Pain: A Precision Signal
for Reinforcement Learning and Control

Ben Seymour[1,2,*]
[1]Center for Information and Neural Networks, National Institute of Information and Communications Technology, 1-4 Yamadaoka, Suita, Osaka 565-0871, Japan
[2]Computational and Biological Learning Lab, Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK
*Correspondence: bjs49@cam.ac.uk
https://doi.org/10.1016/j.neuron.2019.01.055

Since noxious stimulation usually leads to the perception of pain, pain has traditionally been considered sensory nociception. But its variability and sensitivity to a broad array of cognitive and motivational factors have meant it is commonly viewed as inherently imprecise and intangibly subjective. However, the core function of pain is motivational—to direct both short- and long-term behavior away from harm. Here, we illustrate that a reinforcement learning model of pain offers a mechanistic understanding of how the brain supports this, illustrating the underlying computational architecture of the pain system. Importantly, it explains why pain is tuned by multiple factors and necessarily supported by a distributed network of brain regions, recasting pain as a precise and objectifiable control signal.

Despite the advent of brain imaging, a clear picture of how pain is processed in the brain has been much harder to unravel than anticipated, being beset by three problems. First, pain is associated with robust responses in multiple and diverse brain regions, most of which are not specific to pain (at least on a macroscopic scale), and so it has been hard to "pin down" the pain system to any specific brain region. Second, pain is an inherently private percept, but an individual's self-reports of pain can vary widely from moment to moment, and it has remained unclear whether this fluctuation represents irreducible noise and subjectivity or a precise tuning of pain based on hidden factors. Third, pain is exquisitely sensitive to a broad range of emotional, environmental, and cognitive factors—a phenomenon called endogenous modulation. Although this has led to an appreciation that pain is more than a simple readout of nociceptive input, it has not led to any satisfactory unified explanation as to what pain really is. This has left the view that pain is simply a highly variable and malleable representation of assumed actual or potential tissue damage.

In this review, we propose a model of pain that centralizes its role as a learning and control signal and argue that this can solve these problems. We begin with a perspective of how theories of pain have evolved over recent decades, and how insights have emerged that have moved thinking beyond purely sensory accounts of pain. We then argue that current accounts still don't fully capture how pain controls behavior to minimize harm, which is its primary function. Importantly, although this is often achieved by immediate nocifensive responses, a substantial part of this comes from learning—allowing an animal to mitigate or avoid predictable harm long into the future. The foundations of a learning account of pain are rooted in psychological models of animal learning, and we describe how these can be developed in computational terms to provide a mechanistic model of the architecture of the pain system. Critically, we argue that this requires pain to be shaped by a set of factors to optimize its role as a learning and control signal and review evidence that

suggests that a great deal of examples of endogenous control can be explained by this process. Finally, we briefly describe how the model offers potential insights into how pain might become chronic under certain conditions.

Background

There is a long history of theories and constructs that have attempted to capture the complex phenomenology of pain, but a number of models have played a particularly important role in evolving current concepts of pain. Against a historically dominant view that pain could be understood as a sensory system like any other, Melzack and Casey (1968) highlighted what they called the "man-in-the-brain" problem that this seemed to create—the idea that the main function of the pain system was to inform some conscious module of the nature of a particular nociceptive stimulus. Instead, they proposed the tripartite model, in which sensory-discriminative, affective-motivational, and cognitive-evaluative components are processed as part-independent, part-interacting pathways (Melzack and Casey, 1968). In this model, rather than the control of protective behavior being merely downstream to sensory processing, they argued it was an intrinsic and fundamental part of pain experience, not least because pain was clearly sensitive to so many motivational and cognitive factors. A key substrate for modulation, namely descending pathways acting on dorsal horn neurons, had already been proposed in gate control theory (Melzack and Wall, 1965). And in the brain, the model implied that different dimensions would involve multiple different cerebral loci—a premonition of the distributed pattern of cortical and subcortical pain responses that was later revealed by functional neuroimaging (Jones et al., 1991; Treede et al., 1999).

In light of the neurophysiological characterization of many pain-specific receptors and ascending pathways, Craig subsequently proposed the homeostatic model, in which sensory and motivational dimensions are inherently integrated as a single system, involving pain-specific lateral thalamocortical projections to

the insula cortex (Craig, 2002, 2003). Craig placed pain alongside other ''interoceptive'' sensations such as temperature, itch, and pleasant touch, as systems supporting bodily perception with intrinsic motivational value, with this value related to the core homeostatic drive to maintain the integrity of the body. This model proposed a hierarchical sensory processing stream from posterior to anterior regions of insula cortex, with this hierarchy explicitly tied to physiological and behavioral homeostasis. However, it was still largely unclear exactly how homeostatic behaviors were actually implemented, and why pain was modulated by so many factors.

The idea that, as a motivational system, pain was directly modulated as a decision by the system itself was most clearly articulated in Fields' motivation-decision model (Fields, 2006, 2018). The model proposed that pain was inhibited when overshadowed by more important reward- or escape-orientated goals, with descending control mediated by opioid pathways via the periaqueductal gray (PAG) and rostral ventral medulla (RVM) (Basbaum and Fields, 1984). The model also highlighted the fundamental role of pain in learning and the control of avoidance and escape, and this explicitly motivational perspective viewed pain as controlling not just immediate responses to limit tissue damage, but also long-term role harm minimization through learned escape and avoidance (Johansen et al., 2001; Navratilova et al., 2012). This re-conceived pain as an inherently *predictive* system, not simply passively recording nociceptive inputs, in which the generation of pain predictions and expectations are central to the function of central pain circuits.

A central role for expectation and prediction also underlay the idea that pain involves a statistical (e.g., Bayesian) inference resulting from integration of prior expectancy and incoming nociceptive input (Brown et al., 2008; Seymour and Dolan, 2013; Morton et al., 2010; Anchisi and Zanon, 2015; Tabor et al., 2017; Ongaro and Kaptchuk, 2019). More formally, Buchel proposed the predictive coding model (Büchel et al., 2014), involving a hierarchical processing stream from spinal cord to PAG, thalamus, and posterior-to-anterior insula (Geuter et al., 2017; Grahl et al., 2018; Ozawa et al., 2017). In this context, expectations can be acquired through multiple means—through instructed information, learning (i.e., conditioning), and through observation (Wiech, 2016; Tabor et al., 2017)—and suggested that descending control might be implementing top-down predictions and their uncertainties, to be integrated by ascending nociception information and prediction errors. In so doing, this provided an explanation of a set of instances of endogenous control, especially expectancy-based biases and placebo and nocebo responses.

However, inferential theories of pain processing leave open an account of how the motivational function of pain is directed. At an abstract level, concepts such as Friston's Free Energy Framework propose that sensation, motivation, and action are intrinsically related by their drive to understand the causes of unexpected stimuli (Friston, 2010), and the notion of ''active inference'' describes how actions can be conceived to ultimately reflect minimization of future unexpected sensory instances of pain (Tabor and Burr, 2019). But understanding how the brain actually achieves this is much more complex (Pezzulo et al., 2015), and none of the existing models fully capture how the

pain system successfully balances stimulus identification, information seeking, harm minimization, and, perhaps most critically, speed. Below, we outline a computational architecture of the pain system that may achieve this, based on a framework called reinforcement learning (RL). In so doing, this casts light on the fundamental question of what the conscious perception of pain really reflects: an optimal inference of a real or presumed nociceptive stimulus, or an optimal control signal to minimize current and future nociceptive stimuli?

The RL Model

Underlying the evolution of these concepts is the central idea that pain must be understood in the context of behavioral control to minimize current and future harm. Fundamental to this concept is learning: over and above the fact that pain elicits immediate defensive responses (withdrawal, orientation, etc.), it must also guide learning to optimize future responses. To illustrate this, consider a child touching a hot stove: although the immediate response limits the severity of any burn, the main benefit is though the sum of future instances when they don't touch stoves because they have a pain system. Therefore, what has driven the evolution of the architecture of the pain system is its role as a learning signal to *prospectively* reduce harm. But understanding how the brain achieves this exposes a fundamental problem that any experience-based control system must solve: the credit assignment problem (Bellman, 2013).

The Credit Assignment Problem

Harm minimization is both a clearly definable and objectively measurable function and is based on the ability to learn from trial-and-error interaction with the world. This allows actions that terminate (escape) or completely avoid pain and has been well studied in humans and animal learning using Pavlovian (classical) and instrumental (operant) conditioning (Mackintosh, 1983). Most of these paradigms consider simple one-step escape or avoidance, in which pain is predicted by a single preceding cue, which subsequently elicits an appropriate response. However, real-world learning often involves much longer sequences of events, which makes the problem of prediction and avoidance more difficult. For instance, if a series of 5 actions leads to pain at the end, how do you know which of the 5 actions was the mistake (Figure 1A)? This problem is referred to as the credit assignment problem, which is a well-known problem in engineering and control theory (Bellman, 2013).

The credit assignment problem can be solved using a class of learning rule from the field of RL (Sutton and Barto, 1998). RL is effectively an extension of psychological learning models, such as the Rescorla-Wagner model (Rescorla and Wagner, 1972). The Rescorla-Wagner model is usually applied to one-step learning and uses a prediction error term—the difference between what was expected and observed—to update future predictions. But this doesn't work well if the outcome is far into the future. RL models, on the other hand, don't need to wait for the outcome and simply use the next available prediction (formalized as the *value*) as a proxy for the outcome. That is, they store a value term for each action and state and compute the difference between values at each successive time step, taking into account any reward and punishment experienced on the way (Sutton and Barto, 1998). Effectively, this allows pain predictions to
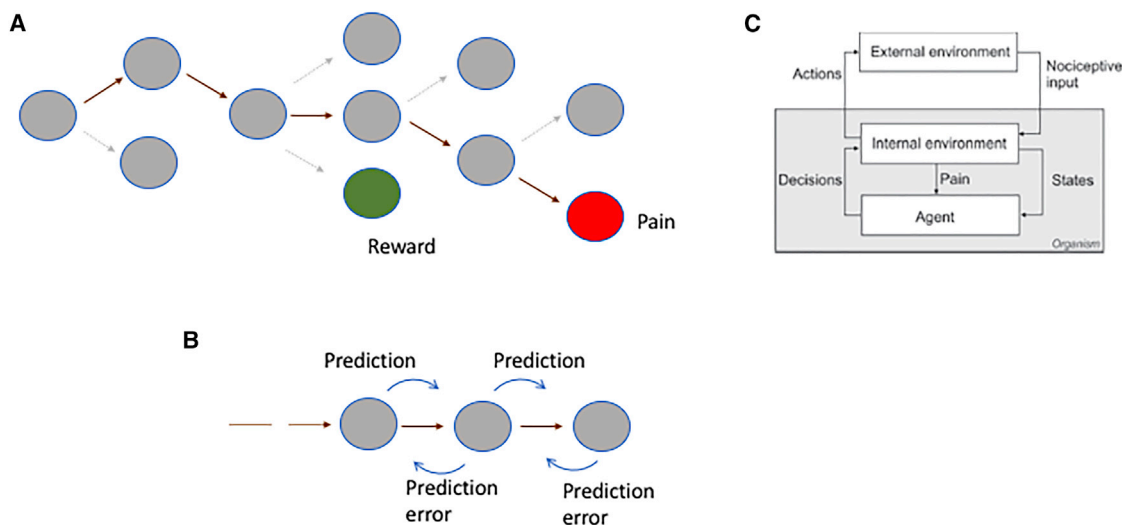
Figure 1. The Credit Assignment Problem and RL Framework
(A) RL provides an algorithmic framework for learning how to make optimal predictions and actions based on trial-and-error interaction with the world, in which salient outcomes (reward and punishment) can be sensed. In particular, it aims to solve the problem of correctly allocating predictive value to preceding states, in terms of the outcomes they eventually predict, when the transition through states of the world is either passive (not under the agent's control, as in Pavlovian learning) or active (determined by the agent's actions, as in instrumental learning).
(B) This is achieved by using a prediction error term to update state or action values, which effectively transfer value predictions back in time to the earliest predictor. Here, the prediction error is equal to the difference between the current prediction, and the sum of the subsequent prediction plus any outcome experienced at this next state.
(C) The agent-environment interface illustrates the basic architecture comprising the agent (which learns values, computes prediction errors, and selects actions), the internal environment (which represents states and outcomes), and the external environment (which contains the sensed objects).

be learned by looking at differences in predictions through time, which passes the prediction back to the earliest reliable predictor (Figure 1B). This error-based mechanism can be applied to passive predictions (learning state values) or active predictions (learning action values).

Figure 1C shows the basic architecture of RL control. By sensing the external environment, the organism's brain generates an internal representation of the current state (e.g., from visual information) and any salient outcomes (e.g., from nociception). This information is passed to an "agent" that decides what responses or actions to emit based on the current stored state and action values and then updates these values based on the next state (Sutton and Barto, 1981). In this architecture, "pain" is the internal reinforcement signal used for learning and is distinct from the nociceptive sensing process (in the same way that reward is distinct from the sensory properties of a reinforcer; Singh et al., 2009).

From its initial demonstration (Seymour et al., 2004), there is now substantial evidence that pain controls behavior using an RL-based strategy, and that this involves a hierarchy of control processes. These are built on a basic system of innate responses that exists across species, and, working together, coordination of these control processes provides a highly effective way of minimizing future harm (Figure 2). Below, we outline the key aspects of each, and how they fit together to control pain behavior.

Innate Responses
Nociceptive stimulation produces a broad and diverse set of motor, autonomic, and behavioral defensive responses that are

stimulus specific, situation specific, and species specific (Bolles, 1971; Fanselow and Lester, 1988). Innate responses are precise, sophisticated, and rapid—driving defensive activity within a few hundred milliseconds. They are also remarkably strong, overwhelming other ongoing behavioral activity (which, as we explain later, has critical implications for the organization of endogenous control mechanisms). These features are reflected in the corresponding neural substrates within a highly complex network of spinal and brainstem connections, including dorsal horn circuits governing motor responses, brainstem autonomic nuclei, and hypothalamic-PAG circuits driving basic behavioral programs (Fanselow, 1994; Craig et al., 1998).

Pavlovian Learning
Pavlovian learning allows innate responses to be activated in advance of a harmful stimulus—offering the chance to prepare for, reduce, or even completely avoid it (Mackintosh, 1983; Bolles, 1972). Any sensory cue that reliably precedes pain can act as a predictor (a "conditioned stimulus"), and it is known that acquisition of the response (the "conditioned response") depends on prediction errors. Importantly, evidence suggests that the brain learns higher-order pain prediction errors—allowing the prediction to be transferred back in time to the earliest reliable predictor, in accordance with an RL solution to the credit assignment problem (Seymour et al., 2004). Pavlovian pain responses can be divided into two categories: pain-specific responses, which tend to be well-timed motor responses thought to be mediated by cerebellar learning (e.g., leg flexion or eye blink to foot or eyelid shock, respectively), and non-specific responses common to many aversive stimuli (such as withdrawal

Figure 2. The RL Model of Pain
This schematic details the computational architecture of the pain system. Basic brain representations of pain receive ascending spinal nociceptive input and are used to generate the internal reinforcement signal that is used for control. These feed into separate control systems that control behavior: (1) an innate pain response system; (2) state-based, Pavlovian learning; (3) associative action-outcome learning (model-free habit learning), and (4) a cognitive (model-based) action learning and planning system. Collectively, learning and responses/actions emerge through pavlovian-instrumental (P-I), and model-based-model-free interactions. Model simulation produces rapid, efficient harm avoidance, and outperforms single system models and conventional control systems for autonomous agents in terms of safe learning (Elfwing and Seymour, 2017). Endogenous control from these controllers reciprocally modulates the sensory pain pathway.

and autonomic arousal), which involve coordinated subcortical network including the amygdala, ventral putamen, ventral PAG, VTA, and dorsal raphe (Groessl et al., 2018; Herry and Johansen, 2014; Zhang et al., 2016). These dissociable systems positively interact with each other (Betts et al., 1996; Pearce et al., 1981) and negatively interact with reward learning circuits (Seymour et al., 2005; Konorski, 1948). Pavlovian predictions are sensitive to uncertainty, which enhances the learning rate and controls autonomic responses such as skin conductance putatively through an amygdala-dependent process (Li et al., 2011; Boll et al., 2013; Zhang et al., 2016). Pavlovian values also *generalize* to perceptually and conceptually similar cues, allowing pain predictions to be made to novel stimuli in an efficient way (Onat and Büchel, 2015; Dunsmoor et al., 2011; Dunsmoor and Kroes, 2019; Koban et al., 2018).

Instrumental Learning

Whereas Pavlovian learning effectively deals with state learning (that is, conditioned responses generally prepare for but don't fundamentally change the probability of pain), instrumental learning allows novel actions to be learned according to their outcomes, and so fundamentally influence the probability of their occurrence (Mackintosh, 1983). For pain, this involves escape (from persistent pain) and avoidance (of phasic pain). Current evidence suggests that the brain employs a parallel system in which action values for relief and pain are simultaneously learned and interact to guide choice (Seymour et al., 2012; Eldar et al., 2016). Indeed, there is a specific advantage to learning these two values separately: relief values can approximate the best-case scenario of future actions ("what to do"), and pain values can approximate the worst-case scenarios ("what not to do"), and the two values can in principle be integrated together to guide action ("multi-attribute RL"). Learning these two values separately in this way conserves information and allows for safer behavior (Elfwing and Seymour, 2017).

Instrumental learning involves reciprocal interactions with the Pavlovian system. Pavlovian values and prediction errors guide both the learning (e.g., two-factor theory [Maia, 2010; Moutoussis et al., 2008]) and expression (e.g., conditioned reinforcement, conditioned suppression, Pavlovian-instrumental transfer [Seymour et al., 2005; Lawson et al., 2014; Talmi et al., 2008]) of

instrumental actions. Neurobiologically, escape and avoidance are implemented in similar circuits to Pavlovian learning and specifically involve posterior putamen and amygdala circuits (Menegas et al., 2018), with connections to ventromedial prefrontal regions encoding action values themselves (Seymour et al., 2012; Roy et al., 2014). Generalization of avoidance actions involves independent process for generalization of pain-predictive (insula cortex) and relief-predictive (ventromedial prefrontal cortex [VMPFC]) action values (Norbury et al., 2018).

Cognitive Learning
In the systems described above, both Pavlovian and instrumental learning involve a process by which the brain learns a simple scalar value of a state or action and use this value to guide responses and choices, respectively, in the absence of any internal model ("model-free learning"). However, humans clearly have enormous capacity to build much more sophisticated representations of pain, events, and contexts and use them to guide deliberative behavior. Under the umbrella term of cognitive learning or "model-based" learning, these cognitive representations can encode specific states, actions, and pain and formulate an internal model of the individual and their environment (Tolman, 1948; Dayan and Daw, 2008). Such an internal model can support explicit planning, evaluation (i.e., the ability to report a pain prediction or intensity judgment), instructed and observational learning, and episodic memory.

Computationally, by encoding an internal map of the world, cognitive learning can infer the presence of "hidden states" and the structure of abstract rules, including the decision policies of other intelligent agents (Behrens et al., 2018). Accordingly, it provides a model-based mechanism that subsumes both Pavlovian and instrumental learning (Gershman et al., 2015). That is, cognitive processes are likely to be routinely involved in simple Pavlovian and instrumental tasks, allowing a more sophisticated representation of task structure than is possible than using simpler "model-free" learning algorithms (Rescorla, 1988). Naturally, however, defining precisely what algorithms are used in cognitive learning is much more difficult to ascertain, simply due to the potential for complexity (Daw and Dayan, 2014). In the case of Pavlovian conditioning, however, Bayesian models can be shown to explain several aspects of learning difficult to explain using simpler RL algorithms (Courville et al., 2006). More generally, the notion of model-free and model-based control across state and action learning scenarios is embedded in a long literature of emotional and habitual behavior, versus deliberative cognition (Daw, 2018). However, the reality may be more complex, and, at least in the case of reward, there is evidence that the brain uses intermediate computational strategies that have features of both model-free and model-based learning (Momennejad et al., 2017).

Neurobiologically, cognitive learning and decision-making structures for pain are likely to include multiple regions of prefrontal cortex, including ventrolateral and dorsolateral PFC, anterior cingulate, and hippocampal regions (Atlas et al., 2016; Olsson and Phelps, 2007; Jeon et al., 2010; Carter et al., 2006; LeDoux and Daw, 2018; Qi et al., 2018). Importantly, there is evidently overlap in many brain regions identified in cognitive learning and simpler learning schemes, especially regions such as the amygdala and striatum (Madarasz et al., 2016; Koban et al., 2017).

Conscious Pain Perception and Interactions between Controllers
In the RL framework, the "pain" signal acts as the primary teaching signal that drives the learning of values. This raises the question as to whether this reflects the conscious perception of pain or exists as a distinct, subconscious entity. The multi-controller architecture does not necessarily mean that each controller uses precisely the same reinforcement signal, and, because innate defense responses need to be rapid, they are inextricably linked to nociceptive input from the very earliest parts of the ascending pathway. However, slower and computationally sophisticated sensory processing of nociceptive signals can be used by the higher cognitive controller, which is clearly directly associated with conscious processing (i.e., working memory, explicit reasoning, and planning). Furthermore, it must be the case that conscious pain can exert a direct effect on cognitive learning and control, because, even if the conscious perception of pain emerged as some sort of epiphenomenon of neural processing, its perception would still lead to avoidance of this state in the future, given the choice. For this reason, at the very least conscious pain must act as a control signal related to cognitive learning systems.

More broadly, cognitive processes are built atop a hierarchy that involves multiple facilitative interactions between layers. Although such an architecture seems complex, computationally it is highly efficient (Piccolo et al., 2018; Lee et al., 2019). For instance, exposure to an unexpected pain stimulus recruits the innate and Pavlovian systems first, to provide rapid, safe defense using evolutionarily learned information. This sets the "action priors" upon which a cognitive system can evaluate the causes of pain based on building a model of what happened, allowing rapid learning and inference of optimal defensive responses. When these responses can lead to the reliable avoidance of pain, the basic (model-free) instrumental system takes over control, which provides computational efficiency and can help protect against unnecessary influence by random noise (Wang et al., 2018; Daw et al., 2005). Overall, each system is optimal in a given situation, and as a whole the architecture balances speed and computational efficiency at one end, with computational sophistication at the other. The caveat, however, may be a necessary susceptibility to impulsiveness and compulsiveness, mostly due to the strength of innate and Pavlovian systems (Lloyd and Dayan, 2018; Millner et al., 2018; Robbins et al., 2012).

Endogenous Control
The RL framework shows how potential harm can be minimized through a nested hierarchy of controllers. This raises the question as to whether we should expect pain to be a fixed, stable signal that faithfully represents the nociceptive signal, or a flexible signal that adapts to the current learning context. As we discuss below, the RL architecture indeed indicates that pain should be modulated by a number of factors if it is to operate optimally as a control signal.

Modulation by Sensory Inference
The available evidence suggests that the RL control hierarchy has a corresponding sensory processing hierarchy, with crude spinal and brainstem nociceptive input feeding into lower

controllers, and conscious pain feeding into higher controllers at the top (Figure 2). The function of sensory processing hierarchies is to allow the optimal estimate of the properties—such as the intensity—of the external stimulus (Seymour and Dolan, 2013; Büchel et al., 2014; Tabor et al., 2017). This estimate is ultimately an inference made based on prior experience and other relevant information, including simple predictive contingencies, multisensory integration, instructed knowledge, or observed knowledge. Based on the assumption that the incoming nociceptive input is inherently noisy, inference will improve the estimate of the true intensity. Computationally, sensory inference is typically proposed to approximate some sort of Bayesian inference (Colombo and Seriès, 2012; Knill and Pouget, 2004) and can in principle explain why pain perception is routinely biased toward prior knowledge (Colloca and Benedetti, 2006; Anchisi and Zanon, 2015; Seymour and Dolan, 2013; Wiech, 2016; Atlas et al., 2010). And it is consistent with the observation that the magnitude of this bias depends on the certainty of the prior information (Brown et al., 2008; Yoshida et al., 2013). The inferred estimate may also be asymmetrically weighted by the cost of errors (i.e., under-estimating pain may be more costly than over-estimating [Rachman and Arntz, 1991]). When errors do occur (for instance, when the discrepancy between prior and incoming nociceptive information reaches a threshold), then the information within the prior may need to be relearned, weakening its capability to bias future pain due to its increased uncertainty, but enhancing the way the information is used in the long run (Yu and Dayan, 2005; Hird et al., 2018). Although evidence suggests that in the case of pain, disconfirming sensory evidence may be relatively under-weighted (Jepma et al., 2018). Overall, however, modulation during perceptual processing creates a more accurate pain signal that is available to higher (cognitive) RL control.

Modulation by Predictive Value

A key feature of the RL model is that it deals with sequential stimuli—predicting outcomes both near and far in the future. This accommodates the fact that pain can both act as both a reinforcer (i.e., an outcome or an unconditioned stimulus, US) and a cue for other motivationally salient outcomes (a predictor, or conditioned stimulus, CS). That is, pain can predict its own termination, explicit reward, or more pain (Gerber et al., 2014; Fields, 2018; Navratilova et al., 2015)—indeed, anything that improves or worsens homeostasis (Keramati and Gutkin, 2014). This means that the aversiveness of pain incorporates *two* quantities—its inherent (aversive) value as an outcome and its value as a predictor. For instance, in Pavlovian counter-conditioning, a subject might learn to predict a food reward following a pain stimulus: after training, the pain stimulus elicits no observable aversive response, only appetitive (positive) responses anticipating food (Eroféeva, 1921). Here, it is clear that the modulation of pain must be at an early level for defensive motor responses to be lost, but also that the pain stimulus continues to act as a discriminative stimulus despite being modulated, implying that, at some level, modulation of pain must have a degree of selectivity to preserve discriminative information (Melzack and Casey, 1968).

Pain can also be a predictor for more pain, which will enhance its aversiveness and raises the issue of how the duration to the next pain influences its aversive valuation. Evidence indicates that two processes are at work—temporal discounting and dread. Temporal discounting is a well-supported assumption in most RL models by which people discount future events as a function of distance into the future (Sutton and Barto, 1998; Frederick et al., 2002). But people also find the process of anticipating pain inherently aversive in its own right—a phenomenon called dread—which often causes them to choose sooner over distant pain (e.g., for a necessary dental procedure, people might want to "get it out of the way" sooner so they don't need to worry about it) (Loewenstein, 1987; Berns et al., 2006). Hence, the net behavior is the combination of a dread function with a discount function, which leads to an "n"-shaped prospective function in which predictions of future pain have an intermediate peak aversive latency (Story et al., 2013).

It is also possible for temporal patterns of pain to act as predictors. In a classic example, decreasing pain is felt as less aversive than increasing pain—the so-called "peak-end" effect (Kahneman et al., 1993, 1997). This implies that either temporal constructs can act as a cues for associative learning, or that people build a more sophisticated internal model (memory) of the episode to support prediction (Fiser et al., 2010).

Modulation by Decision Conflict

The RL model mediates a broad array of responses and actions, from autonomic and physiological responses, reflexive motor responses (e.g., limb withdrawal), innate behavioral programs (freeze, fight, or flight), communicative responses (facial expression and vocalizations), and any type of instrumental motor action (such as pressing a keyboard in a pain experiment). Within this set, emission of some types of response can occur independently of others—for instance, pressing a key will not interfere with a heart rate response or facial expression. But other types of response will: for instance, innate motor responses may well interfere with instrumental actions that might be more important, such as escape from danger or acquisition of a large reward that outweighs the magnitude of the pain (Maier et al., 1982; Fields, 2006, 2018; Dayan et al., 2006). The problem the pain system has in managing this decision conflict is that because the innate responses are relatively hard-wired early in the ascending pain pathway, which is necessary to elicit rapid responses, the only way to suppress innate responses is to suppress nociceptive afferents when or soon after they enter the dorsal horn of the spinal cord. That is, it may not be feasible to selectively modulate innate responses without modulating ascending pain signals at the same time. This means that when instrumental decision circuits prioritize reward seeking over pain avoidance or escape, or when active instrumental avoidance or escape involves actions different from innate avoidance or escape, then pain is endogenously reduced (Dum and Herz, 1984).

Decision conflict may also operate at the level of cognition, since pain inherently drives attention, learning, and planning (Legrain et al., 2009; Van Damme et al., 2010). Thus, in the face of a competing and more significant goal (i.e., escape or large reward), pain may interfere with and disrupt more important cognitive processes (Eccleston and Crombez, 1999). Just as it is the case that two physical actions may be incompatible (such as simultaneously moving in two directions), two "mental" actions may also be incompatible (such as planning to simultaneously move in two different directions) (Brown et al., 2016). Thus,

<div style="border:1px solid #000">

Box 1. Key Predictions of the RL Model

- **Opposite effects of controllability and uncertainty on phasic and tonic pain.** The model proposes that uncertainty and controllability relate to a greater marginal benefit to learning, and phasic pain should be enhanced accordingly. However, in the case of tonic pain, its relief acts as the teaching signal, and ongoing pain has a direct suppressive effect on cognition, and so ongoing pain should be reduced to enhance relief learning. Existing support for this is mixed (Yoshida et al., 2013; Brown et al., 2008; Zaman et al., 2017; Bräscher et al., 2016; Wiech et al., 2006; Salomons et al., 2007; Zhang et al., 2018a), partly because uncertainty is not always studied in the context of controllability, and increased controllability often results in reduced uncertainty, leaving this issue to be fully demonstrated.
- **Endogenous control should drive exploration.** If a core function of endogenous control is to facilitate information acquisition, this should be reflected in choice, and manifest by a direct relationship between pain modulation and exploratory action. Furthermore, this should also be sensitive to the expected benefit of this information to future behavior—i.e., a greater opportunity to exploit information relates to greater endogenous control (cf. Wilson et al., 2014).
- **Pain discrimination is preserved during endogenous analgesia.** If pain can act as both a cue and an outcome, it is important that the capacity to discriminate pain remains unimpaired in the context of endogenous control. Although evidence suggests that descending modulatory pathways have a selective end target in the dorsal horn (Heinricher et al., 2009), there is yet no behavioral evidence, for instance, that fine discrimination (e.g., spatial or intensity) is preserved in endogenous analgesia.

</div>

decision conflict may invoke endogenous control at both the level of action and cognition.

Modulation by Informational Value

The fact that attention and controllability reliably modulate pain (Eccleston and Crombez, 1999; Wiech et al., 2006; Yoshida et al., 2013; Salomons et al., 2007, 2015; Taylor et al., 2017; Bräscher et al., 2016), beyond that which can be explained by mechanisms above, suggests that factors intrinsic to learning and control specifically modulate pain. Although the goal of RL is to learn to minimize pain as an objective function, performance can be enhanced by intrinsically modulating pain according to its informational value in learning. This is because the prospective benefit of learning is not a fixed quantity but varies according to how much there is to learn (uncertainty), how long there is to exploit learnable information (opportunity), and the capability to exploit it (controllability). In the case of reward, such intrinsic modulation of decision value is well recognized (for instance, in novelty seeking and uncertainty seeking) and helps solve the exploration-exploitation problem of trial-and-error learning (i.e., information sampling [Wilson et al., 2014; Wittmann et al., 2008]).

In the case of pain, therefore, the magnitude of a phasic pain stimulus should be enhanced if uncertainty, opportunity, and controllability are high, because the marginal benefit of learning is higher. More precisely, the model predicts these factors should interact, because the benefit of learning is only manifest if the opportunity and controllability are both significant (Zhang et al., 2018b) (see Box 1). In the case of learning relief from tonic pain, the opposite effect should occur (i.e., background pain should be reduced if the benefit to learning about relief is high), because the object of learning is relief, not pain (and persistent pain exerts a tonic control effect on behavior, as we discuss below). This appears to be the case—relief uncertainty reduces background tonic pain when relief uncertainty is high, when relief is controllable (Zhang et al., 2018a).

Across these demonstrations, uncertainty-based modulation reflects the mechanism underlying what is conventionally considered attention or salience (Eccleston and Crombez, 1999). This spans attention that is driven by bottom-up processes learned through trial and error (i.e., lots of errors equate

to high uncertainty), or top-down processes provided by external cues or instruction. In all cases, the effect is to enhance learning and guide choice in a way that benefits long-run prospects. Thus, any relatively unexpected change in persistent or repetitive pain will have a modulatory effect: reductions in persistent nociceptive stimulation will cause an exaggerated reduction in pain perception, and increases in nociceptive stimulation will cause an exaggerated increase in perceived pain. These effects are well recognized in studies of relative valuation of pain (Winston et al., 2014; Vlaev et al., 2009) and offset hypoalgesia and onset hyperalgesia (Sprenger et al., 2018; Grill and Coghill, 2002; Yelle et al., 2009).

Neural Implementation of Endogenous Analgesia

The primary effector pathway for endogenous control (both hypo- and hyperalgesia) is known to involve descending control via the PAG to rostral ventral medulla, to the dorsal horn of the spinal cord (Heinricher et al., 2009). What has been harder to ascertain is which higher brain sites instruct this pathway, and where and how the amount of descending control is computed. One of the difficulties is that many classic paradigms of endogenous control may actually involve several distinct mechanisms, so it is difficult to relate computational mechanisms to specific neural loci without a considerable degree of uncertainty. For instance, placebo analgesia can involve all four of the above mechanisms. However, several cortical regions seem to play a key role, including regions of anterior cingulate cortex, dorsolateral prefrontal cortex, and insula (Wiech, 2016; Tracey, 2010). More specifically, the pregenual anterior cingulate cortex (pgACC) has emerged as the most consistently implicated cortical region in human endogenous control paradigms, including in placebo and expectancy hypoalgesia (Wager et al., 2004; Bingel et al., 2006; Eippert et al., 2009), uncertainty-based analgesia (Zhang et al., 2018a, 2018b), controllability (Salomons et al., 2007, 2015), habituation (Bingel et al., 2007), stress-induced analgesia (Yilmaz et al., 2010), and even analgesia induced by motor cortex stimulation (Peyron et al., 2007). The pgACC is highly opioid rich and sits within an anatomical network with connections to PAG and other subcortical regions involved in pain and learning, including amygdala, VMPFC, hippocampus,

lateral orbitofrontal cortex (OFC), and PFC (Margulies et al., 2007; Vogt, 2005). These sites are central to pain and reward learning and directly implicated in control by decision conflict (Fields, 2018), sensory inference (Büchel et al., 2014), and value learning (Craig, 2003; Seymour et al., 2004; Ploghaus et al., 1999). Furthermore, the pgACC is also implicated in both rodent models and human clinical cases of chronic pain (Qu et al., 2011; Segerdahl et al., 2018; Mano et al., 2018).

What has been less clear is the specificity of modulation of ascending pathways in the dorsal horn. An inherent paradox of endogenous control paradox is that it risks degrading the information that pain carries as a predictive stimulus. Hence, it is likely that at least some aspects of discriminitive information must be selectively preserved in endogeonous control in the ascending pathways (Box 1). Indeed, psychologically, preservation of discriminative perception accompanying analgesia with opioids and cingulotomy is well described (Melzack and Casey, 1968) and forms the basis of a conventional notion of dissociability of affective and discriminative pain processing in putative "medial and lateral pain systems," respectively (Vogt and Sikes, 2000; Corder et al., 2019). Recent evidence indicates that this selectivity may be mediated by preferential control of C-fibers over A-delta fibers in the dorsal horn (Heinricher et al., 2009), which fits with the notion that the A-delta fiber pathways carry more refined discriminative information.

Translation to Chronic Pain

Tonic or persistent pain after injury serves several physiological functions. First, it has a direct effect on mood and cognition, encouraging rest and recuperation by reducing motivation to engage in non-essential reward-guided activity, which is less important for homeostatic priorities. Second, it represents a state from which reduction or cessation of pain becomes a new motivational goal and hence frames relief as an objective function for appetitive RL. Third, when accompanied by hyperalgesia and allodynia, it sensitizes otherwise less or non-noxious stimuli to drive pain learning, which is clearly adaptive given that the area of injury may be more prone to further injury than normal.

This raises the question as to why physiologically persistent pain outlives its usefulness to become pathologically persistent pain in some individuals. Clearly, chronic pain is heterogeneous, and many forms of chronic pain could simply reflect a normal brain response to increased nociceptive input at peripheral or spinal levels. More commonly, however, it is likely that peripheral and central factors interact to generate the chronic pain state: persistent nociceptive signals are further amplified and maintained by aberrant brain processes.

The RL model of pain illustrates many specific computational mechanisms that could hypothetically contribute to this process. These could be perceptual: with persistent pain reflecting an inference of the state of injury but subject to an excessive and irrefutable belief that this is the case biasing perceptual inference (a sort of "self-fulfilling prophecy" [Jepma et al., 2018]). Or it could be motivational: such as excessive aversive valuation, elevated or asymmetrical aversive learning rates, over-generalization, loss of any component of endogenous control, reduced extinction of movement-related fear, excessive dread, aberrant

sequential learning, and so on. This may have either a direct effect on pain perception or an indirect effect: in the fear-avoidance model of musculoskeletal pain, excessive fear learning of movement leads to inactivity, which itself causes increased tissue injury through secondary means (Vlaeyen and Linton, 2000).

From a computational perspective, therefore, it is likely that individual factors (i.e., the parameters of the RL model) act as individual risk factors for chronic pain, that subsequently interact to generate the chronic pain phenotype given appropriate external events (such as a precipitating tissue injury). This would define a "computome" of chronic pain risk and illustrates many ways in which some of the individual factors might be shared with co-incident psychiatric conditions, such as depression and anxiety (Figure 3). This framework also offers a computational framework to start to address some of the neurobiological differences seen in RL-linked circuits in chronic pain patients, including VMPFC and nucleus accumbens (Baliki et al., 2010, 2012; Mano et al., 2018). However, given the complexity of the RL model (in terms of its architecture and large number of parameters), it strongly appeals to simulation methods to help predict how different factors might conspire together to generate chronic pain risk (Seymour and Lee, 2019).

Conclusions

Both theory and evidence point to a view of pain as a precision signal that guides prospective behavior to minimize harm through learning. The pain system has been shaped through evolution by the complexity and diversity of actual threats in the natural world, but, in particular, it has faced four problems that have had a dominant impact on its architecture. The first is how to learn about harm both near and far into the future (the credit assignment problem), which is solved by the predictive value learning system defined by RL. The second is how to balance speed of response with processing sophistication (a type of speed-accuracy dilemma), which is solved by having a nested hierarchical architecture that spans rapid reflexes to internal models, and which interact through endogenous control. The third is how to balance information acquisition about threat with the concurrent need to avoid it (the information sampling dilemma), which is solved by endogenous fine-tuning of pain to maximize its value as a learning signal. And the fourth is how to suppress pain when needed, while not suppressing the information it carries, which is solved by having dissociable discriminative and affective subcomponents of pain. Overall, the RL model of the pain system illustrates computationally how these solutions are implemented in the brain, and how this drives safe, efficient, rapid, and effective pain behavior.

The model also offers insight into the three broad issues in pain neuroscience raised in the introduction. From the perspective of the representation of pain in the brain (the "pain matrix"), it is clear that pain is constructed not only from nociceptive input, but also from a set of cortical and subcortical components that compute the effective magnitude of pain as a control signal. That this implies that subjective pain will be best estimated from responses in multiple regions is consistent with general network and connectivity (Kucyi and Davis, 2015) and multivariate "signatures" of pain (Wager et al., 2013; Woo et al., 2015;
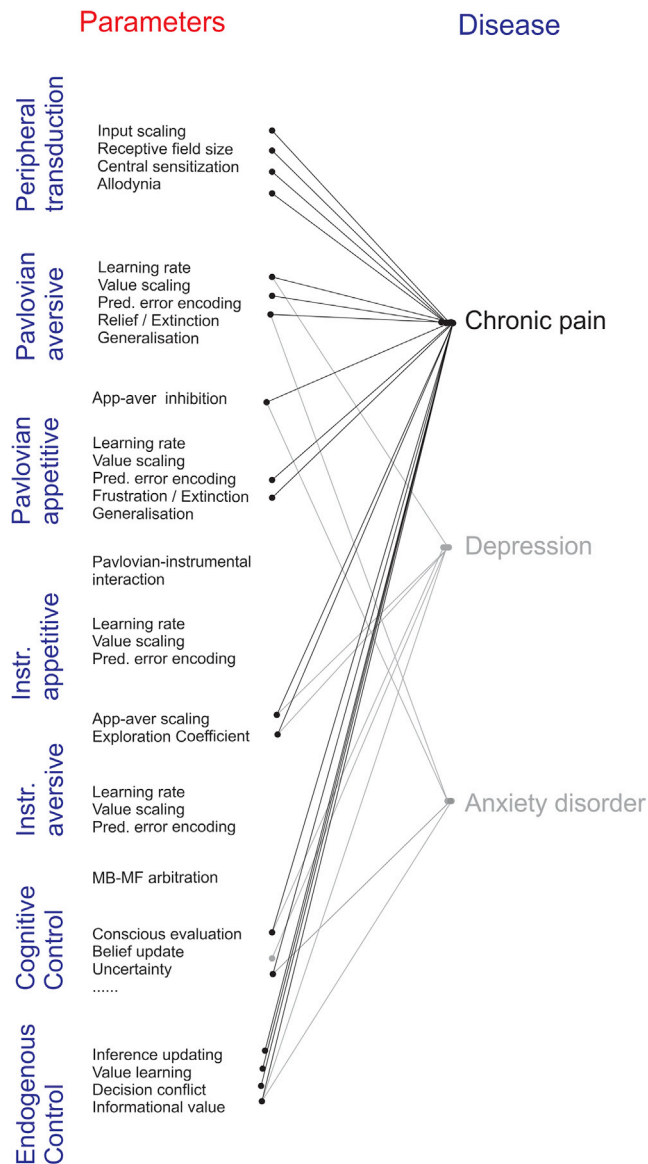
Figure 3. The Hypothetical Chronic Pain Computome
The RL model involves a large set of parameters that determine the way in which learning and decision making operate within the control framework. Several of these have been proposed to be involved with not only chronic pain, but also other disorders associated with aversive learning, including anxiety and depression. The figure shows a schematic of how a series of hypothetical factors might operate together to create an overall risk for chronic pain, given an appropriate peripheral drive from an injury, and these factors might have different roles at different points in the pain chronification process. Importantly, the way these factors interact is determined by the complexity of the computational (RL) model, which defines the information-processing operations they govern, which means predicting how different combinations determine risk may require advanced simulation platforms (Seymour and Lee, 2019).

Marquand et al., 2010; Zunhammer et al., 2018) but goes beyond these by highlighting the importance of understanding exactly what each of the nodes in the pain network do (i.e., pain as a computational network). However, it is also consistent with cortical specificity models, because as long as endogenous

modulation is primarily descending, there should still be a restricted cortical response that primarily reflects pain aversive intensity after modulation (for example, in posterior insula or mid-anterior cingulate cortex [Segerdahl et al., 2015; Kragel et al., 2018; Craig and Craig, 2009]). In other words, although there is always the most information available from a broad set of brain regions (involving processes that are not individually unique to pain), it must also be the case that unique and fundamental representations of discriminative features and pain value are bound together to yield the unique subjective experience of pain. However, key questions remain, and perhaps the most important is knowing where in the brain internal representations of pain used for cognitive planning and control are coded (i.e., where is the "cognitive map" of pain?).

From the perspective of the subjectivity of pain, the RL model challenges the primacy of self-report. This is because at a fundamental level pain concerns control, and so control behavior should serve as the ultimate measure of pain. Irrespective of the problems associated with self-report scales (Stewart et al., 2005), pain leads to a broad set of learning and control behaviors that can be objectively measured, and the conscious perception of pain merely serves these functions. This speaks to Melzack's and Casey's man-in-the-brain problem—pain perception is merely a link (albeit it a critical one) in a self-organizing control circuit, rather than a terminal node that informs an elusive higher controller. On this basis, the RL model's prediction is that all aspects of endogenous control should be reflected in subsequent choice behavior, and this remains an important prediction for future studies (see Box 1).

Finally, the model yields a concept of pain as a signal that is tuned precisely to its function as a control signal. The concept that pain is modulated presupposes that pain is a sensory nociceptive signal whose primary role is to retrospectively estimate the objective intensity of a stimulus and then needs to be tuned to support whatever behavior is required at the time, i.e., the view of modulation of pain as post-perceptual processing. But as a prospective control signal, pain behaves in precisely the way it needs, and hence should not be considered to be modulated at all. Although the factors that lead to the tuning of pain may be difficult to experimentally observe and objectify (such as complex social information about forthcoming pain), this doesn't mean that the brain doesn't estimate and represent these quantities in a precise manner. Ultimately, endogenous control illustrates how pain is constructed on a moment-by-moment basis based on an complex but objectively definable integration of broad sources of information.

REFERENCES

Anchisi, D., and Zanon, M. (2015). A Bayesian perspective on sensory and cognitive integration in pain perception and placebo analgesia. PloS One 10, e0117270.

Atlas, L.Y., Bolger, N., Lindquist, M.A., and Wager, T.D. (2010). Brain mediators of predictive cue effects on perceived pain. J. Neurosci. *30*, 12964–12977.

Atlas, L.Y., Doll, B.B., Li, J., Daw, N.D., and Phelps, E.A. (2016). Instructed knowledge shapes feedback-driven aversive learning in striatum and orbitofrontal cortex, but not the amygdala. eLife *5*, 5.

Baliki, M.N., Geha, P.Y., Fields, H.L., and Apkarian, A.V. (2010). Predicting value of pain and analgesia: nucleus accumbens response to noxious stimuli changes in the presence of chronic pain. Neuron *66*, 149–160.

Baliki, M.N., Petre, B., Torbey, S., Herrmann, K.M., Huang, L., Schnitzer, T.J., Fields, H.L., and Apkarian, A.V. (2012). Corticostriatal functional connectivity predicts transition to chronic back pain. Nat. Neurosci. *15*, 1117–1119.

Basbaum, A.I., and Fields, H.L. (1984). Endogenous pain control systems: brainstem spinal pathways and endorphin circuitry. Annu. Rev. Neurosci. *7*, 309–338.

Behrens, T.E.J., Muller, T.H., Whittington, J.C.R., Mark, S., Baram, A.B., Stachenfeld, K.L., and Kurth-Nelson, Z. (2018). What is a cognitive map? Organising knowledge for flexible behaviour. Neuron *100*, 490–509.

Bellman, R. (2013). Dynamic Programming (Courier Corporation).

Berns, G.S., Chappelow, J., Cekic, M., Zink, C.F., Pagnoni, G., and Martin-Skurski, M.E. (2006). Neurobiological substrates of dread. Science *312*, 754–758.

Betts, S.L., Brandon, S.E., and Wagner, A.R. (1996). Dissociation of the blocking of conditioned eyeblink and conditioned fear following a shift in US locus. Anim. Learn. Behav. *24*, 459–470.

Bingel, U., Lorenz, J., Schoell, E., Weiller, C., and Büchel, C. (2006). Mechanisms of placebo analgesia: rACC recruitment of a subcortical antinociceptive network. Pain *120*, 8–15.

Bingel, U., Schoell, E., Herken, W., Büchel, C., and May, A. (2007). Habituation to painful stimulation involves the antinociceptive system. Pain *131*, 21–30.

Boll, S., Gamer, M., Gluth, S., Finsterbusch, J., and Büchel, C. (2013). Separate amygdala subregions signal surprise and predictiveness during associative fear learning in humans. Eur. J. Neurosci. *37*, 758–767.

Bolles, R.C. (1971). Species-specific defense reactions. In Aversive Conditioning and Learning, R.F. Brush, ed. (Elsevier), pp. 183–233.

Bolles, R.C. (1972). The avoidance learning problem. Psychol. Learn. Motiv. *6*, 97–145.

Bräscher, A.-K., Becker, S., Hoeppli, M.E., and Schweinhardt, P. (2016). Different brain circuitries mediating controllable and uncontrollable pain. J. Neurosci. *36*, 5013–5025.

Brown, C.A., Seymour, B., Boyle, Y., El-Deredy, W., and Jones, A.K. (2008). Modulation of pain ratings by expectation and uncertainty: Behavioral characteristics and anticipatory neural correlates. Pain *135*, 240–250.

Brown, T.I., Carr, V.A., LaRocque, K.F., Favila, S.E., Gordon, A.M., Bowles, B., Bailenson, J.N., and Wagner, A.D. (2016). Prospective representation of navigational goals in the human hippocampus. Science *352*, 1323–1326.

Büchel, C., Geuter, S., Sprenger, C., and Eippert, F. (2014). Placebo analgesia: a predictive coding perspective. Neuron *81*, 1223–1239.

Carter, R.M., O'Doherty, J.P., Seymour, B., Koch, C., and Dolan, R.J. (2006). Contingency awareness in human aversive conditioning involves the middle frontal gyrus. Neuroimage *29*, 1007–1012.

Colloca, L., and Benedetti, F. (2006). How prior experience shapes placebo analgesia. Pain *124*, 126–133.

Colombo, M., and Seriès, P. (2012). Bayes in the brain—on Bayesian modelling in neuroscience. Br. J. Philos. Sci. *63*, 697–723.

Corder, G., Ahanonu, B., Grewe, B.F., Wang, D., Schnitzer, M.J., and Scherrer, G. (2019). An amygdalar neural ensemble that encodes the unpleasantness of pain. Science *363*, 276–281.

Courville, A.C., Daw, N.D., and Touretzky, D.S. (2006). Bayesian theories of conditioning in a changing world. Trends Cogn. Sci. *10*, 294–300.

Craig, A.D. (2002). How do you feel? Interoception: the sense of the physiological condition of the body. Nat. Rev. Neurosci *3*, 655.

Craig, A.D. (2003). A new view of pain as a homeostatic emotion. Trends Neurosci. *26*, 303–307.

Craig, A.D., and Craig, A. (2009). How do you feel—now? The anterior insula and human awareness. Nat. Rev. Neurosci *10*, 59–70.

Craig, A., Bowsher, D., Tasker, R.R., Lenz, F., Dougherty, P.M., and Wiesenfeld-Hallin, Z. (1998). A new version of the thalamic disinhibition hypothesis of central pain. Pain Forum *7*, 1–28.

Daw, N.D. (2018). Are we of two minds? Nat. Neurosci *21*, 1497.

Daw, N.D., and Dayan, P. (2014). The algorithmic anatomy of model-based evaluation. Phil. Trans. R. Soc. B: Biol. Sci. *369*, 1655.

Daw, N.D., Niv, Y., and Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat. Neurosci *8*, 1704.

Dayan, P., and Daw, N.D. (2008). Decision theory, reinforcement learning, and the brain. Cogn. Affect. Behav. Neurosci. *8*, 429–453.

Dayan, P., Niv, Y., Seymour, B., and Daw, N.D. (2006). The misbehavior of value and the discipline of the will. Neural Netw. *19*, 1153–1160.

Dum, J., and Herz, A. (1984). Endorphinergic modulation of neural reward systems indicated by behavioral changes. Pharmacol. Biochem. Behav. *21*, 259–266.

Dunsmoor, J.E., and Kroes, M.C. (2019). Episodic memory and Pavlovian conditioning: ships passing in the night. Curr. Opin. Behav. Sci. *26*, 32–39.

Dunsmoor, J.E., White, A.J., and LaBar, K.S. (2011). Conceptual similarity promotes generalization of higher order fear learning. Learn. Mem. *18*, 156–160.

Eccleston, C., and Crombez, G. (1999). Pain demands attention: A cognitive-affective model of the interruptive function of pain. Psychol. Bull. *125*, 356.

Eippert, F., Bingel, U., Schoell, E.D., Yacubian, J., Klinger, R., Lorenz, J., and Büchel, C. (2009). Activation of the opioidergic descending pain control system underlies placebo analgesia. Neuron *63*, 533–543.

Eldar, E., Hauser, T.U., Dayan, P., and Dolan, R.J. (2016). Striatal structure and function predict individual biases in learning to avoid pain. Proc. Natl. Acad. Sci. USA *113*, 4812–4817.

Elfwing, S., and Seymour, B. (2017). Parallel reward and punishment control in humans and robots: safe reinforcement learning using the MaxPain algorithm. In 7th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics, ICDL-EpiRob 2017 (IEEE), pp. 140–147.

Eroféeva, M.N. (1921). Further observations upon conditioned reflexes to nocuous stimuli. Bulletin of the Institute of Lesgaft *3*.

Fanselow, M.S. (1994). Neural organization of the defensive behavior system responsible for fear. Psychon. Bull. Rev. *1*, 429–438.

Fanselow, M.S., and Lester, L.S. (1988). A functional behavioristic approach to aversively motivated behavior: Predatory imminence as a determinant of the topography of defensive behavior. In Evolution and Learning, R.C. Bolles and M.D. Beecher, eds. (Lawrence Erlbaum Associates), pp. 185–212.

Fields, H.L. (2006). A motivation-decision model of pain: the role of opioids. In Proceedings of the 11th World Congress on Pain, H. Flor, ed., pp. 449–459.

Fields, H.L. (2018). How expectations influence pain. Pain *159* (*Suppl 1*), S3–S10.

Fiser, J., Berkes, P., Orbán, G., and Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. Trends Cogn. Sci. *14*, 119–130.

Frederick, S., Loewenstein, G., and O'donoghue, T. (2002). Time discounting and time preference: A critical review. J. Econ. Lit. *40*, 351–401.

Friston, K. (2010). The free-energy principle: a unified brain theory? Nat. Rev. Neurosci *11*, 127.

Gerber, B., Yarali, A., Diegelmann, S., Wotjak, C.T., Pauli, P., and Fendt, M. (2014). Pain-relief learning in flies, rats, and man: basic research and applied perspectives. Learn. Mem. 21, 232–252.

Gershman, S.J., Norman, K.A., and Niv, Y. (2015). Discovering latent causes in reinforcement learning. Curr. Opin. Behav. Sci. 5, 43–50.

Geuter, S., Boll, S., Eippert, F., and Büchel, C. (2017). Functional dissociation of stimulus intensity encoding and predictive coding of pain in the insula. eLife 6, https://doi.org/10.7554/eLife.24770.

Grahl, A., Onat, S., and Büchel, C. (2018). The periaqueductal gray and Bayesian integration in placebo analgesia. eLife 7, e32930.

Grill, J.D., and Coghill, R.C. (2002). Transient analgesia evoked by noxious stimulus offset. J. Neurophysiol. 87, 2205–2208.

Groessl, F., Munsch, T., Meis, S., Griessner, J., Kaczanowska, J., Pliota, P., Kargl, D., Badurek, S., Kraitsy, K., Rassoulpour, A., et al. (2018). Dorsal tegmental dopamine neurons gate associative learning of fear. Nat. Neurosci 21, 952.

Heinricher, M.M., Tavares, I., Leith, J.L., and Lumb, B.M. (2009). Descending control of nociception: Specificity, recruitment and plasticity. Brain Res. Brain Res. Rev. 60, 214–225.

Herry, C., and Johansen, J.P. (2014). Encoding of fear learning and memory in distributed neuronal circuits. Nat. Neurosci. 17, 1644–1654.

Hird, E.J., Charalambous, C., El-Deredy, W., Jones, A.K., and Talmi, D. (2018). Boundary effects of expectation in human pain perception. bioRxiv, 467738.

Jeon, D., Kim, S., Chetana, M., Jo, D., Ruley, H.E., Lin, S.Y., Rabah, D., Kinet, J.P., and Shin, H.S. (2010). Observational fear learning involves affective pain system and Ca v 1.2 Ca 2+ channels in ACC. Nat. Neurosci 13, 482–488.

Jepma, M., Koban, L., van Doorn, J., Jones, M., and Wager, T.D. (2018). Behavioural and neural evidence for self-reinforcing expectancy effects on pain. Nat. Hum. Behav 2, 838–855.

Johansen, J.P., Fields, H.L., and Manning, B.H. (2001). The affective component of pain in rodents: direct evidence for a contribution of the anterior cingulate cortex. Proc. Natl. Acad. Sci. USA 98, 8077–8082.

Jones, A.K., Brown, W.D., Friston, K.J., Qi, L.Y., and Frackowiak, R.S. (1991). Cortical and subcortical localization of response to pain in man using positron emission tomography. Proc. Biol. Sci. 244, 39–44.

Kahneman, D., Fredrickson, B.L., Schreiber, C.A., and Redelmeier, D.A. (1993). When more pain is preferred to less: Adding a better end. Psychol. Sci. 4, 401–405.

Kahneman, D., Wakker, P.P., and Sarin, R. (1997). Back to Bentham? Explorations of experienced utility. Q. J. Econ. 112, 375–406.

Keramati, M., and Gutkin, B. (2014). Homeostatic reinforcement learning for integrating reward collection and physiological stability. eLife 3, e04811.

Knill, D.C., and Pouget, A. (2004). The Bayesian brain: the role of uncertainty in neural coding and computation. Trends Neurosci. 27, 712–719.

Koban, L., Jepma, M., Geuter, S., and Wager, T.D. (2017). What's in a word? How instructions, suggestions, and social information change pain and emotion. Neurosci. Biobehav. Rev. 81 (Pt A), 29–42.

Koban, L., Kusko, D., and Wager, T.D. (2018). Generalization of learned pain modulation depends on explicit learning. Acta Psychol. (Amst.) 184, 75–84.

Konorski, J. (1948). Conditioned reflexes and neuron organization (CUP Archive).

Kragel, P.A., Kano, M., Van Oudenhove, L., Ly, H.G., Dupont, P., Rubio, A., Delon-Martin, C., Bonaz, B.L., Manuck, S.B., Gianaros, P.J., et al. (2018). Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. Nat. Neurosci 21, 283.

Kucyi, A., and Davis, K.D. (2015). The dynamic pain connectome. Trends Neurosci. 38, 86–95.

Lawson, R.P., Seymour, B., Loh, E., Lutti, A., Dolan, R.J., Dayan, P., Weiskopf, N., and Roiser, J.P. (2014). The habenula encodes negative motivational value associated with primary punishment in humans. Proc. Natl. Acad. Sci. USA 111, 11858–11863.

LeDoux, J., and Daw, N.D. (2018). Surviving threats: neural circuit and computational implications of a new taxonomy of defensive behaviour. Nat. Rev. Neurosci. 19, 269–282.

Lee, J.H., Seymour, B., Leibo, J.Z., An, S.J., and Lee, S.W. (2019). Toward high-performance, memory-efficient, and fast reinforcement learning—Lessons from decision neuroscience. Science Robotics 4.

Legrain, V., Damme, S.V., Eccleston, C., Davis, K.D., Seminowicz, D.A., and Crombez, G. (2009). A neurocognitive model of attention to pain: behavioral and neuroimaging evidence. Pain 144, 230–232.

Li, J., Schiller, D., Schoenbaum, G., Phelps, E.A., and Daw, N.D. (2011). Differential roles of human striatum and amygdala in associative learning. Nat. Neurosci. 14, 1250–1252.

Lloyd, K., and Dayan, P. (2018). Pavlovian-instrumental interactions in active avoidance: The bark of neutral trials. Brain Res. Published online October 9, 2018. https://doi.org/10.1016/j.brainres.2018.10.011.

Loewenstein, G. (1987). Anticipation and the valuation of delayed consumption. Econ. J. (Lond.) 97, 666–684.

Mackintosh, N.J. (1983). Conditioning and Associative Learning (Clarendon Press Oxford).

Madarasz, T.J., Diaz-Mataix, L., Akhand, O., Ycu, E.A., LeDoux, J.E., and Johansen, J.P. (2016). Evaluation of ambiguous associations in the amygdala by learning the structure of the environment. Nat. Neurosci 19, 965.

Maia, T.V. (2010). Two-factor theory, the actor-critic model, and conditioned avoidance. Learn. Behav. 38, 50–67.

Maier, S.F., Drugan, R.C., and Grau, J.W. (1982). Controllability, coping behavior, and stress-induced analgesia in the rat. Pain 12, 47–56.

Mano, H., Kotecha, G., Leibnitz, K., Matsubara, T., Sprenger, C., Nakae, A., Shenker, N., Shibata, M., Voon, V., Yoshida, W., et al. (2018). Classification and characterisation of brain network changes in chronic back pain: A multi-center study. Wellcome Open Res. 3, 19.

Margulies, D.S., Kelly, A.M., Uddin, L.Q., Biswal, B.B., Castellanos, F.X., and Milham, M.P. (2007). Mapping the functional connectivity of anterior cingulate cortex. Neuroimage 37, 579–588.

Marquand, A., Howard, M., Brammer, M., Chu, C., Coen, S., and Mourão-Miranda, J. (2010). Quantitative prediction of subjective pain intensity from whole-brain fMRI data using Gaussian processes. Neuroimage 49, 2178–2189.

Melzack, R., and Casey, K.L. (1968). Sensory, motivational, and central control determinants of pain. In International Symposium on the Skin Senses, D. Kenshalo, ed. (C.C. Thomas), pp. 423–435.

Melzack, R., and Wall, P.D. (1965). Pain mechanisms: a new theory. Science 150, 971–979.

Menegas, W., Akiti, K., Amo, R., Uchida, N., and Watabe-Uchida, M. (2018). Dopamine neurons projecting to the posterior striatum reinforce avoidance of threatening stimuli. Nat. Neurosci. 21, 1421–1430.

Millner, A.J., Gershman, S.J., Nock, M.K., and den Ouden, H.E.M. (2018). Pavlovian control of escape and avoidance. J. Cogn. Neurosci 30, 1379–1390.

Momennejad, I., Russek, E.M., Cheong, J.H., Botvinick, M.H., Daw, N., and Gershman, S.J. (2017). The successor representation in human reinforcement learning. Nat. Hum. Behav 1, 680.

Morton, D.L., El-Deredy, W., Watson, A., and Jones, A.K. (2010). Placebo analgesia as a case of a cognitive style driven by prior expectation. Brain Res. 1359, 137–141.

Moutoussis, M., Bentall, R.P., Williams, J., and Dayan, P. (2008). A temporal difference account of avoidance learning. Network 19, 137–160.

Navratilova, E., Xie, J.Y., Okun, A., Qu, C., Eyde, N., Ci, S., Ossipov, M.H., King, T., Fields, H.L., and Porreca, F. (2012). Pain relief produces negative reinforcement through activation of mesolimbic reward-valuation circuitry. Proc. Natl. Acad. Sci. USA 109, 20709–20713.

Navratilova, E., Atcherley, C.W., and Porreca, F. (2015). Brain circuits encoding reward from pain relief. Trends Neurosci. *38*, 741–750.

Norbury, A., Robbins, T.W., and Seymour, B. (2018). Value generalization in human avoidance learning. eLife *7*, e34779.

Olsson, A., and Phelps, E.A. (2007). Social learning of fear. Nat. Neurosci *10*, 1095.

Onat, S., and Büchel, C. (2015). The neuronal basis of fear generalization in humans. Nat. Neurosci. *18*, 1811–1818.

Ongaro, G., and Kaptchuk, T.J. (2019). Symptom perception, placebo effects, and the Bayesian brain. Pain *160*, 1–4.

Ozawa, T., Ycu, E.A., Kumar, A., Yeh, L.F., Ahmed, T., Koivumaa, J., and Johansen, J.P. (2017). A feedback neural circuit for calibrating aversive memory strength. Nat. Neurosci *20*, 90.

Pearce, J.M., Montgomery, A., and Dickinson, A. (1981). Contralateral transfer of inhibitory and excitatory eyelid conditioning in the rabbit. Q. J. Exp. Psychol. Sec. B *33*, 45–61.

Peyron, R., Faillenot, I., Mertens, P., Laurent, B., and Garcia-Larrea, L. (2007). Motor cortex stimulation in neuropathic pain. Correlations between analgesic effect and hemodynamic changes in the brain. A PET study. Neuroimage *34*, 310–321.

Pezzulo, G., Rigoli, F., and Friston, K. (2015). Active inference, homeostatic regulation and adap-tive behavioural control. Prog. Neurobiol. *134*, 17–35.

Piccolo, L., Libera, F., Bonarini, A., Seymour, B., and Ishiguro, H. (2018). Pain and self-preservation in autonomous robots: From neurobiological models to psychiatric disease. In 7th Joint IEEE International Conference on Development and Learning and on Epigenetic Robotics, ICDL-EpiRob 2017, pp. 263–270.

Ploghaus, A., Tracey, I., Gati, J.S., Clare, S., Menon, R.S., Matthews, P.M., and Rawlins, J.N. (1999). Dissociating pain from its anticipation in the human brain. Science *284*, 1979–1981.

Qi, S., Hassabis, D., Sun, J., Guo, F., Daw, N., and Mobbs, D. (2018). How cognitive and reactive fear circuits optimize escape decisions in humans. Proc. Natl. Acad. Sci. USA *115*, 3186–3191.

Qu, C., King, T., Okun, A., Lai, J., Fields, H.L., and Porreca, F. (2011). Lesion of the rostral anterior cingulate cortex eliminates the aversiveness of spontaneous neuropathic pain following partial or complete axotomy. Pain *152*, 1641–1648.

Rachman, S., and Arntz, A. (1991). The overprediction and underprediction of pain. Clin. Psychol. Rev. *11*, 339–355.

Rescorla, R.A. (1988). Pavlovian conditioning: It's not what you think it is. Am. Psychol *43*, 151.

Rescorla, R.A., and Wagner, A.R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In Classical Conditioning II: Current Research and Theory, A.H. Black and W.F. Prokasy, eds. (Appleton-Century-Crofts), pp. 64–99.

Robbins, T.W., Gillan, C.M., Smith, D.G., de Wit, S., and Ersche, K.D. (2012). Neurocognitive endophenotypes of impulsivity and compulsivity: towards dimensional psychiatry. Trends Cogn. Sci. *16*, 81–91.

Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G.E., and Wager, T.D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. Nat. Neurosci. *17*, 1607–1612.

Salomons, T.V., Johnstone, T., Backonja, M.M., Shackman, A.J., and Davidson, R.J. (2007). Individual differences in the effects of perceived controllability on pain perception: critical role of the prefrontal cortex. J. Cogn. Neurosci. *19*, 993–1003.

Salomons, T.V., Nusslock, R., Detloff, A., Johnstone, T., and Davidson, R.J. (2015). Neural emotion regulation circuitry underlying anxiolytic effects of perceived control over pain. J. Cogn. Neurosci. *27*, 222–233.

Segerdahl, A.R., Mezue, M., Okell, T.W., Farrar, J.T., and Tracey, I. (2015). The dorsal posterior insula subserves a fundamental role in human pain. Nat. Neurosci *18*, 499–500.

Segerdahl, A.R., Themistocleous, A.C., Fido, D., Bennett, D.L., and Tracey, I. (2018). A brain-based pain facilitation mechanism contributes to painful diabetic polyneuropathy. Brain *141*, 357–364.

Seymour, B., and Dolan, R. (2013). Emotion, motivation, and pain. In Wall & Melzack's Textbook of Pain, S. McMahon, ed. (Elsevier Health Sciences), pp. 248–255.

Seymour, B., and Lee, S.W. (2019). Decision-making in brains and robots: the case for an interdisciplinary approach. Curr. Opin. Behav. Sci. *26*, 137–145.

Seymour, B., O'Doherty, J.P., Dayan, P., Koltzenburg, M., Jones, A.K., Dolan, R.J., Friston, K.J., and Frackowiak, R.S. (2004). Temporal difference models describe higher-order learning in humans. Nature *429*, 664–667.

Seymour, B., O'Doherty, J.P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., and Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. Nat. Neurosci. *8*, 1234–1240.

Seymour, B., Daw, N.D., Roiser, J.P., Dayan, P., and Dolan, R. (2012). Serotonin selectively modulates reward value in human decision-making. J. Neurosci. *32*, 5833–5842.

Singh, S., Lewis, R.L., and Barto, A.G. (2009). Where do rewards come from. In Proceedings of the Annual Conference of the Cognitive Science Society, pp. 2601–2606.

Sprenger, C., Stenmans, P., Tinnermann, A., and Büchel, C. (2018). Evidence for a spinal involvement in temporal pain contrast enhancement. Neuroimage *183*, 788–799.

Stewart, N., Brown, G.D., and Chater, N. (2005). Absolute identification by relative judgment. Psychol. Rev. *112*, 881.

Story, G.W., Vlaev, I., Seymour, B., Winston, J.S., Darzi, A., and Dolan, R.J. (2013). Dread and the disvalue of future pain. PLoS Comput. Biol *9*, e1003335.

Sutton, R.S., and Barto, A.G. (1981). Toward a modern theory of adaptive networks: expectation and prediction. Psychol. Rev. *88*, 135.

Sutton, R.S., and Barto, A.G. (1998). Reinforcement learning: An introduction, *Vol. 1. 1* (Cambridge: MIT press).

Tabor, A., and Burr, C. (2019). Bayesian Learning Models of Pain: A Call to Action. Curr. Opin. Behav. Sci. *26*, 54–61.

Tabor, A., Thacker, M.A., Moseley, G.L., and Körding, K.P. (2017). Pain: a statistical account. PLoS Comput. Biol *13*, e1005142.

Talmi, D., Seymour, B., Dayan, P., and Dolan, R.J. (2008). Human pavlovian-instrumental transfer. J. Neurosci. *28*, 360–368.

Taylor, V.A., Chang, L., Rainville, P., and Roy, M. (2017). Learned expectations and uncertainty facilitate pain during classical conditioning. Pain *158*, 1528–1537.

Tolman, E.C. (1948). Cognitive maps in rats and men. Psychol. Rev. *55*, 189.

Tracey, I. (2010). Getting the pain you expect: mechanisms of placebo, nocebo and reappraisal effects in humans. Nat. Med *16*, 1277.

Treede, R.-D., Kenshalo, D.R., Gracely, R.H., and Jones, A.K.P. (1999). The cortical representation of pain. Pain *79*, 105–111.

Van Damme, S., Legrain, V., Vogt, J., and Crombez, G. (2010). Keeping pain in mind: a motivational account of attention to pain. Neurosci. Biobehav. Rev. *34*, 204–213.

Vlaev, I., Seymour, B., Dolan, R.J., and Chater, N. (2009). The price of pain and the value of suffering. Psychol. Sci. *20*, 309–317.

Vlaeyen, J.W., and Linton, S.J. (2000). Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. Pain *85*, 317–332.

Vogt, B.A. (2005). Pain and emotion interactions in subregions of the cingulate gyrus. Nat. Rev. Neurosci *6*, 533.

Vogt, B.A., and Sikes, R.W. (2000). The medial pain system, cingulate cortex, and parallel processing of nociceptive information. Prog. Brain Res. *122*, 223–235.

Wager, T.D., Rilling, J.K., Smith, E.E., Sokolik, A., Casey, K.L., Davidson, R.J., Kosslyn, S.M., Rose, R.M., and Cohen, J.D. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. Science *303*, 1162–1167.

Wager, T.D., Atlas, L.Y., Lindquist, M.A., Roy, M., Woo, C.W., and Kross, E. (2013). An fMRI-based neurologic signature of physical pain. N. Engl. J. Med. *368*, 1388–1397.

Wang, O., Lee, S.W., O'Doherty, J., Seymour, B., and Yoshida, W. (2018). Model-based and model-free pain avoidance learning. Brain Neurosci. Adv. *2*, 2398212818772964.

Wiech, K. (2016). Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. Science *354*, 584–587.

Wiech, K., Kalisch, R., Weiskopf, N., Pleger, B., Stephan, K.E., and Dolan, R.J. (2006). Anterolateral prefrontal cortex mediates the analgesic effect of expected and perceived control over pain. J. Neurosci. *26*, 11501–11509.

Wilson, R.C., Geana, A., White, J.M., Ludvig, E.A., and Cohen, J.D. (2014). Humans use directed and random exploration to solve the explore-exploit dilemma. J. Exp. Psychol. Gen *143*, 2074.

Winston, J.S., Vlaev, I., Seymour, B., Chater, N., and Dolan, R.J. (2014). Relative valuation of pain in human orbitofrontal cortex. J. Neurosci. *34*, 14526–14535.

Wittmann, B.C., Daw, N.D., Seymour, B., and Dolan, R.J. (2008). Striatal activity underlies novelty-based choice in humans. Neuron *58*, 967–973.

Woo, C.-W., Roy, M., Buhle, J.T., and Wager, T.D. (2015). Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. PLoS Biol *13*, e1002036.

Yelle, M.D., Oshiro, Y., Kraft, R.A., and Coghill, R.C. (2009). Temporal filtering of nociceptive information by dynamic activation of endogenous pain modulatory systems. J. Neurosci. *29*, 10264–10271.

Yilmaz, P., Diers, M., Diener, S., Rance, M., Wessa, M., and Flor, H. (2010). Brain correlates of stress-induced analgesia. Pain *151*, 522–529.

Yoshida, W., Seymour, B., Koltzenburg, M., and Dolan, R.J. (2013). Uncertainty increases pain: evidence for a novel mechanism of pain modulation involving the periaqueductal gray. J. Neurosci. *33*, 5638–5646.

Yu, A.J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron *46*, 681–692.

Zaman, J., Vanpaemel, W., Aelbrecht, C., Tuerlinckx, F., and Vlaeyen, J.W.S. (2017). Biased pain reports through vicarious information: A computational approach to investigate the role of uncertainty. Cognition *169*, 54–60.

Zhang, S., Mano, H., Ganesh, G., Robbins, T., and Seymour, B. (2016). Dissociable learning processes underlie human pain conditioning. Curr. Biol. *26*, 52–58.

Zhang, S., Mano, H., Lee, M., Yoshida, W., Kawato, M., Robbins, T.W., and Seymour, B. (2018a). The control of tonic pain by active relief learning. eLife *7*, e31949.

Zhang, S., Yoshida, W., Mano, H., Yanagisawa, T., Shibata, K., Kawato, M., and Seymour, B. (2018b). Endogenous Controllability of Closed-loop Brain Machine Interfaces for Pain. bioRxiv. https://doi.org/10.1101/369736.

Zunhammer, M., Bingel, U., and Wager, T.D.; Placebo Imaging Consortium (2018). Placebo effects on the neurologic pain signature: a meta-analysis of individual participant functional magnetic resonance imaging data. JAMA Neurol. *75*, 1321–1330.