off at regular intervals, showing startle responses when it turned off early or late (Broom 1968). Anticipatory modelling is of course central to predictive processing and active inference theories and the generation of associative learning. When animals learn about associations between two environmental changes, as in classical conditioning, or between an action and subsequent environmental event, as in instrumental conditioning, they form mental models that allow distinction between chance coincidence and true causal relations between neutral events and subsequent events of biological significance (Dickinson 1980). This process can be complex, even in invertebrates such as insects (reviewed by Perry and Barron 2013). Numerous groups of cells and circuits, and their interactions, have been identified as involved in learning about rewards and punishments. Further, the roles of octopamine and dopamine in reward and punishment learning have been discovered, as have more complex interplay between these two control circuits. Of particular interest, however, is the suggestion that with rewards there are subjective feelings of "liking" as separate from "wanting", and again separate but interacting circuits seem to be involved (Perry and Barron 2013). That is, the hedonic value of an unconditioned stimulus plays a role in the learning about a conditioned stimulus, such as an odour,

29

and how it predicts the arrival of the unconditioned stimulus, such as sucrose or a sweet substitute. Hedonic value may also play a part in avoidance of punishment, such as electric shock. The conclusion from these studies, and many others, is that animals gather information about themselves and about the environment and use these for building models that show expectation of future events so that behaviour may be modified to better gain rewards and avoid risks. These models of self and environment are key to the success of metazoans.

3.2 Flexibility of responses to stimuli

Identifying flexibility of response, (following the classical definition of free will: "able to do otherwise") depends on there being available response options and a demonstration of more than one response to the same stimulus from the same individual. The first criterion can be established within a species by observing inter-individual differences in response, because such differences would result from differences of internal state among individuals, i.e. a complex of genetic, developmental and accumulated experiences (Stamps 2016). Appel and Elwood (2009) demonstrated this with hermit crabs undergoing a standardised noxious stimulus (i.e. with minimum variation in magnitude and site of application). Crabs were induced to occupy empty gastropod shells wired to apply electric shocks to the abdomen of the crab within its shell. Of the 123 crabs that received a standardised shock treatment 61 evacuated the shell and 29 of those groomed and tended to their abdomen at the site of the shock application. Four crabs attempted to climb the wall of the observation chamber and three engaged in shell-rapping, an activity normally seen in fights for ownership of shells. After evacuation, 57 crabs re-entered the shell, leaving four that stayed away from it. None of these activities were observed in unshocked controls. Evidently the observed behaviours show a variety of individual responses to the same noxious stimulus, demonstrating options for action selection.

To identify flexibility within the individual (endogenous placticity), we must first distinguish between sources of variation in their response. We reject random as it negates autonomy. Developmental shifts (maturation and e.g. role differentiation in social insects) do not indicate coincident options for the animal (discussed by Jeanson 2019). Changes in response to a change of the environment might be generated by an automaton algorithm, so not free. For example, (Czaczkes et al. 2018) observed task switching between exploration and exploitation in forager ants in a T-maze with sucrose rewards at the end of each arm. Following a period of training in which ants learned to associate reward levels with various cues, ants were free to choose either arm over multiple trials. They showed little switching between arms, irrespective of their reward levels as long as rewards remained constant. When reward levels, along with associated cues, in both arms were simultaneously increased, or decreased, then switching rate also increased between trials, showing a change from exploitation to exploration behaviour. If ants have an exploration algorithm, but otherwise default to exploitation, then a simple threshold switch, sensitive to reward change (Wilson 1976), would suffice to produce this apparent behavioural flexibility. Conversely, within-individual changes of behaviour, without a change in environmental stimulus, may result from learning, in which case internalised information gained by the organism, not inbuilt, is the source of change; hence free-choice flexibility is demonstrated. Jeanson (2019) discussed learning-dependent flexibility in social insects, e.g. that in ants able to perform multiple tasks, a successful foraging experience can increase the likelihood of repeated foraging (Ravary et al. 2007). Representing cephalopods, Chung et al. (2022) showed that cuttlefish changed their response to ambiguous prey choice following the experience of receiving an unexpected food reward. They interpreted this change as foraging strategy selection mediated by an internal state they identified as an emotion-like state. Magurran (1993) reviewed a substantial body of evidence of context-dependent behaviours within teleosts, not least the ability of male guppies to choose between overt display for a mate, or "sneaky" mat-

ing tactics. Most supportive of ABA, Earley et al. (2013) found that mangrove killifish (*Kryptolebias marmoratus*) 'perceive' their own fighting ability (implying a self-model) and they "adjust contest strategy" when that perception is updated following wins or losses. The authors identified this behavioural flexibility with changes in three hormones, concluding it is "modulated by internal state".

3.3 Proactive choice and forward planning

Examples of forward planning in action selection are available among invertebrates (Elwood 2022). One such study used terrestrial hermit crabs and allowed them to walk along a corridor with obstacles that partially blocked the passage of the crab's shell (Sonoda et al. 2012). The obstacles varied in the degree to which they made passage difficult. However, crabs were proactive in this task, turning their shells to avoid the obstacle before encountering it. Further, they turned the shell to a greater degree if the space between the shell and obstacle was narrow. When crabs had plastic plates attached to the shells making the obstacle course more difficult, they turned the shell to a greater degree on their first attempt, enabling pasaage without the plastic extension colliding with obstacles. The behaviour (degree of turning) was selected prior to collision experience and in response to a novel situation, thus showing proactive selection of behaviour for optimal outcome. Other examples with hermit crabs showing proactive flexible responses are discussed in Elwood (2022). Ants following cues that predict a valuable reward show more pheromone marking of the trail than do those following cues to a weaker reward, which again suggests forward planning (Czaczkes et al. 2018). Spiders that live in a complex 3-dimensional environment can detect prey from a distance but reaching the prey might not be achieved in a straight line. Spiders have been seen to plan the route and on occasions might move further away from the prey in order to get to a branch that will then lead them closer (Tarsitano 2006). All together, action choice with forward planning has been documented for arthropods. Among vertebrates, these

faculties are well known, for example through reversal learning experiments, especially with birds (e.g. Bond et al. 2007).

3.4 Non-algorithmic, selection of response to a noxious stimulus

Probably the best evidence for proactive choice in response to a noxious stimulus comes from examples of trade-offs between avoidance of a noxious stimulus and any other goal (e.g. Balasko and Cabanac 1998). Sneddon (2019) reviewed evidence of this in fish (and cephalopod) species. Millsopp and Laming (2008) found that goldfish (*Carassius auratus*), which they trained to feed in one region of an experimental aquarium, and where they were subsequently subjected to electric shocks, would spend more time in this feeding/shock zone the more food-deprived they were, and that this trade-off shifted away from feeding attempts towards escape as the shock intensity was increased. Crook et al. (2014) showed direct fitness benefits for squid (*Doryteuthis pealeii*) as they put extra effort into escape from predator cues when they were experimentally injured, leading to an almost doubling of survival rate compared to those that had been anaesthetised during the injury process. Hermit crabs evacuate their shell after an electric shock with a probability that depends on the quality of the shell (Elwood and Appel 2009) and also the presence of a predator odour Magee and Elwood (2016). Thus, these crabs displayed a flexible trade-off when responding to a noxious stimulus with respect to keeping a high-quality shell and avoidance of predation. Further, a recent study on bumblebees demonstrated a trade-off between avoiding a high temperature and obtaining a high-quality food source, with the bees using learned colour cues for their decisions, indicating both flexible responses and associative learning based on contextual information (Gibbons et al. 2022b). These demonstrations of trade-offs suggest proactive choice following noxious stimuli. It is doubtful if they could result from an inbuilt algorithm because of the complexity required and because they seem to occur in novel situations. Further, some authors have put considerable weight on trade-offs as a key criterion of

33

pain (Crump et al. 2022). Intriguingly, Puri and Faulkes (2015), who examined responses of crayfish to a heat stimulus from a soldering iron that touched the animal, showed (in videos published with their paper) that some animals grabbed the shaft of the soldering iron in response. Evidently the animal was selecting a response that was dependent upon available information. This unexpected protective response replaced the reflex withdrawal, seen in some individuals, with a co-ordinated attack that seems to use freedom of action-selection to manifest.

3.5 Evolution of pain experience

So far, there is evidence for pain in three major phyla, the chordates (Sneddon et al. 2003), molluscs (Cooke 2021) and arthropods (Elwood 2019). These three phyla arose during the Cambrian explosion and the most recent common ancestor for these is likely to be a free-living worm-like organism from about 530-550 million years ago (Elwood 2011). The parsimonious explanation for the evolution of pain in the three phyla is that there was one evolutionary step that occurred in or before the most recent common ancestor. Against this, evidence for pain is restricted to specific groups, such as the cephalopods within molluscs, the decapod crustaceans and some insects and arachnids within the arthropods and the vertebrates within the chordates. Evidence for pain among many phyla remains weak or absent, but that might simply reflect lack of relevant studies. However, a patchy distribution of pain might occur if pain was lost in some lineages. For example, taxa that evolved from a free-living form to a sedentary lifestyle may have reduced their behavioural choices and, thus, there may be no need for free choice and pain. For example, bivalve molluscs, such as oysters and mussels, which are fixed to hard substrates and therefore limited in how they might respond to noxious stimuli, might not benefit from a pain system.

We might reasonably expect pain to be found in basal groups of these three phyla but within the arthropods, identifying basal groups with extant examples has proved difficult

34

(Edgecombe and Legg 2014). Because pain has been suggested for the decapods, we should examine basal crustaceans, for which ostracods or branchiopods represent extant early taxa, but we are not aware of any studies that might indicate sentience in these groups. There is also a paucity of relevant studies on primitive insects (Gibbons et al. 2022a). Basal molluscs, such as the worm-like aplacophorans, provide no evidence for sentience because these are deep-sea burrowing animals and we are not aware of suitable studies on live specimens (Wanninger and Wollesen 2019). There is more information on early chordates, for example the protochordates, including *Amphioxus* (Lacalli 2022). There has been detailed comparison of the CNS of *Amphioxus* with those of vertebrates. This indicates that the brain of *Amphioxus* has some of the areas found in vertebrates, but *Amphioxus* lacks the major areas involved in the sensory experience of vertebrates. Lacalli (2022) concludes that sentience developed within the vertebrates rather than being a feature of the protochordates.

In general, the evidence points to the less parsimonious multiple origins of sentience and pain. One reason for this is suggested by Lacalli (2022), specifically for the chordates, but which might apply to the molluscs and arthropods. Early groups in these taxa lack well developed sensory systems. For example, light-sensitive cells may be found in early forms, but they likely only provide information about light levels. Whilst they might provide warning due to the shadow of a predator, they do not provide an image that came with the evolution of eyes. Eyes have evolved in some groups of molluscs, vertebrates and arthropods, and although these differ in composition, they are able to form images of distant objects and thus gather vast amounts of information (Godfrey-Smith 2020). If that is processed efficiently, it may be used to predict what will happen next. For example, improved sensory ability provides information about potential mates, potential competitors, potential predators and a myriad of other environmental changes that might impact fitness. This improved sensing is not restricted to vision but involves other modalities for which there has been marked development in the appropriate sense organs. The inte-

35

grated processing of this much larger amount of information has necessitated a parallel development of nervous systems. This was particularly likely in those animals that developed a highly mobile predatory lifestyle and the requirement for swift decision-making. Thus development of special senses leading to a substantial enlargement of information and potential action space may have stimulated the parallel development of AUL and sentience. In engineering terms, the state-space of sensory information and potential actions completely outstripped the capabilities of automaton-based systems, necessitating autonomous affect-driven decision-making. We suggest that the resulting flexibility of behaviour and the vastly improved ability to predict has brought about the requirement for pain in the context of affect-driven decision-making. That is, pain may be a consequence of mobility and behavioural choice (anticipatory behavioural autonomy) that we see in fish (and other vertebrates), cephalopods and decapods and some insects and arachnids. Animals outside of these specific groups, but with similar sensory and behavioural properties, may be considered likely to also experience pain. One suggestion for this is the crustacean group of stomatopods, commonly called mantis shrimps, which so far appear to have been excluded from a consideration of pain-like states. Based on our arguments, we might also expect to find evidence for pain in other arthropods such as spiders, scorpions, millipedes and centipedes, and we encourage studies of such animals.

4   Conclusions

We have considered pain as an experiential phenomenon emergent from the neural processing of nociceptive signals in the context of a self-model which is integrated with a neurohormonal system that provides emotional valence. Pain causes suffering because the pained state is remote from the goal (pain-free) state. In this view, the biological systems needed to cause suffering are no more than those needed to cause pain, so where

36

pain is established, suffering is likely too. We consider the term psychological stress to be equivalent to suffering and note that psychological stress is the primary measure for animal welfare studies and normally quantified by surrogate stress-hormone assays.

Our proposition can be put rather simply: pain is adaptive only for organisms capable of anticipatory behavioural autonomy (ABA), which is the freedom to choose among available behaviours based on model-derived anticipation of the outcomes, so pain could reasonably be attributed to any organism capable of that. The hypothetical system we propose for achieving ABA is just one of presumably many biologically plausible systems, but the components and architecture of their assembly into a working action-selection system are all testable. It is consistent with previous models of anticipatory behavioural control (Hoffmann 2003), the imperativist account of pain (Martinez 2015), the 'organisational approach' (Mossio et al. 2009) explanation of autonomy (Bich and Damiano 2012, Froese et al. 2007, Farnsworth 2018), proposed hallmarks of consciousness (Ginsburg and Jablonka 2019) and empirical findings in neuroanatomy Barajas-Azpeleta et al. (2021), Jiang and Pan (2022) and ideas about animal behaviour Budaev et al. (2020), Clayton et al. (2003), Crump et al. (2022), Elwood (2019), Ponte et al. (2022), Sneddon et al. (2014). One advantage of the concept we propose is that it succeeds in explaining apparent free choice as well as the role of emotional pain (suffering) in the control system of organisms possessing it. Another important advantage is that it is in principle testable using animal behaviour experiments.

Appendix

On the philosophical analysis of pain

Corns (2018) reviewed some of the main currents in the modern philosophy of pain, especially the evaluativist versus imperativist debate (we note that earlier perceptual and sense-datum theories have fallen largely obsolete (Aydede 2019)). The evaluative

37

thesis holds that pain is a representation of bodily damage that is experienced and interpreted normatively (Jacobson 2019, Bain 2013). Being strictly subjective and phenomenological (essentially perceptualist), the evaluative thesis is not well connected with mechanistic explanation, nor does it easily deal with empirical evidence showing rather weak correlation between bodily damage (or disturbance) and pain (Wright 2011, Garland 2012). Though the perceptualist approach is probably consistent with the sensory-descriminatory aspect of pain, for many, especially imperativists, it remains unclear how it can address the affective-motivational aspect of pain. However, that criticism was rejected by Bain (2013) and was addressed by psycho-functionalism (Aydede and Fulkerson 2019) which claims to explain all aspects of pain by combining features of both evaluativist and motivationalist accounts. The simple (and common) idea that pain is an informing signal that is accompanied by an affective state related to its normative valuation is not easy to explain in terms of biological mechanisms or even as an evolutionary adaptation. We believe the imperativist account, in which pain is a command for action intended to avoid it (Klein 2007; 2015b) is more useful to understanding the biological basis for pain.

Casser (2021) explains that imperativists believe "*there is no biological evidence to support the notion that pain was originally selected for its informative capacities, nor that it currently contributes to the fitness of organisms in this specific capacity*". According to the imperativist account, pain, he says, like other "*homeostatic sensations, such as hunger or thirst, [is] best interpreted as an action-guiding, imperative signal that serves the biological function of bringing about appropriate protective behaviours*", citing leading proponents of the imperativist theory: (Martinez 2011, Klein 2015a, Martinez and Klein 2016). Although Casser (2021) also warns that imperativists "*need to worry about the adequacy of their assumptions concerning pain's biological function*", it seems he finds more support for pain as an imperative than as an informant.

Most studies of pain so far have concentrated on perception and attempts to explain

it as a higher level of cognition beyond nociception (Aydede 2019). As such, pain is a construct of the mind. This idea gains support from reports of vicarious pain, in which some humans experience the pain of those they observe (Osborn and Derbyshire 2010), a phenomenon taken to be part of mirror-touch synesthesia (Banissy and Ward 2013). Pain, if considered as an example of *qualia*, is explained by proponents of active inference and predictive processing theories as an 'intermediate-level construct' (statistical model) of the world used by the mind to account for the cause of perceived signals (Clark et al. 2019), which may include the whole neuro-endocrine-immune system in 'embodied predictive processing' (Kiverstein et al. 2022).

Such cognitive constructs are described as models of the self and the environment and in general, models are instantiations of information about objects and their, especially causal, relationships (Rosen 1985). The necessary information can be innate (in the case of the algorithmic control system); otherwise it must be acquired by the organism, i.e. learned. Of particular relevance to pain, reinforcement learning (aversion) is ubiquitous among animals with a nervous system and can be achieved by direct stimulus-response modulation (Elwood 2019). Proponents of predictive processing and active inference theories (following Friston et al. 2013) consider brains (presumably nervous systems in general) as prediction engines that use learning, in the form of Bayesian updating of a statistical model, for determining the behaviour of the organism. In AI engineering, adaptive learning systems have been augmented by incorporating a self-representing model to enable planning (additional to adaptive response). This self-model, which is internal to the system, provides a means of anticipating the outcomes of different actions, extending anticipation of stimuli to become evaluative action selection (Butz and Hoffmann 2002). It is these systems that we take to be the primary reason for pain being adaptive.

39

On Autonomy

The idea of information controlling the behaviour of a system derives from a theory of physical causation, which is able to account for the unique attributes of biological systems: the appearance of downward causation and nested hierarchy (see e.g. Jaeger and Calkins (2012), Noble (2012)) and most relevant here, the emergence of autonomy from circular causation (Bich et al. (2016), Farnsworth (2018), Froese et al. (2007), Marshall et al. (2017), Moreno and Mossio (2015), Varela (1979)).

Autonomy is achieved through the information embodied within the organism constraining physical forces that are produced by it (Farnsworth 2022). Agent causation further implies a degree of causal isolation that is probably unique to life (Friston et al. 2013, Juel et al. 2019, Albantakis et al. 2019), the philosophical implications of which are discussed by Kauffman and Clayton (2006), Meincke (2019), Mossio et al. (2013).

Three levels of autonomy can be identified using an 'organisational approach to biological systems' (Mossio et al. 2009): a) none (system states are entirely determined by the environment, as in most natural systems); b) automaton (system states are at least partly determined by internal (embodied) control, as in clockwork automata, which may additionally be predetermined by a decision-making algorithm, as in assembly-line robots) and c) proactive autonomy (system states are at least partly determined by internal control that is a-priori indeterminate and to that extent free). All life, by virtue of being closed to efficient causation (Rosen 1991) has at least the automaton level, e.g. the automaton algorithm has been identified in plants (Kawano et al. 2012) and bacteria (Lan and Tu 2016). The idealised behaviour of an organism in a Skinner box coincides with that of an automaton having Bayesian updating, but only because it constrains the system to make a necessary causal connection between input (stimuli) and output (behaviour). However, many organisms, free of such artificial constraints, are able to select and enact behaviours that are not causally connected to the stimulus presented. Such behaviours may be responses to internal states, or even the result of high-level

cognition, e.g. the organism may play with the apparatus or seek attention in some way. These are displays of autonomous behaviour, in that they are not entirely caused by exogenous stimuli and are therefore not a-priori predictable from a knowledge of the inputs.

Declarations:

Ethical Approval

not applicable.

Competing interests

None.

Author's Contributions

KDF contributed to the conception and writing; RWE contributed to conception and writing.

Funding

None

Availability of data and materials

No data or materials were used.

References

Adamo, S. A. (2019). Is it pain if it does not hurt? on the unlikelihood of insect pain. *Can. Entomol.* 151, 685–695. doi:https://doi.org/10.4039/tce.2019.49

Albantakis, L., Marshall, W., Hoel, E., and Tononi, G. (2019). What causes what? A quantitative account of Actual Causation using dynamic causal networks. *Entropy* 21, doi:10.3390/e21050459

Antonopoulos, C. G., Fokas, A. S., and Bountis, T. C. (2016). Dynamical complexity in the c.elegans neural network. *Eur. Phys. J. Spec. Top.* 225, 1255–1269. doi:https://doi.org/10.1140/epjst/e2016-02670-3

Appel, M. and Elwood, R. W. (2009). Gender differences, responsiveness and memory of a potentially painful event in hermit crabs. *Anim. Behav.* 78, 1373–1379. doi:https://doi.org/10.1016/j.anbehav.2009.09.008

Aydede, M. (2019). Pain. In *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta (Metaphysics Research Lab, Stanford University). Spring 2019 edn.

Aydede, M. and Fulkerson, M. (2019). *The Philosophy of Pain: Unpleasantness, Emotion, and Deviance* (New York, USA.: Routledge), chap. Reasons and Theories of Sensory Affect. 27–59

Bain, D. (2013). What makes pains unpleasant? *Philos. Stud.* 166, S69–S89. doi:https://doi.org/10.1007/s11098-012-0049-7

Balasko, M. and Cabanac, M. (1998). Behavior of juvenile lizards (*Iguana iguana*) in a conflict between temperature regulation and palatable food. *Brain Behav. Evol.* 52, 257–262. doi:https://doi.org/10.1159/000006570

Baltzer-Jaray, K. (2018). *Homunculus* (John Wiley & Sons, Ltd), chap. 31. 165–167. doi:https://doi.org/10.1002/9781119165811.ch31

Banissy, M. and Ward, J. (2013). Mechanisms of self-other representations and vicarious experiences of touch in mirror-touch synesthesia. *Front. Hum. Neurosci.* 7. doi:https://doi.org/10.3389/fnhum.2013.00112

Barajas-Azpeleta, R., Tastekin, I., and Ribeiro, C. (2021). Neuroscience: How the brain prioritizes behaviors. *Curr. Biol.* 31, R1125–R1127. doi:https://doi.org/10.1016/j.cub.2021.08.048

Bargmann, C. (2012). Beyond the connectome: How neuromodulators shape neural circuits. *Bioessays* 34, 458–465. doi:https://doi.org/10.1002/bies.201100185

Barlassina, L. and Hayward, M. K. (2019). More of me! less of me!: Reflexive imperativism about affective phenomenal character. *Mind* 128, 1013–1044. doi:https://doi.org/10.1093/mind/fzz035

Barron, A. B. and Klein, C. (2016). What insects can tell us about the origins of consciousness. *Proc. Natl. Acad. Sci. U. S. A.* 113, 4900–4908. doi:https://doi.org/10.1073/pnas.1520084113

Berg, H. (2004). */emphE. coli in Motion.* (New York, USA.: Springer), chap. Optimal Control. No. 6 in Biological and Medical Physics, Biomedical Engineering. 61–67. doi:https://doi.org/10.1007/978-0-387-21638-6_7

Best, J., Duncan, W., Sadre-Marandi, F., Hashemi, P., Nijhout, H. F., and Reed, M. (2020). Autoreceptor control of serotonin dynamics. *BMC Neurosci.* 21. doi:https://doi.org/10.1186/s12868-020-00587-z

Bich, L. and Damiano, L. (2012). Life, autonomy and cognition: An organizational approach to the definition of the universal properties of life. *Orig. Life Evol. Biosph.* 42, 389–397. doi:https://doi.org/10.1007/s11084-012-9300-7

Bich, L., Mossio, M., Ruiz-Mirazo, K., and Moreno, A. (2016). Biological regulation: controlling the system from within. *Biol. Philos.* 31, 237–265. doi:https://doi.org/10.1007/s10539-015-9497-8

Bich, L., Mossio, M., and Soto, A. M. (2020). Glycemia regulation: From feedback loops to organizational closure. *Front. Physiol.* 11. doi:https://doi.org/10.3389/fphys.2020.00069

Birch, J., Ginsburg, S., and Jablonka, E. (2020a). Unlimited associative learning and the origins of consciousness: a primer and some predictions. *Biol. Philos.* 35. doi:https://doi.org/10.1007/s10539-020-09772-0

Birch, J., Schnell, A. K., and Clayton, N. S. (2020b). Dimensions of animal consciousness. *Trends Cognit. Sci.* 24, 789–801. doi:https://doi.org/10.1016/j.tics.2020.07.007

Bond, A. B., Kamil, A. C., and Balda, R. P. (2007). Serial reversal learning and the evolution of behavioral flexibility in three species of north american corvids (*Gymnorhinus cyanocephalus, Nucifraga columbiana, Aphelocoma californica*). *J. Comp. Psychol.* 121, 372–379

Brette, R. (2021). Integrative neuroscience of paramecium, a "swimming neuron". *eNeuro* 8. doi:https://doi.org/10.1523/ENEURO.0018-21.2021

Broekens, J., Jacobs, E., and Jonker, C. M. (2015). A reinforcement learning model of joy, distress, hope and fear. *Conn. Sci.* 27, 215–233. doi:https://doi.org/10.1080/09540091.2015.1031081

Broom, D. M. (1968). Specific habituation by chicks. *Nature* 217, 880–881. doi:https://doi.org/10.1038/217880a0

Brown, G. E., Ferrari, M. C. O., Elvidge, C. K., Ramnarine, I., and Chivers, D. P. (2013). Phenotypically plastic neophobia: a response to variable predation risk. *Proc. Roy. Soc. London. B.* 280. doi:https://doi.org/10.1098/rspb.2012.2712

Brown, R., LeDoux, J., and Rosenthal, D. (2021). The extra ingredient. *Biol. Philos.* 36. doi:https://doi.org/10.1007/s10539-021-09797-z

Browning, H. and Birch, J. (2022). Animal sentience. *Philos. Compass* 17. doi:https://doi.org/10.1111/phc3.12822

Budaev, S., Kristiansen, T. S., Giske, J., and Eliassen, S. (2020). Computational animal welfare: towards cognitive architecture models of animal sentience, emotion and wellbeing. *Royal Soc. Open Sci.* 7, 201886

Butz, M. V. and Hoffmann, J. (2002). Anticipations control behavior: Animal behavior in an anticipatory learning classifier system. *Adapt. Behav.* 10, 75–96. doi:https://doi.org/10.1177/1059-712302-010002-01

Cabanac, M. (1992). Pleasure: the common currency. *J. Theor. Biol.* 155, 173–200. doi:https://doi.org/10.1016/S0022-5193(05)80594-6

Cabanac, M. (2002). What is emotion? *Behav. Processes* 60, 69–83. doi:https://doi.org/10.1016/s0376-6357(02)00078-5

Casser, L. C. (2021). The function of pain. *Aust. J. Philos.* 99, 364–378. doi:https://doi.org/10.1080/00048402.2020.1735459

Cerqueira, M., Millot, S., Silva, T., Felix, A. S., Castanheira, M. F., Rey, S., et al. (2021). Stressor controllability modulates the stress response in fish. *BMC Neurosci.* 22. doi:https://doi.org/10.1186/s12868-021-00653-0

Chang, E. (2005). Stressed-out lobsters: Crustacean hyperglycemic hormone and stress proteins. *Integr. Comp. Biol.* 45, 43–50. doi:https://doi.org/10.1093/icb/45.1.43. Symposium on EcoPhysiology and Conservation - The Contribution of Endocrinology and Immunology, New Orleans, LA, JAN 05-09, 2004

Cheng, K. (2021). Learning in cnidaria: A systematic review. *Learning and Behavior* 49, 175–189. doi:https://doi.org/10.3758/s13420-020-00452-3

Cheng, K. (2022). Oscillators and servomechanisms in orientation and navigation, and sometimes in cognition. *Proc. Roy. Soc. London. B.* 289. doi:https://doi.org/10.1098/rspb.2022.0237

Cheriyamkunnel, S. J., Rose, S., Jacob, P. F., Blackburn, L. A., Glasgow, S., Moorse, J., et al. (2021). A neuronal mechanism controlling the choice between feeding and sexual behaviors in drosophila. *Curr. Biol.* 31, 4231–4245. doi:https://doi.org/10.1016/j.cub.2021.07.029

Cheung, B., Cohen, M., Rogers, C., and Albayram O., e. a. (2005). Experience-dependent modulation of *C. elegans* behavior by ambient oxygen. *Curr. Biol.* 15, 905–917. doi:https://doi.org/10.1016/j.cub.2005.04.017

Chung, T.-T., Darmaillacq, A.-S., Dickel, L., and Chiao, C.-C. (2022). The effect of unexpected rewards on decision making in cuttlefish. *Sci. Rep.* 12, 2514. doi:https://doi.org/10.1038/s41598-022-06443-w

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

Clark, A., Friston, K., and Wilkinson, S. (2019). Bayesing qualia consciousness as inference, not raw datum. *J. Conscious. Stud.* 26, 19–33

Clayton, N., Bussey, T., and Dickinson, A. (2003). Can animals recall the past and plan for the future? *Nat. Rev. Neurosci.* 4, 685–691. doi:https://doi.org/10.1038/nrn1180

Cooke, J. E. (2021). What is consciousness? integrated information vs. inference. *Entropy* 23. doi:https://doi.org/10.3390/e23081032

Corns, J. (2014). The inadequacy of unitary characterizations of pain. *Philos. Stud.* 169, 355–378. doi:https://doi.org/10.1007/s11098-013-0186-7

Corns, J. (2018). Recent work on pain. *Analysis* 78, 737–753. doi:https://doi.org/10.1093/analys/any055

Crook, R. J. (2021). Behavioral and neurophysiological evidence suggests affective pain experience in octopus. *ISci.* 24. doi:https://doi.org/10.1016/j.isci.2021.102229

Crook, R. J., Dickson, K., Hanlon, R. T., and Walters, E. T. (2014). Nociceptive sensitization reduces predation risk. *Curr. Biol.* 24, 1121–1125. doi:https://doi.org/10.1016/j.cub.2014.03.043

Crump, A., Browning, H., Schnell, A., Burn, C., and Birch, J. (2022). Sentience in decapod crustaceans: A general framework and review of the evidence. *Anim. Scentience* 32. doi:https://doi.org/10.51291/2377-7478.1691

Cunningham, J. (2001). Towards an axiomatic theory of consciousness. *Logic J. IGPL* 9, 325–331. doi:https://doi.org/10.1093/jigpal/9.2.325

Curran, K. P. and Chalasani, S. H. (2012). Serotonin circuits and anxiety: what can invertebrates teach us? *Invert. Neurosci.* 12, 81–92. doi:https://doi.org/10.1007/s10158-012-0140-y

Czaczkes, T. J., Koch, A., Froeber, K., and Dreisbach, G. (2018). Voluntary switching in an invertebrate: The effect of cue and reward change. *J. Exp. Psychol. Animal Learn. Cogn.* 44, 247–257. doi:https://doi.org/10.1037/xan0000171

de Abreu, M. S., Maximino, C., Banha, F., Anastacio, P. M., Demin, K. A., Kalueff, A. V., et al. (2020). Emotional behavior in aquatic organisms? lessons from crayfish and zebrafish. *J. Neurosci. Res.* 98, 764–779. doi:https://doi.org/10.1002/jnr.24550

Dennett, D. C. (1991). *Consciousness Explained* (Boston, M.A.: Little, Brown and Co.)

Dickinson, A. (1980). *Contemporary animal learning theory* (Cambridge, UK.: Cambridge University Press)

Earley, R. L., Lu, C.-K., Lee, I.-H., Wong, S. C., and Hsu, Y. (2013). Winner and loser effects are modulated by hormonal states. *Front. Zoo.* 10, 6. doi:https://doi.org/10.1186/1742-9994-10-6

Edgecombe, G. D. and Legg, D. A. (2014). Origins and early evolution of arthropods. *Palaeontology* 57, 457–468. doi:https://doi.org/2https://doi.org/10.1111/pala.12105

Eliezer, Y., Deshe, N., Hoch, L., Iwanir, S., Pritz, C. O., and Zaslaver, A. (2019). A memory circuit for coping with impending adversity. *Curr. Biol.* 29, 1573+. doi:https://doi.org/10.1016/j.cub.2019.03.059

Ellis, G. F. R. and Kopel, J. (2019). The dynamical emergence of biology from physics: Branching causation via biomolecules. *Front. Psychol.* 9. doi:https://doi.org/10.3389/fphys.2018.01966

Elwood, R. W. (2011). Pain and suffering in invertebrates? *ILAR J.* 52, 175–184. doi:https://doi.org/10.1093/ilar.52.2.175

Elwood, R. W. (2019). Discrimination between nociceptive reflexes and more complex responses consistent with pain in crustaceans. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 374. doi:https://doi.org/10.1098/rstb.2019.0368

Elwood, R. W. (2021). Potential pain in fish and decapods: Similar experimental approaches and similar results. *Front. Vet. Sci.* 8. doi:https://doi.org/10.3389/fvets.2021.631151

Elwood, R. W. (2022). Hermit crabs, shells, and sentience. *Animal Cognition* doi:https://doi.org/10.1007/s10071-022-01607-7

Elwood, R. W. and Adams, L. (2015). Electric shock causes physiological stress responses in shore crabs, consistent with prediction of pain. *Biol. Lett.* 11. doi:https://doi.org/10.1098/rsbl.2015.0800

Elwood, R. W. and Appel, M. (2009). Pain experience in hermit crabs? *Anim. Behav.* 77, 1243–1246. doi:https://doi.org/10.1016/j.anbehav.2009.01.028

Farnsworth, K. D. (2017). Can a robot have free will? *Entropy* 19, 237. doi:https://doi.org/10.3390/e19050237

Farnsworth, K. D. (2018). How organisms gained causal independence and how it might be quantified. *Biology* 7, 38. doi:https://doi.org/10.3390/biology7030038

Farnsworth, K. D. (2022). How an information perspective helps overcome the challenge of biology to physics. *BioSystems* , 104683doi:https://doi.org/10.1016/j.biosystems.2022.104683

Farnsworth, K. D., Nelson, J., and Gershenson, C. (2013). Living is information processing: From molecules to global systems. *Acta Biotheor.* 61, 203–222. doi:https://doi.org/10.1007/s10441-013-9179-3

Feinberg, T. E. and Mallatt, J. (2013). The evolutionary and genetic origins of consciousness in the cambrian period over 500 million years ago. *Front. Psychol.* 4. doi:https://doi.org/10.3389/fpsyg.2013.00667

Fossat, P., Bacque-Cazenave, J., De Deurwaerdere, P., Delbecque, J.-P., and Cattaert, D. (2014). Comparative behavior anxiety-like behavior in crayfish is controlled by serotonin. *Science* 344, 1293–1297. doi:https://doi.org/10.1126/science.1248811

Friston, K. J., Schwartenbeck, P., FitzGerald, T., Moutoussis, M., Behrens, T., and Dolan, R. J. (2013). The anatomy of choice: active inference and agency. *Front. Hum. Neurosci.* 7. doi:https://doi.org/10.3389/fnhum.2013.00598

Froese, T., Virgo, N., and Izquierdo, E. (2007). Autonomy: a review and a reappraisal. In *European Conference on Artificial Life* (Springer Berlin Heidelberg), 455–464

Fulkerson, M. (2021). How thirst compels: An aggregation model of sensory motivation. *Mind Lang.* doi:https://doi.org/10.1111/mila.12369

Garland, E. L. (2012). Pain processing in the human nervous system a selective review of nociceptive and biobehavioral pathways. *Prim. Care* 39, 561+. doi:https://doi.org/10.1016/j.pop.2012.06.013

Gibbons, M., Crump, A., Barrett, M., Sarlak, S., Birch, J., and Chittka, L. (2022a). *Advances in Insect Physiology* (Academic Press), vol. 63, chap. Can Insects Feel Pain? 155–229. doi:https://doi.org/10.1016/bs.aiip.2022.10.001

Gibbons, M., Versace, E., Crump, A., Baran, B., and Chittka, L. (2022b). Motivational trade-offs and modulation of nociception in bumblebees. *Proc. Natl. Acad. Sci. U. S. A.* 31. doi:0.1073/pnas.2205821119

Ginsburg, S. and Jablonka, E. (2019). *The Evolution of the Sensitive Soul: Learning and the Origins of Consciousness* (Cambridge, MA: MIT Press). doi:https://doi.org/10.7551/mitpress/11006.001.0001

Godfrey-Smith, P. (2020). *Metazoa: Animal minds and the birth of consciousness.* (London: William Collins)

Graves, A., Wayne, G., and Danihelka, I. (2014). Neural turing machines. *CoRR* abs/1410.5401

Greve, R. B., Jacobsen, E. J., and Risi, S. (2016). Evolving neural turing machines for reward-based learning. In *Proceedings of the genetic and evolutionary computation conference 2016.* 117–124. doi:https://doi.org/10.1145/2908812.2908930

Hall, R. J. (2008). If it itches, scratch! *Aust. J. Philos.* 86, 525–535. doi:https://doi.org/10.1080/00048400802346813

Helm, B. W. (2002). Felt evaluations: A theory of pleasure and pain. *Am Philos Q* 39, 13–30

Hoffmann, J. (2003). Anticipatory behavioral control. In *Anticipatory behavior in adaptive learning systems*, eds. M. Butz, O. Sigaud, and P. Gérard (Springer), vol. 2684 of *Lecture Notes in Computer Science*, chap. (). 44–65. doi:https://doi.org/10.1007/978-3-540-45002-3_4

Jacobson, H. (2019). Not only a messenger: Towards an attitudinal-representational theory of pain. *Philos. Phenomenol. Res.* 99, 382–408. doi:https://doi.org/10.1111/phpr.12493

Jaeger, L. and Calkins, E. R. (2012). Downward causation by information control in micro-organisms. *Interface Focus* 2, 26–41

Jeanson, R. (2019). Within-individual behavioural variability and division of labour in social insects. *J. Exp. Biol.* 222. doi:https://doi.org/10.1242/jeb.190868. Jeb190868

Jékely, G., Godfrey-Smith, P., and Keijzer, F. (2021). Reafference and the origin of the self in early nervous system evolution. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 376, 20190764. doi:https://doi.org/10.1098/rstb.2019.0764

Jiang, X. and Pan, Y. (2022). Neural control of action selection among innate behaviors. *Neurosci. Bull.* doi:https://doi.org/10.1007/s12264-022-00886-x

Juel, B., Comolatti, R., Tononi, G., and Albantakis, L. (2019). When is an action caused from within? Quantifying the causal chain leading to actions in simulated agents. *Entropy* 21, doi: 10.3390/e21050459

Kaplan, H. S., Nichols, A. L., and Zimmer, M. (2018). Sensorimotor integration in *Caenorhabditis elegans*: a reappraisal towards dynamic and distributed computations.

Philos. Trans. R. Soc. Lond., B, Biol. Sci. 373, 20170371. doi:https://doi.org/10.1098/rstb.2017.0371

Kauffman, S. A. and Clayton, P. (2006). On emergence, agency, and organization. *Biol. Philos.* 21, 501–521. doi:https://doi.org/10.1007/s10539-005-9003-9

Kawano, T., Bouteau, F., and Mancuso, S. (2012). Finding and defining the natural automata acting in living plants. *Commun. Integr. Biol.* 5, 519–526

Key, B. (2015). Fish do not feel pain and its implications for understanding phenomenal consciousness. *Biol. Philos.* 30, 149–165. doi:https://doi.org/10.1007/s10539-014-9469-4

Key, B. and Brown, D. (2018). Designing brains for pain: Human to mollusc. *Front. Physiol.* 9. doi:https://doi.org/10.3389/fphys.2018.01027

Key, B., Zalucki, O., and Brown, D. J. (2021). Neural design principles for subjective experience: Implications for insects. *Front. Behav. Neurosci.* 15. doi:https://doi.org/10.3389/fnbeh.2021.658037

Key, B., Zalucki, O., and Brown, D. J. (2022). A first principles approach to subjective experience. *Front. Syst. Neurosci.* 16. doi:https://doi.org/10.3389/fnsys.2022.756224

Kiverstein, J., Kirchhoff, M., and Thacker, M. (2022). An embodied predictive processing theory of pain experience. *Rev. Philos. Psychol.* doi:0.1007/s13164-022-00616-2

Klein, C. (2007). An imperative theory of pain. *J. Philos.* 104, 517–532. doi:https://doi.org/10.5840/jphil2007104104

Klein, C. (2015a). What pain asymbolia really shows. *Mind* 124, 493–516. doi:https://doi.org/10.1093/mind/fzu185

Klein, C. (2015b). *What the body commands: The impertive theory of pain* (MIT Press)

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

Lacalli, T. (2022). An evolutionary perspective on chordate brain organization and function: insights from amphioxus, and the problem of sentience. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 377. doi:https://doi.org/10.1098/rstb.2020.0520

Lan, G. and Tu, Y. (2016). Information processing in bacteria: memory, computation, and statistical physics: a key issues review. *Rep. Prog. Phys.* 79

Liu, P., Chen, B., and Wang, Z.-W. (2020). Gabaergic motor neurons bias locomotor decision-making in c. elegans. *Nat. Commun.* 11, 5076. doi:https://doi.org/10.1038/s41467-020-18893-9

Magee, B. and Elwood, R. W. (2013). Shock avoidance by discrimination learning in the shore crab (carcinus maenas) is consistent with a key criterion for pain. *J. Exp. Biol.* 216, 353–358. doi:https://doi.org/10.1242/jeb.072041

Magee, B. and Elwood, R. W. (2016). Trade-offs between predator avoidance and electric shock avoidance in hermit crabs demonstrate a non-reflexive response to noxious stimuli consistent with prediction of pain. *Behav. Processes* 130, 31–35. doi: https://doi.org/10.1016/j.beproc.2016.06.017

Magurran, A. (1993). *Behaviour of teleost fishes* (London.: Chapman & Hall), chap. Individual differences and alternative behaviours. 441–477

Marshall, W., Kim, H., Walker, S., Tononi, G., and Albantakis, L. (2017). How causal analysis can reveal autonomy in models of biological systems. *Philos. Trans. R. Soc. A-Math. Phys. Eng. Sci.* 375, DOI: 10.1098/srta.2016.0358

Martinez, M. (2011). Imperative content and the painfulness of pain. *Phenomenol Cogn Sci* 10, 67–90. doi:https://doi.org/10.1007/s11097-010-9172-0

Martinez, M. (2015). Pains as reasons. *Philos. Stud.* 172, 2261–2274. doi:https://doi. org/10.1007/s11098-014-0408-7

Martinez, M. and Klein, C. (2016). Pain signals are predominantly imperative. *Biol. Philos.* 31, 283–298. doi:https://doi.org/10.1007/s10539-015-9514-y

Matsumoto, T. and Tani, J. (2020). Goal-directed planning for habituated agents by active inference using a variational recurrent neural network. *Entropy* 22. doi:https: //doi.org/10.3390/e22050564

McFarland, D., Sibly, R., and Thorpe, W. H. (1975). The behavioural final common path. *Phil. Trans. R. Soc. B.* 270, 265–293. doi:https://doi.org/10.1098/rstb.1975.0009

McNamara, J. and Houston, A. (1986). The common currency for behavioural decisions. *Am. Nat.* 127, 358–378. doi:doi.org/10.1086/284489

Meincke, A. S. (2019). Autopoiesis, biological autonomy and the process view of life. *Eur. J. Philos. Sci.* 9. doi:https://doi.org/10.1007/s13194-018-0228-2

Melzack, R. (2001). Pain and the neuromatrix in the brain. *J Dent Educ* 65, 1378–1382. doi:https://doi.org/10.1002/j.0022-0337.2001.65.12.tb03497.x

Melzack, R. and Casey, K. (1968). *The Skin Senses* (Springfield IL.: Thomas), chap. Sensory, motivational, and central control determinants of pain: A new conceptual model. pp. 423–443

Mendl, M. and Paul, E. S. (2020). Animal affect and decision-making. *Neurosci. Biobe- hav. Rev.* 112, 144–163. doi:https://doi.org/10.1016/j.neubiorev.2020.01.025

Michel, M. (2020). Consciousness science underdetermined: A short history of endless debates. *Ergo* 6, 771–809. doi:https://doi.org/10.3998/ergo.12405314.0006.028

Millsopp, S. and Laming, P. (2008). Trade-offs between feeding and shock avoidance in goldfish (carassius auratus). *App. Anim. Behav. Sci.* 113, 247–254. doi:https://doi.org/10.1016/j.applanim.2007.11.004

Montévil, M. and Mossio, M. (2015). Biological organisation as closure of constraints. *J. Theor. Biol.* 372, 179–191. doi:https://doi.org/10.1016/j.jtbi.2015.02.029

Moreno, A. and Mossio, M. (2015). *Biological Autonomy - a philosophical and theoretical enquiry* (Heidelberg: Springer)

Mossio, M., Bich, L., and Moreno, A. (2013). Emergence, Closure and Inter-level Causation in Biological Systems. *Erkenntnis* 78, 153–178. doi:https://doi.org/10.1007/s10670-013-9507-7

Mossio, M., Montevil, M., and Longo, G. (2016). Theoretical principles for biology: Organization. *Prog. Biophys. Mol. Biol.* 122, 24–35. doi:https://doi.org/10.1016/j.pbiomolbio.2016.07.005

Mossio, M., Saborido, C., and Moreno, A. (2009). An organisational account of biological functions. *Br J Philos Sci* 60, 813–841

Nani, A., Volpara, G., and Faggio, A. (2021). Sentience with or without consciousness. *J. Conscious. Stud.* 28, 60–79

Noble, D. (2012). A theory of biological relativity: no privileged level of causation. *Interface Focus* 2, 55–64

Ortony, A. (2022). Are all "basic emotions" emotions? a problem for the (basic) emotions construct. *Perspect Psychol Sci* 17, 41–61. doi:https://doi.org/10.1177/1745691620985415. PMID: 34264141

Osborn, J. and Derbyshire, S. (2010). Pain sensation evoked by observing injury in others. *Pain* 148, 268–274. doi:https://doi.org/10.1016/j.pain.2009.11.007

Perry, C. J. and Barron, A. B. (2013). Neural mechanisms of reward in insects. In *Annu. Rev. Entomol.*, ed. M. Berenbaum, vol. 58 of *Annual Review of Entomology*. 543–562. doi:https://doi.org/10.1146/annurev-ento-120811-153631

Pezzulo, G., Parr, T., and Friston, K. (2022). The evolution of brain architectures for predictive coding and active inference. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* 377. doi:https://doi.org/10.1098/rstb.2020.0531

Pietri, E., Fazio, R., and Shook, N. (2013). Weighting positive versus negative: the fundamental nature of valence asymmetry. *J. Person.* 81, 196–208. doi:https://doi.org/10.1111/j.1467-6494.2012.00800.x

Ponte, G., Chiandetti, C., Edelman, D. B., Imperadore, P., Pieroni, E. M., and Fiorito, G. (2022). Cephalopod behavior: From neural plasticity to consciousness. *Front. Syst. Neurosci.* 15. doi:https://doi.org/10.3389/fnsys.2021.787139

Posner, J., Russell, J. A., and Peterson, B. S. (2005). The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Dev. Psychopathol.* 17, 715–734. doi:https://doi.org/10.1017/S0954579405050340

Puri, S. and Faulkes, Z. (2015). Can crayfish take the heat? *Procambarus clarkii* show nociceptive behaviour to high temperature stimuli, but not low temperature or chemical stimuli. *Biol. Open* 4, 441–448. doi:https://doi.org/10.1242/bio.20149654

Raja, S. N., Carr, D. B., Cohen, M., Finnerup, N. B., Flor, H., Gibson, S., et al. (2020). The revised international association for the study of pain definition of pain: concepts, challenges, and compromises. *Pain* 161, 1976–1982. doi:https://doi.org/10.1097/j.pain.0000000000001939

Ravary, F., Lecoutey, E., Kaminski, G., Chaline, N., and Jaisson, P. (2007). Individual

experience alone can generate lasting division of labor in ants. *Curr. Biol.* 17, 1308–1312. doi:https://doi.org/10.1016/j.cub.2007.06.047

Ridderinkhof, K. R. (2017). Emotion in action: A predictive processing perspective and theoretical synthesis. *Emot Rev* 9, 319–325. doi:https://doi.org/10.1177/1754073916661765

Rosen, R. (1985). *Anticipatory systems: Philosophical, mathematical, and methodological foundations* (Oxford; Toronto: Pergamon Press)

Rosen, R. (1991). *Life itself: A comprehensive enquiry into the nature, origin and fabrication of life* (New York, USA.: Columbia University Press)

Russell, J. (1978). Evidence of convergent validity on the dimensions of affect. *J Pers Soc Psychol* 36, 1152–1168. doi:https://doi.org/10.1037/0022-3514.36.10.1152

Russell, J. and Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *J Pers Soc Psychol* 76, 805–819. doi:https://doi.org/10.1037/0022-3514.76.5.805

Scheffer, L. K. and Meinertzhagen, I. A. (2019). The fly brain atlas. *Annu. Rev. Cell Dev. Biol.* 35, 637–653. doi:https://doi.org/10.1146/annurev-cellbio-100818-125444. PMID: 31283380

Scheffer, L. K., Xu, C. S., Januszewski, M., Lu, Z., Takemura, S.-Y., Hayworth, K. J., et al. (2020). A connectome and analysis of the adult drosophila central brain. *eLife* 9, e57443. doi:https://doi.org/10.7554/eLife.57443

Scherer, K. (2022). Theory convergence in emotion science is timely and realistic. *Cogn Emot* 36, 154–170. doi:https://doi.org/10.1080/02699931.2021.1973378

Scherer, K., Bänziger, T., and Roesch, E. (eds.) (2001). *Appraisal processes in emotion: theory, methods, research* (Oxford University Press)

Schneider, M. (2018). A cybernetical perspective on emotion-cognition interactions as self-directed adaptation in situated decision-makers. *Cogn. Syst. Res.* 52, 400–428. doi:https://doi.org/10.1016/j.cogsys.2018.07.005

Sequeira, P., Melo, F. S., and Paiva, A. (2015). Emergence of emotional appraisal signals in reinforcement learning agents. *Auton Agent Multi Agent Syst* 29, 537–568. doi:https://doi.org/10.1007/s10458-014-9262-4

Sneddon, L. U. (2009). Pain perception in fish: Indicators and endpoints. *ILAR J.* 50, 338–342. doi:https://doi.org/10.1093/ilar.50.4.338

Sneddon, L. U. (2019). Evolution of nociception and pain: evidence from fish models. *Phil. Trans. R. Soc. B.* 374. doi:https://doi.org/10.1098/rstb.2019.0290

Sneddon, L. U., Braithwaite, V., and Gentle, M. (2003). Do fishes have nociceptors? evidence for the evolution of a vertebrate sensory system. *Proc. Roy. Soc. London. B.* 270, 115–1121. doi:https://doi.org/10.1098/rspb.2003.2349

Sneddon, L. U., Elwood, R. W., Adamo, S. A., and Leach, M. C. (2014). Defining and assessing animal pain. *Anim. Behav.* 97, 201–212. doi:https://doi.org/10.1016/j.anbehav.2014.09.007

Sonoda, K., Asakura, A., Minoura, M., Elwood, R. W., and Gunji, Y.-P. (2012). Hermit crabs perceive the extent of their virtual bodies. *Biol. Lett.* 8, 495–497. doi:https://doi.org/10.1098/rsbl.2012.0085

Stafford, J. and Mellor, D. (2005). Dehorning and disbudding distress and its alleviation in calves. *Vet. J.* 169, 337–349

Stamps, J. (2016). Individual differences in behaviouralplasticities. *Biol. Rev.* 91, 534–567. doi:https://doi.org/10.1111/brv.12186

Sterling, P. (2012). Allostasis: a model of predictive regulation. *Physiol. Behav.* 106, 5–15. doi:https://doi.org/10.1016/j.physbeh.2011.06.004

Suzuki, D. G. (2021). Consciousness in jawless fishes. *Front. Syst. Neurosci.* 15. doi: https://doi.org/10.3389/fnsys.2021.751876

Tarsitano, M. (2006). Route selection by a jumping spider (*Portia labiata*) during the locomotory phase of a detour. *Anim. Behav.* 72. doi:https://doi.org/10.1016/j.anbehav.2006.05.007

Tate, A. J. M. (2021). A predictive processing theory of motivation. *Synthese* 198, 4493–4521. doi:https://doi.org/10.1007/s11229-019-02354-y

Temizer, I., Donovan, J. C., Baier, H., and Semmelhack, J. L. (2015). A visual pathway for looming-evoked escape in larval zebrafish. *Curr. Biol.* 25, 1823–1834. doi:https://doi.org/10.1016/j.cub.2015.06.002

Tononi, G. (2008). Consciousness as integrated information: A provisional manifesto. *Biol. Bull.* 215, 216–242

Tye, M. (2021). Qualia. In *The Stanford Encyclopedia of Philosophy (Fall 2021)*, ed. E. N. Zalta (Metaphysics Research Lab, Stanford University)

Tymoshchuk, P. and Shatnyi, S. (2015). Hardware implementation of discrete-time neural circuit of largest/smallest signal identification. 226–230. doi:https://doi.org/10.1109/CADSM.2015.7230842

Varela, F. (1979). *Principles of Biological Autonomy* (Holland and N.Y. USA.: Elsevier)

Wagner, A. E. (2010). Effects of stress on pain in horses and incorporating pain scales for equine practice. *Vet. Clin. North Am. Equine Pract.* 26, 481+. doi:https://doi.org/10.1016/j.cveq.2010.07.001

1330 Wanninger, A. and Wollesen, T. (2019). The evolution of molluscs. *Biol. Rev.* 94,

1331   102–115. doi:https://doi.org/10.1111/brv.12439

1332 Wilson, E. (1976). Behavioral discretization and number of castes in an ant species.

1333   *Behav. Ecol. Sociobiol.* 1, 141–154. doi:https://doi.org/10.1007/BF00299195

1334 Wright, A. (2011). A criticism of the IASP's definition of pain. *J. Conscious. Stud.* 18,

1335   19–44