HOMEOSTASIS AFTER INJURY: HOW INTERTWINED INFERENCE AND CONTROL UNDERPIN POST-INJURY PAIN AND BEHAVIOUR

Pranav Mahajan
Nuffield Department of Clinical Neurosciences
University of Oxford
pranav.mahajan@ndcn.ox.ac.uk

Peter Dayan
Max Planck Institute for Biological Cybernetics
University of Tübingen
dayan@tue.mpg.de

Ben Seymour
Nuffield Department of Clinical Neurosciences
University of Oxford
ben.seymour@ndcn.ox.ac.uk

May 16, 2025

ABSTRACT

Several arguments suggest that the brain might have a dedicated representation of the state of injury. This would provide an internal control system to modulate behaviour given the changed homeostatic priorities associated with injury, including pain, anxiety and mood changes appropriate to the need for heightened protection and recuperation during healing. Here, we propose a computational architecture for how this might be constructed, treating the injury as a partially observable Markov decision process (POMDP), and proposing a Bayesian decision-theoretic solution that combines inference with optimal control. We show how this offers an explanation of two core paradoxical observations: behaviours such as rubbing an injured area (conventionally viewed under the lens of gate control theory), and high propensity of transition to pathological chronic pain states. Overall, this provides a quantitative framework for mapping injury homeostasis to neural substrates, with potential for identifying novel chronic pain targets.

1 Introduction

Injuries are common across the lifespan of many species, and often lead to a period of vulnerability and reduced functionality whilst healing occurs. In both ecological and laboratory studies, characteristic behavioural changes such as increased anxiety and reduced activity are also often observed. They are considered to reflect adaptive changes and altered motivational priorities appropriate for safe recovery (Wall, 1979; Bolles and Fanselow, 1980; Williams, 2019). These changes are notoriously accompanied, and indeed potentially mediated, by forms of pain. Noting that organising appropriate behaviour is, by definition, a problem of control (Sutton and Barto, 2018), Seymour et al. (2023) offered a substantially new perspective on tonic pain. Here, we centre Seymour et al. (2023) in its natural reinforcement learning context, and provide the first concrete computational realisation of these ideas.

The critical innovation in Seymour et al. (2023) was to note that the brain suffers inevitable uncertainties about the nature and extent of the injury and the status of the recovery process. They suggested that the brain continually integrates multisensory and physiological inputs to infer, and thereby represent, the uncertain injury state. This representation would then be tied to the appropriate choice of action or inaction, and generate internal signals interpreted as pain. This framework extends Bayesian models of pain perception (Seymour et al., 2013; Büchel et al., 2014; Wiech, 2016) to address not only inferential aspects (for instance, expectancy effects) but also control under uncertainty. Crucially, they proposed that protective behaviours might restrict access to informative signals about recovery, potentially leading to persistent or maladaptive injury beliefs and contributing to chronic pain.

We formalize the suggestion of Seymour et al. (2023) as a partially observable Markov decision process (POMDP; Kaelbling et al. (1998)). This situates injury, recovery, and pain within the explanatory framework of neural reinforcement learning (Sutton and Barto, 2018; Dayan and Daw, 2008), where the solution is a policy mapping observations to actions or inaction. In doing so, our work addresses several gaps in the literature: it (i) unifies Bayesian inference and control approaches in pain research (Seymour and Mancini, 2020), (ii) mathematically formalises in a minimal model, how information restriction about whether injury has resolved, may drive chronic pain states (Seymour et al., 2023), and (iii) makes explicit how the value of information influences pain-related behaviour (Seymour, 2019).

We focus on the belief about the injury state constructed by the brain, proposing that this belief underlies the experience of tonic pain and modulates phasic pain responses. While this belief is informed by ascending sensory signals—particularly sustained nociceptor firing and hyperexcitability—such inputs, especially those from small-diameter, unmyelinated C-fibres (Debanne, 2004), may be unreliable or imprecise, especially during the later stages of healing. Accurate injury state estimation therefore requires integrating multiple sources of information, including autonomic and physiological signals, exteroceptive cues such as vision (Höfle et al., 2010), and prior expectations (Mancini et al., 2022; Seymour and Mancini, 2020; Seymour et al., 2023), accumulated over time to construct a coherent, control-relevant internal representation (Büchel et al., 2014; Wiech, 2016).

Although our simulations are inevitably limited, they show how tonic pain, as the belief about the state of the injury in the first instance, is a critical, internally constructed and interpreted signal which has broadcast properties, reorganizing macroscopic patterns of behaviour. In our simulation results, we will dwell on a particular facet of the injury POMDP, which is the potentially large cost of the conventional policy solution to uncertainty, namely information gathering to determine which options are better, or at least less worse. This cost could accrue if, for instance, the only way to learn about the current state of the injury is to attempt to use the injured body part. We examine such counter-intuitive actions of injury investigation (e.g. rubbing injured areas conventionally interpreted in the context of Gate Control Theory) in the light of our theory, and also suggest how the costs might lead to the development of pathological states of chronic pain, even when the periphery is apparently no longer presenting nociceptive input (Fitzcharles et al., 2021).

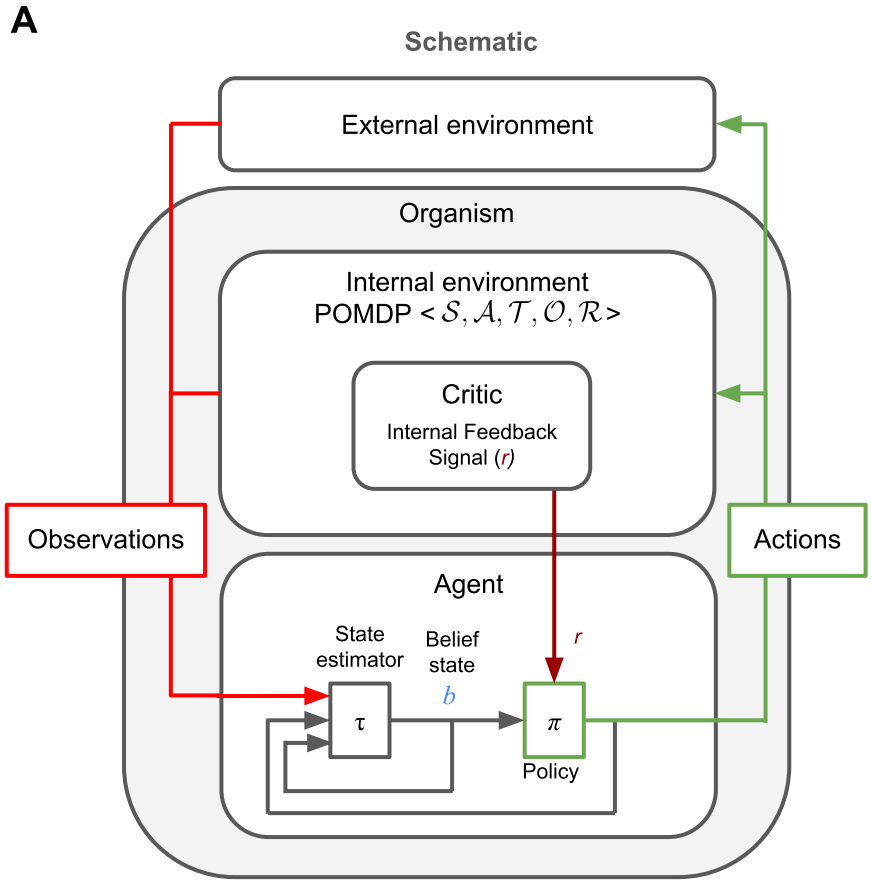
We proceed by building a simple POMDP model with just a handful of states and actions, and studying the properties of its optimal policy for different parameter values and starting states. The POMDP may seem rather abstract; we therefore tie it as closely as possible to phenomena observed in the context of pain.

2 Results

2.1 Theory sketch

We propose that the brain treats injury in terms of a partially observable Markov Decision Problem (POMDP; Fig. 1A) (Drake, 1962; Astrom et al., 1965; Sondik, 1971; Kaelbling et al., 1998). A POMDP comprises states $\mathcal{S}$, actions $\mathcal{A}$, transitions $\mathcal{T}$, observations $\mathcal{O}$ and utilities $\mathcal{R}$; we make substantial simplifications in our construction of all of these. The decision-theoretic task for the brain is then to determine a policy which maps the history of observations to the probability of taking an action to maximize the so-called future *return*, which is a long-run measure of cumulative utility. The general Bayesian Decision Theoretic framework is further described in the Methods section.

For the purpose of our simulations, we construct a simplified injury POMDP describing the following situation: A patient contemplating whether or not to commit to a demanding activity (e.g. getting groceries), whilst being uncertain about their injury. The true state characterizes the current circumstance of the injury. We consider just two possible states $s_t \in \{0, 1\}$ for healthy or injured at time t. However, crucially, the brain lacks full information about the state – interoception is incomplete, as well as being noisy. The agent thus only has a so-called *belief* state $b_t \in [0, 1]$, which is a probabilisic distribution over all states (Kaelbling et al., 1998). The actions include everything that could be done; however, we make the radical simplification of considering just three to four actions: a_{act}, which involves anything physically demanding but can collect resources; $a_{\text{r\&r}}$, which allows time for recovery and recuperation; and a_{que}, which involves assessing the injury, for instance trying to walk on a recently broken leg (Fig. 1B). We also allow for a null action a_{nul} that does nothing (we omit this from a few experiments for simplicity). Transitions specify how states change, probabilistically, based on the actions – in general, with recovery and worsening of the injury, although here, we restrict ourselves to the case that the state actually remains constant (and so drop the time index for s). Observations include all the exteroceptive and interoceptive information available to the brain about the injury, requiring multisensory integration, but we reduce it to a single observation channel for simplicity in our simulations. Investigatory actions such as probing the injured area might provide both types of information.



B

Simulation Action Set

a_{act}	Commit to a demanding activity e.g. get groceries	
$a_{r\&r}$	Not committing to the activity e.g. rest at home	
a_{que}	Investigating the injury e.g. probing the injured area	
a_{nul}	Avoiding to disturb the injury (omitted in some experiments)	

Figure 1: (A) Schematic of the injury POMDP, with an internal environment generating observations and conferring utilities and an internal agent inferring a belief state (Kaelbling et al., 1998) to choose optimal actions. (B) The POMDP action set used for simulations.

Utilities are central to determining the optimal policy and are a contentious aspect of the POMDP. In standard RL, utilities are externally provided by the environment (e.g., points in a video game). However, in nature, animals must infer or construct affective consequences from observations alone. This has inspired work on homeostatic reinforcement learning (Keramati and Gutkin, 2014) and intrinsic rewards (Chentanez et al., 2004; Barto and Simsek, 2005; Singh et al., 2009; Dayan, 2022). Here, as a simplification, we assume the agent has an intrinsic reinforcement function $r(s, a)$, defining immediate affective consequences of action a in true state s. For instance, this function might be large and negative/positive for activity/rest actions when injured (the former as a convenient proxy for the very long-run costs of incurring extra damage); small and negative for investigating while injured (due to potential harm); and negative for

resting while uninjured, reflecting unmet resource needs (again, as a proxy for the very long run effects of homeostatic deterioration). Since the agent only knows its belief b_t about s, the expected utility is internally constructed.

We illustrate the injury POMDP in a simplified setting where the agent chooses a single, final, choice of a_{act} or $a_{\text{r\&r}}$, but can choose to a_{que} along the way if it is sufficiently uncertain about the injury state. The true state remains fixed throughout the decision-making episode, and the return is the cumulative utility of the trajectory of choices. The one-step expected utilities in the belief space of the injury POMDP used in our simulations are shown in Fig. 2A. Note, that these expected utilities differ from long-run return for actions a_{que} and a_{nul}. Choosing a_{que} and a_{nul} also results in sampling observations from distributions shown in Fig. 2B. We see a_{que} provides more discriminating or discerning observations about the internal state and is therefore more informative than a_{nul}. In our results, we often report state-action values $Q(b_t, a)$, which represent the expected long-run return for belief state b_t and action a. The optimal action in any b_t is the one with the highest Q value.

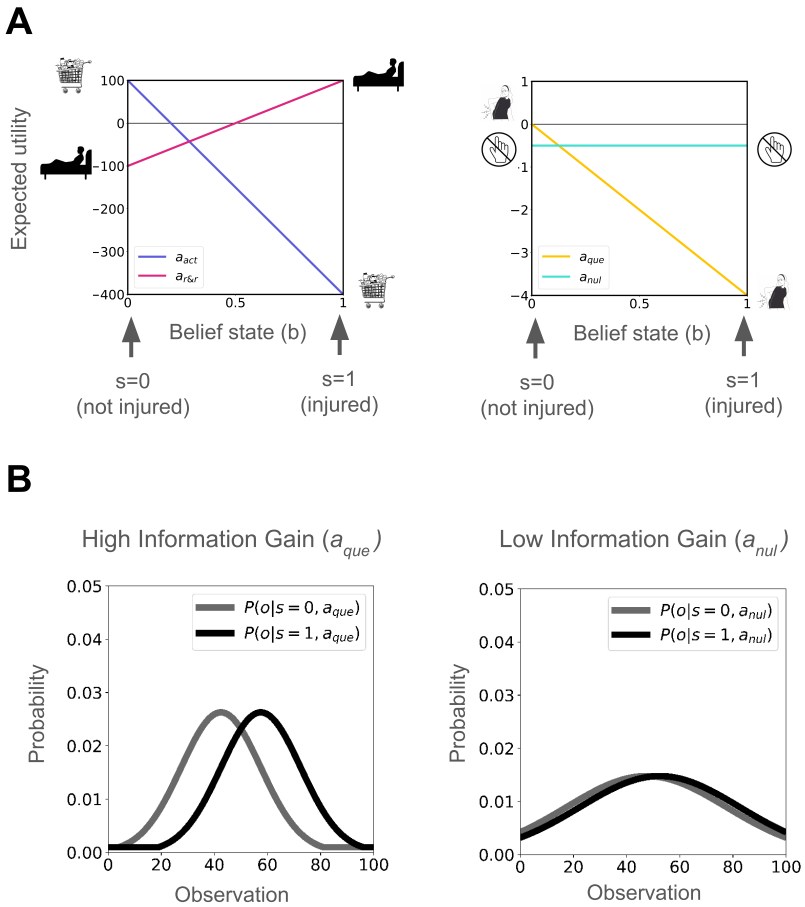


Figure 2: Injury POMDP: (A) Expected utilties of actions a_{act}, $a_{\text{r\&r}}$, a_{que} and a_{nul} in belief space. Utilities: when injured: $r(s = 1, a_{\text{act}}) = -400, r(s = 1, a_{\text{r\&r}}) = +100, r(s = 1, a_{\text{que}}) = -4$; when uninjured: $r(s = 0, a_{\text{act}}) = +100, r(s = 0, a_{\text{r\&r}}) = -100, r(s = 0, a_{\text{que}}) = 0$. $r(a_{\text{nul}}) = -0.5$ regardless of the internal state as a minor opportunity cost. (B) Observation processes for actions a_{que} and a_{nul}. (Further details in Methods - model simulation)

For concreteness, as a substantial simplification, we associate the belief state b_t with tonic pain – higher belief in being injured corresponds to greater pain. This pain becomes chronic if the agent fails to act or gather evidence to revise its belief. We link the expected negative reinforcement from a_{que} to phasic pain caused by injury investigation during the episode. This expectation averages over the belief state (see Methods), providing a mechanism for precisely tuning the feedback (Seymour, 2019) in case of correct inferences, but is also susceptible to incorrect or underinformed inferences.

2.2 Normative consequences

Why do we investigate injury despite it being painful?

Our model provides a normative explanation for the behaviour as to why we investigate our injury despite it being painful. This is difficult to explain through a simple inferential or a control-theoretic approach alone.

In this instance, the feedback for a_{que} when injured is $r(s = 1, a_{\text{que}}) = -4$, denoting the negative consequence of probing the injury when injured. However, it also results in sampling observations using an observation process that can help infer the internal state more accurately. For this simulation demonstrating the value of information gained from injury investigation, we omit a_{nul} for simplicity, but including it does not affect the results.

When starting from an uncertain belief state of $b_0 = 0.5$, the agent progressively samples observations by choosing a_{que} at the cost of some (self-constructed) phasic pain. This is because there is a value to the information that can be gained given this uncertainty (see Fig. 3). With every sample, the agent accumulates evidence about the internal state and updates its belief until it can commit to either action a_{act} or $a_{\text{r\&r}}$ (Fig. 3, red and blue arrows). The belief state acts as the context for driving optimal behaviour. When the true internal state is not injured $s = 0$, the agent updates its beliefs (red arrows) and chooses the optimal decision a_{act} in the end. When the true internal state is injured $s = 1$, the agent updates its beliefs (blue arrows) and chooses the optimal decision $a_{\text{r\&r}}$ in the end. In conclusion, this demonstrates agent choosing injury investigating/probing actions despite associated with costs so as to accrue enough evidence to take the optimal decision. Note, actions investigating an injury extend beyond "rubbing the injured area" (explained by Gate Control Theory as "touch inhibits pain") to include non-contact explorations like bending a painful back or moving a painful joint etc.

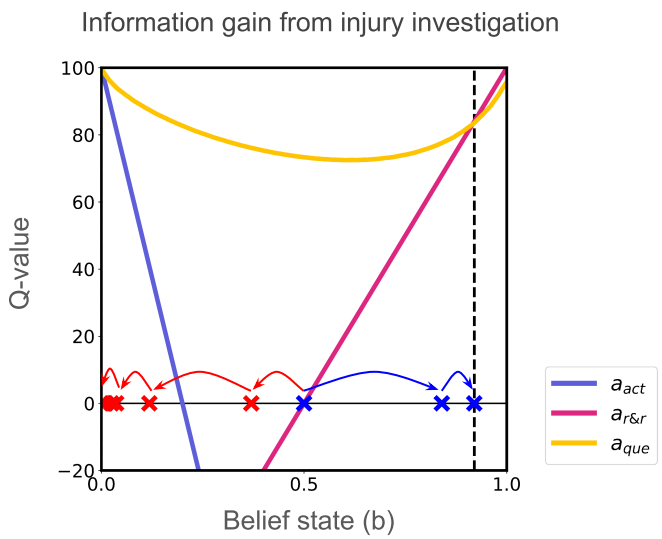


Figure 3: Action values after value iteration. Red arrows show belief updates under true state $s = 0$ (not injured), whereas blue arrows show belief updates under true state $s = 0$ (injured). The agent takes multiple a_{que} actions, reaching either belief thresholds, where either choosing a_{act} or $a_{\text{r\&r}}$ is more valuable than a_{que} and thus terminating the episode.

Trade-off between information gain and phasic pain:

Consider the example of assessing your back after an injury. Mild discomfort from actions like stretching (e.g., "pins and needles") can provide valuable information about the injury's state, outweighing the discomfort. In contrast, persistent tonic pain from bending offers little clarity about healing, as the sensory feedback may not distinguish between injury or recovery, making such actions less valuable.

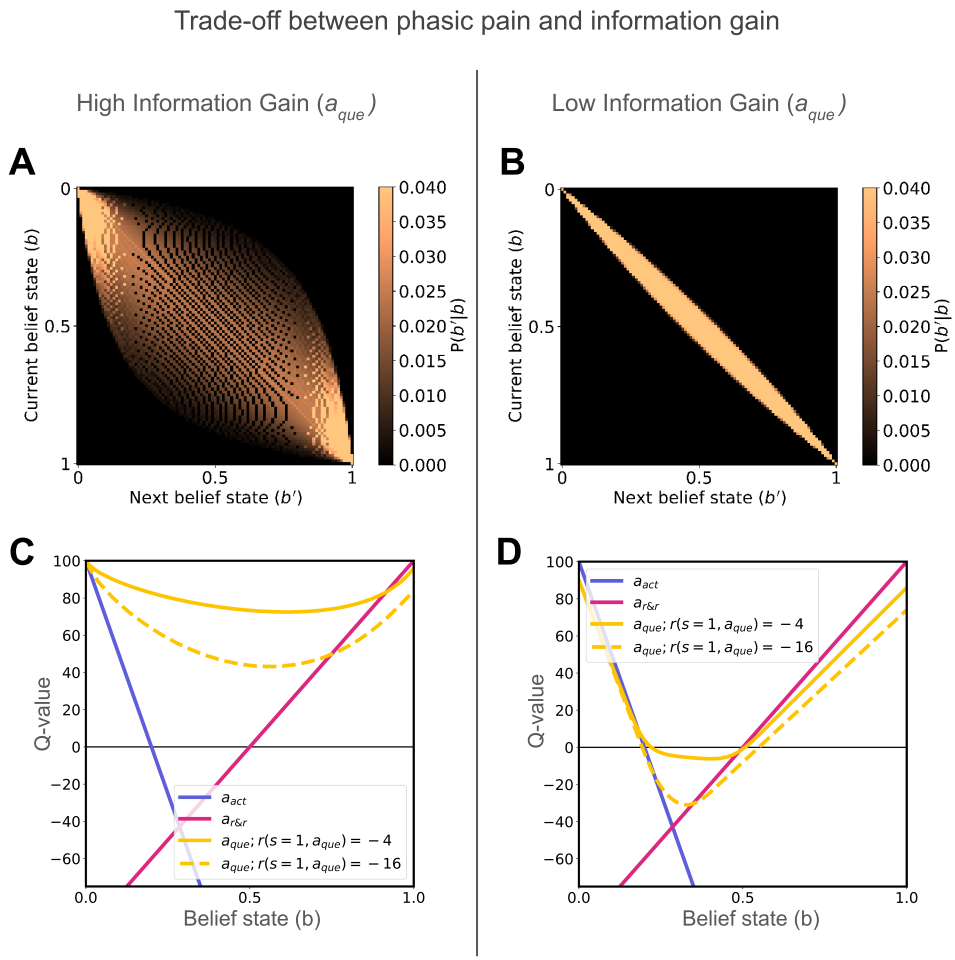


Figure 4: Didactic demonstration of a 2×2 factorial design. (A, B) Belief transitions from current belief state (b) to subsequent belief state (b') reflect the influence of high vs. low informativeness of the observation process on belief updating. (C, D) Action value plots show how phasic pain ($r(s = 1, a_{\text{que}}) = -4$ vs. $r(s = 1, a_{\text{que}}) = -16$) and high vs low information gain influence a_{que}. Higher phasic pain reduces action value, while higher information gain increases value when the expected cost remains constant.

We explored this trade-off between information gain and phasic pain using a 2×2 factorial simulation experiment. We manipulated the information gain by altering the observation process for a_{que}, using the two observation processes from Fig. 2B. In this simulation, we omit a_{nul} for simplicity of our didactic demonstration.

We find higher information gain of the observation process results in greater belief updates (Fig. 4A), and thus confers a higher action value of a_{que} across various uncertain belief states (Fig. 4C). This contrasts with less informative observation process which results in small belief updates (Fig. 4B) and lower value (Fig. 4D). Further, when observation process is the same, higher cost of a_{que} decreases the value of choosing a_{que} (Fig. 4C,D). This dynamic illustrates the trade-off between exploratory behaviour and pain avoidance. This experiment also highlights a key facet of our framework: the informativeness of observations and the motivational (phasic) punishment used for credit assignment are disentangled/treated independently. In conclusion, the trade-off between information gain and phasic pain determines the value of an action in our model.

2.3 Dysfunctional consequences

Having illustrated the normative consequences of the model, we will next consider how it might predict pathways to post-injury chronic pain. We describe two such pathways: (1) Information restriction and (2) Aberrant priors.

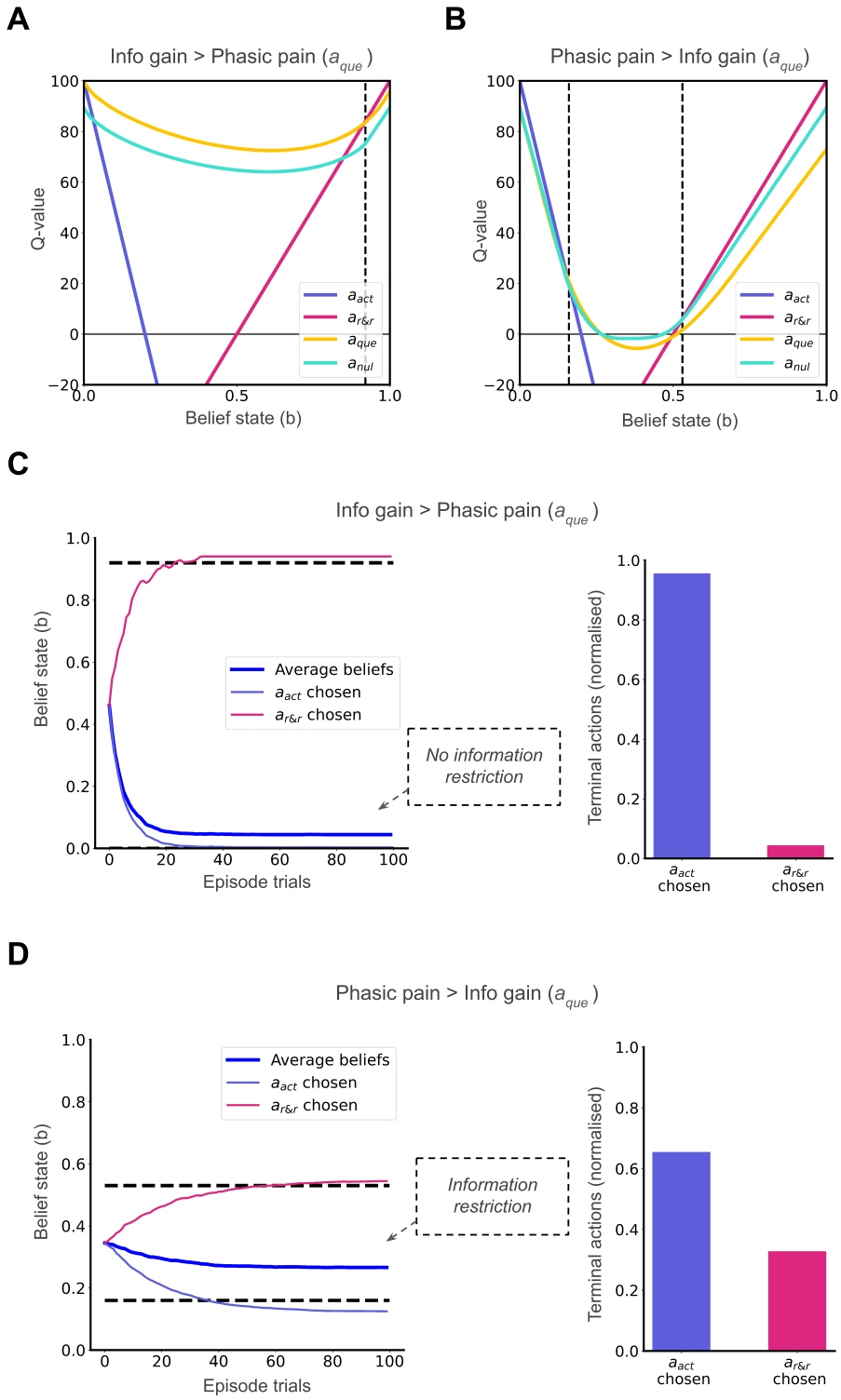


Figure 5: Two possible scenarios along the recovery where the internal state is $s = 0$: (A) Q-values when information gain from a_{que} outweighs phasic pain costs; (B) Q-values when phasic pain costs outweigh information gain. (C) Faster adaptive hidden state resolution is seen when more informative a_{que} is chosen over less informative a_{nul} (D) Maladaptive information restriction occurs when less informative a_{nul} is chosen over more informative a_{que}. Belief state trajectories are averaged over 1000 simulation episodes (dark blue) or only the episodes ending in a_{act} (violet) and $a_{\text{r\&r}}$ (pink). 'Episode trials' represent steps within a decision-making episode.

Information restriction:

A person recovering from a back injury might reasonably avoid movement due to fear of pain. This would hinder them from gathering valuable information that could signal the resolution of the injury. Such a Fear-Avoidance (Crombez et al., 2012) route to pain chronification has been elaborated in the path-dependent (Dayan et al., 2020) suggestion of the information restriction hypothesis (Seymour et al., 2023).

To investigate this route to pain chronification, we simulated the injury POMDP in which the true internal state is no longer injured ($s = 0$), as in the late stages of recovery. In this scenario, the action to investigate the injury (a_{que}) is nevertheless slightly painful when $b_t > 0$, due to the uncertainty of the injury. Action a_{que} is more informative than a_{nul}, using the observation processes shown in Fig. 2B. The action a_{nul} has no cost associated with it, but is less informative than a_{que}. The difference from $a_{\text{r\&r}}$ is that it is not a terminating action - the agent could subsequently change its mind and execute one of the three other action.

We varied the the cost of a_{que} from $r(s = 1, a_{\text{que}}) = -4$ to $r(s = 1, a_{\text{que}}) = -16$, while keeping the observation processes fixed to explore different action values where information gain exceeds phasic pain and vice versa respectively (Fig. 5A and B). We set the initial belief at the mid-point between the thresholds for choosing actions a_{act} and $a_{\text{r\&r}}$. When the information gained from a_{que} outweighs its associated pain, the individual's beliefs adaptively update, correctly resolving the hidden state to $b_t = 0$, leading ultimately to the optimal action ($a_{\text{r\&r}}$) over time (Fig. 5C, averaged belief state trajectories). However, when the phasic pain of a_{que} exceeds the information gain, it is less valued than a_{nul} and a_{nul} is preferred. Due to the less informative nature of a_{nul}, the beliefs do not change greatly. This leads to information restriction, as observed in belief state trajectories averaged over all simulation episodes (Fig. 5D). Thus, we can observe a prolonged incorrect belief that the injury persists, leading to slower recovery or persistent injury mis-inference, which in our model correlates with increased tonic pain (as per eqn. 11). Additionally, this behaviour results in more suboptimal avoidance behaviour - we see an increase in agent choosing $a_{\text{r\&r}}$ (Fig. 5D, right). This simulation illustrates how restricted information due to pain avoidance can contribute to the prolonged incorrect belief that the injury persists, providing a computational account for the chronic pain pathway described by (Seymour et al., 2023).

Aberrant priors:

We next study the adverse effects of having a strongly held incorrect belief about the severity of an injury e.g. an aberrant prior. A prior (belief at the start of the decision-making episode) that underestimates the injury leads to further unnecessary harm, whereas a prior that overestimates the injury leads to over-protective behaviour. Both of our simulations use the same Q-values as Fig. 5A.

In the first experiment, we use an underestimating prior when the true state is injured ($s = 1$). With aberrant starting belief states closer to zero (but greater than zero, as this is a decision threshold for a_{act}), we observe that the agent chooses a_{que} over a_{nul}. This results in greater average cumulative expected costs (interpreted as cumulative phasic pain) accrued during the episode in the interest of accurate state inference and subsequent optimal decision-making (Fig. 6A). These excess costs decrease as the starting belief becomes closer to $b_0 = 1$, i.e., being better aligned with the true state (Fig. 6A). This demonstrates a pathway for excess phasic pain through injury investigation due to maladaptive priors underestimating an injury when the true state is injured.

In our second experiment, we use an overestimating prior when the true state is not injured ($s = 0$). If the starting belief state is aberrantly closer to 1 (itself a possible consequence of probability distortions in the context of risk aversion; Gagne and Dayan, 2022), then it contributes to prolonged incorrect belief that the injury persists, seen in belief trajectories in Fig. 6B. Such prolonged incorrect belief decreases as the starting belief becomes closer to $b_0 = 0$, i.e., again being better aligned with the true state (Fig. 6B). This demonstrates as pathway for persistent tonic and potentially chronic pain, due to maladaptive priors overestimating the injury when the true state is recovered.

3 Discussion

We present a theoretical framework for understanding the computational logic of a dedicated homeostatic state for injury (Seymour et al., 2023). Central to this framework is the notion that internal states are partially observable and require inference. This perspective provides a normative explanation for behaviours such as probing an injury despite immediate phasic pain, as these actions prioritise information acquisition. Additionally, we identify two broad "fault lines" that can lead to suboptimal behaviour and chronic pain, offering insights into how the homeostatic injury system can go awry.

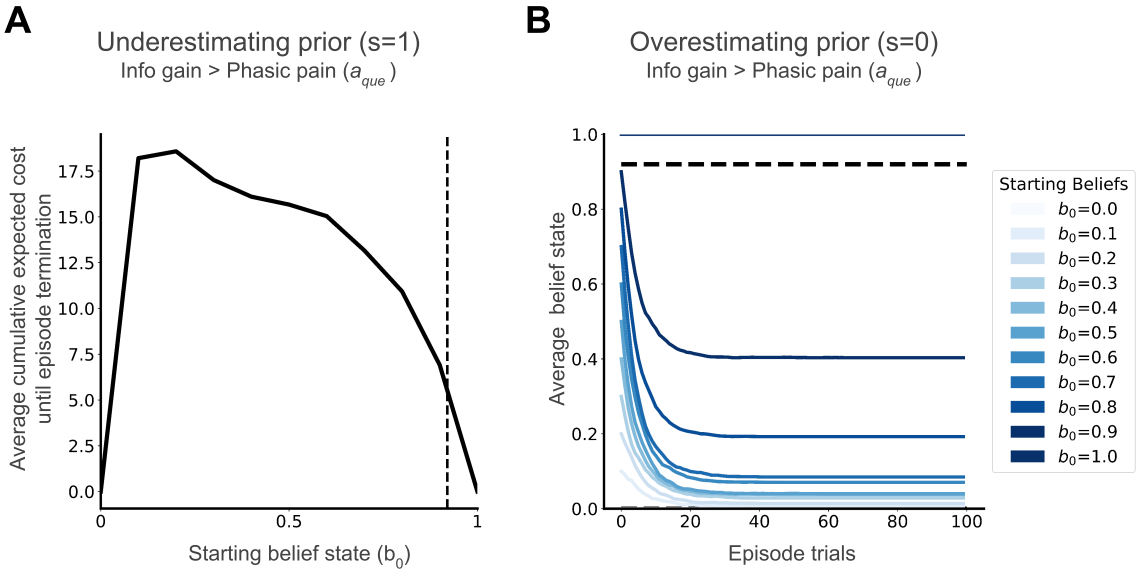


Figure 6: (A) Underestimation of injury when injured ($s = 1$), modelled as start belief states closer to 0, increases average phasic pain when injured. (B) Overestimation of injury when not injured ($s = 0$), modelled as start belief states closer to 1 but short of the decision threshold, can lead to prolonged incorrect beliefs, as seen in mean belief state trajectories. Belief state trajectories are averaged over 1000 simulation episodes.

Our framework builds on prior models of homeostatic motivation (Chentanez et al., 2004; Barto and Simsek, 2005; Singh et al., 2009) in which reinforcement signals are generated within the agent from observations, rather than being provided externally. However, we highlight the necessity of inferring internal states based on noisy observations and thus partial observability.

Our results align with aspects of the Fear-Avoidance model of pain chronification (Vlaeyen and Linton, 2000; Vlaeyen et al., 2016), offers a rich psychological framework grounded in fear learning and patient experience, positing that fear of movement leads to avoidance and physical deconditioning, thereby reducing opportunities to update beliefs about injury resolution. We offer the complementary approach (see Appendix Figure) of formalising information restriction and avoidance dynamics, providing a normative, computational, inference-based account that quantifies when and why avoidance could sustains maladaptive beliefs, consistent with theories of ex-consequentia reasoning (Arntz et al., 1995; van Vliet et al., 2018; Meulders, 2019). Thus our model makes a precise, testable prediction about how avoidance behaviour (or its absence) can lead to persistent chronic pain (or recovery), respectively. This information-restriction pathway can be assessed in terms of stagnated belief updates about injury resolution. In contrast, the Fear-Avoidance model remains largely qualitative. While it clearly links avoidance to persistent fear, and lack of avoidance to fear extinction — drawing on insights from fear conditioning — its connection to chronic pain is more indirect. It relies on the narrative that fear of pain or re-injury discourages activity, leading to physical deconditioning of muscles and disability and, ultimately, persistent pain (Vlaeyen and Linton, 2000). Crucially, it thereby lacks such a formal account of how injury-related beliefs are inferred or updated.

Our simulations with aberrant priors, particularly those reflecting overestimation, also align with self-fulfilling prophecy and catastrophisation models (Jepma et al., 2018; Flink et al., 2013), which highlight how catastrophic thinking enhances attentional demand and hinders disengagement from pain (Van Damme et al., 2002, 2004).

Looking ahead, the "fault lines" we identify can be categorised as the agent: solving the "wrong problem," solving the correct problem with the "wrong solution," or solving the correct problem correctly in an unfortunate, "wrong environment" (Huys et al., 2015). For example, aberrant priors illustrate solving the wrong problem, while maladaptive behaviours, such as excessive avoidance due to Pavlovian biases (Mahajan et al., 2024) or habitual traits (Ball and Gunaydin, 2022), represent wrong solutions that could lead to information restriction. A wrong environment may yield priors or utilities that were once adaptive but have become maladaptive, creating scenarios that appear to involve the wrong problem or solution. This categorisation provides a theoretical framework for linking current behavioural theories of chronic pain under a unified computational perspective, akin to computational nosology (Mathys et al., 2016; Friston et al., 2017).

Our model is particularly relevant to clinical chronic pain conditions involving ambiguous symptoms, such as sciatica after lumbar radiculopathy. For instance, it may explain observations with sciatica muscle weakness patients, where baseline muscle weakness (a worse symptom) is counter-intuitively associated with improved outcomes (Konstantinou et al., 2018; Vroomen et al., 2002). Recovery of muscle strength could provide objective and unambiguous observations thereby providing high-information feedback for accurate injury inference and self-improvement. Similarly, our model could potentially explain phenomena such as boom-bust activity cycles (Moseley, 2003; Antcliff et al., 2016) via faulty injury inference feeding into maladaptive behaviour. Potential interventions could involve improving injury inference, for example, through external cues or reducing maladaptive behaviour by activity pacing. Moreover, our framework may offer insights into the positive outcomes of Pain Reprocessing Therapy (Gordon and Ziv, 2021), particularly the role of somatic tracking and guided touch (Kim et al., 2022; McParlin et al., 2022).

The model is just a first step. In particular, we did not include transitions in internal states, and thus actual recovery or exacerbation of injuries. We rather arbitrarily generated the utility terms in our simulations; in the future, we plan to derive them from both physiological costs and the opportunity costs associated with the threats to homeostasis that forced inaction imposes. We have yet to model concrete neural or behavioural data (although we have identified ACL-tear as an ideal test case). More subtly, there are uncertainties about the exact modeling construct of tonic pain. Here, we modelled it directly using the belief over an injury state, which is in line with previous lines of work treating pain in terms of Bayesian inference. However, if an experiment is able to demonstrate that actions associated with information gain reduce tonic (as well as phasic) pain, as in endogenous analgesia (Basbaum and Fields, 1978; Bannister and Dickenson, 2017) then it would deem value over belief states $V(b)$ as a more accurate correlate of tonic pain. Alternatively, one may also choose to represent tonic pain and its multiple attributes as a predictive state representation, as in general value functions (Sutton et al., 2011). Each of these various approaches implies a different meaning to tonic pain and its role in guiding behaviour.

Finally, and perhaps of greatest translational importance, chronic pain likely arises from complex interactions between learning processes, internal and external environments, and behavioural feedback loops. By offering a computational framework that spans from phasic to chronic pain, our account provides a mathematical foundation for making theories of this condition explicit.

4 Methods

Bayesian decision theoretic approach to homeostasis

We build upon the Bayesian Decision Theoretic (BDT) framework (Dayan and Daw, 2008; Huys et al., 2015) utilising POMDPs (Drake, 1962; Astrom et al., 1965; Sondik, 1971; Kaelbling et al., 1998; Ross et al., 2008) and extend it to the problem of homeostasis.

Our internal environment POMDP is formally represented as a tuple $\langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{O}, \mathcal{R} \rangle$, where: $\mathcal{S}$ is the set of all internal states (s). $\mathcal{A}$ is the set of all possible actions (a). $\mathcal{T}$ is the transition function, where $\mathcal{T}(s, a, s') = P(s'|s, a)$ is the probability of ending in s' by performing action a in state s. $\mathcal{O}$ is the set of all observations (o), where the observation function is $\mathcal{O}(s', a, o) = P(o|a, s')$ is the probability of observing o if action a is performed and resulting state is s'. $\mathcal{R}$ are the utilities, where $r(s, a)$ is the internal feedback signal obtained by taking action a in internal state s.

At any time t, the agent does not necessarily have access to its complete internal state s_t, but has access to observations o_t from the internal environment, which provide incomplete (noisy) information about the internal state. The agent therefore infers a belief state b_t over internal state space $\mathcal{S}$ (Astrom et al., 1965). The belief state is defined as the posterior probability density of being in each internal state s, given the complete history h_t of actions and observations at any time t, $h_t = \{a_0, o_1, ...o_{t-1}, a_{t-1}, o_t\}$ and initial belief b_0

$$b_t(s) = P(s_t = s|h_t, b_0) \tag{1}$$

The belief state b_t is a sufficient statistic for the history h_t (Smallwood and Sondik, 1973; Sondik, 1978). At any time t, the belief state b_t can be computed from the previous belief state b_{t-1}, using the previous action a_{t-1} and the current observation o_t using a state estimator function (τ) derived using Bayes rule (Kaelbling et al., 1998; Ross et al., 2008)

$$\begin{aligned} b_t(s') &= \tau(b_{t-1}, a_{t-1}, o_t)(s') \\ &= \frac{\mathcal{O}(s', a_{t-1}, o_t) \sum_{s \in \mathcal{S}} \mathcal{T}(s, a_{t-1}, s') b_{t-1}(s)}{P(o_t|b_{t-1}, a_{t-1})} \end{aligned} \tag{2}$$

where the $P(o_t|b_{t-1}, a_{t-1})$ is the marginal distribution over observations o_t (independent of s', acting as a normalising factor for b_t).This belief transition function (τ) is used to obtain belief transition plots and belief trajectories.

A critical concept for the POMDP is a stationary policy $\pi(b, a)$ which is a probability distribution over the action a that the agent will take given that its current belief is b. Define $V^\pi(b_t)$ as the expected discounted long-term reward following policy π as the value of this belief state:

$$V^\pi(b) = \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \sum_{a_t \in \mathcal{A}} \rho(b_t, a_t)\pi(b_t, a_t)|b_0 = b\right] \tag{3}$$

where γ is a temporal discount function and

$$\rho(b_t, a_t) = \sum_{s \in \mathcal{S}} b_t(s) r(s, a_t) \tag{4}$$

is the feedback function on belief states, constructed from the original feedback signal on internal states $r(s, a)$. Then, it can be shown that the value satisfies the Bellman equation:

$$V^\pi(b) = \sum_{a \in \mathcal{A}} \pi(b, a)\Big[\rho(b, a) + \gamma \sum_{o \in \mathcal{O}} P(o|b, a) V^\pi(\tau(b, a, o))\Big] \tag{5}$$

where the sum over $\mathcal{O}$ is interpreted as the expected future return over the infinite horizon of executing action a, assuming the policy π is followed afterwards. It is known that there is a deterministic policy $\pi^*(b, a)$ that optimizes the value of all belief states (Ross et al., 2008)

$$\pi^*(b) = \arg\max_{\pi \in \Pi} \mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t \sum_{a \in \mathcal{A}} \rho(b_t, a)\pi(b_t, a)|b_0 = b\right] \tag{6}$$

The value function V^* of the optimal policy π^* is the fixed point of the Bellman's equation (Bellman, 1954)

$$V^*(b) = \max_{a \in \mathcal{A}}\Big[\rho(b, a) + \gamma \sum_{o \in \mathcal{Z}} P(o|b, a) V^*(\tau(b, a, o))\Big] \tag{7}$$

and one can define a corresponding optimal Q-value function, which refers to the value of taking action a and then following the optimal policy, as

$$Q^*(b, a) = \rho(b, a) + \gamma \sum_{o \in \mathcal{O}} P(o|b, a) V^*(\tau(b, a, o)) \tag{8}$$

Model implementation and simulation details

We do not aim to model the entire lifetime of an agent but rather an episode in the agent's lifetime. Here, we maintain a binary hidden state $s = \{0, 1\}$ and we do not model transitions between states. Since we do not model transitions in s_t within the POMDP episode, the state estimator equation 2 defining the belief transitions is reduced to the following,

$$b_t(s) = \tau(b_{t-1}, a_{t-1}, o_t)(s) = \frac{\mathcal{O}(s, a_{t-1}, o_t) b_{t-1}(s)}{P(o_t|b_{t-1}, a_{t-1})} \tag{9}$$

This is similar to the Tiger POMDP environment by (Kaelbling et al., 1998), except that the a_{que} is also belief-dependent. Further we refer to the belief state over s as $P(s = 1) = b_t(s = 1) = b$ and $P(s = 0) = b_t(s = 0) = 1 - b$. If the POMDP has more states, the belief state provides the probability distribution of being in each of those states.

To solve the POMDP, we treat it as a belief MDP, and of the several approaches to perform value approximation (Hauskrecht, 2000; Kochenderfer et al., 2022), we utilise belief grid value iteration (Lovejoy, 1991)leading to piecewise linear and convex (PWLC) value functions. Here, we discretise the belief space in steps of 0.01 and the observation distribution, $o \in (0, 100] = \mathcal{O}$, in steps of 1 and assume the observation function to be known.

From the action set $\mathcal{A} = \{a_{\text{act}}, a_{\text{r\&r}}, a_{\text{que}}, a_{\text{nul}}\}$ described in Fig. 1, a_{act} and $a_{\text{r\&r}}$ result in episode termination and a_{que} and a_{nul} do not. We then used the Q values of each action in each belief state to determine the decision thresholds for choosing actions a_{act} and $a_{\text{r\&r}}$ which terminate the episode, defined as the most uncertain belief state b where the optimal action is no longer a non-terminal action.

Simulation parameters

Experiment	True state	Outcome for a_{que} when injured	Observation model for a_{que}
Injury investigation	$s = 1$	$r(s = 1, a_{\text{que}}) = -4$	$O(s, a_{\text{que}}, o) = HI$
Phasic pain - information gain trade-off (High information gain condition)	$s = 1$	$r(s = 1, a_{\text{que}}) = -4, -16$	$O(s, a_{\text{que}}, o) = HI$
Phasic pain - information gain trade-off (Low information gain condition)	$s = 1$	$r(s = 1, a_{\text{que}}) = -4, -16$	$O(s, a_{\text{que}}, o) = LI$
Information restriction (Info gain >phasic pain)	$s = 0$	$r(s = 1, a_{\text{que}}) = -4$	$O(s, a_{\text{que}}, o) = HI$, $O(s, a_{\text{nul}}, o) = LI$
Information restriction (Phasic pain >info gain)	$s = 0$	$r(s = 1, a_{\text{que}}) = -16$	$O(s, a_{\text{que}}, o) = HI$, $O(s, a_{\text{nul}}, o) = LI$
Underestimating prior	$s = 1$	$r(s = 1, a_{\text{que}}) = -4$	$O(s, a_{\text{que}}, o) = HI$, $O(s, a_{\text{nul}}, o) = LI$
Overestimating prior	$s = 0$	$r(s = 1, a_{\text{que}}) = -4$	$O(s, a_{\text{que}}, o) = HI$, $O(s, a_{\text{nul}}, o) = LI$

The utilities mentioned in Fig. 2 were fixed throughout, with the exception of experiments where phasic pain exceeds information gain for a_{que}, therefore, we vary $r(s = 1, a_{\text{que}}) = -4$ to $r(s = 1, a_{\text{que}}) = -16$. The observation process for a_{que} and a_{nul} were always high information gain (HI) and low information gain (LI) respectively, as shown in Fig. 2, with the exception of the Fig. 4, where we vary the information gain of a_{que}. For the HI process, the means of observation distributions for $s = 0$ and $s = 1$, were apart by 15 units and the observation distributions had the std. dev. of 15 units. For the LI process, the means of observation distributions for $s = 0$ and $s = 1$, were apart by 5 units and the observation distributions had the std. dev. of 30 units. Action a_{nul} was omitted from Normative Consequences results for simplicity but included in Dysfunctional Consequences results. The maximum trials within a POMDP episode were set to 100. Results were averaged over 1000 independent POMDP episodes for a fixed utility function. The discount factor was set to $\gamma = 1$.

Definition of phasic and tonic pain in our model simulations

In the first instance, we define phasic pain based on equation 4, as belief-dependent negative feedback (expected negative costs of an action), only when concerning the pain/injury systems.

$$\text{Phasic Pain} \propto \sum_{s \in \mathcal{S}} b(s)\zeta(s, a) \tag{10}$$

Here, $\zeta(s, a)$ is the negative rewards or punishments concerning the pain systems and excludes other kinds of punishments such as opportunity costs or loss in resources. In our simulations, we only require and report the phasic pain for action a_{que}, where the $\zeta(s, a_{\text{que}}) = r(s, a_{\text{que}})$.

We define tonic pain to be proportional to the belief-weight drive in the injury state space. This further contributes to the affective value, which is a function of the belief state $V^{\pi}(b)$.

$$\text{Tonic Pain} \propto b(s = 1) \tag{11}$$

5 Acknowledgments

PM would like to thank Michael Browning, Rafal Bogacz, Suyi Zhang, Charlie Yan, Shuangyi Tong, Danielle Hewitt for their feedback on presentations or earlier drafts of the manuscript. The work was funded by Wellcome Trust (214251/Z/18/Z, 203139/Z/16/Z and 203139/A/16/Z), IITP (MSIT 2019-0-01371) and JSPS (22H04998). This research was also partly supported by the NIHR Oxford Health Biomedical Research Centre (NIHR203316). PD is funded by the Max Planck Society and the Humboldt Foundation. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

Deborah Antcliff, Philip Keeley, Malcolm Campbell, Steve Woby, and Linda McGowan. Exploring patients' opinions of activity pacing and a new activity pacing questionnaire for chronic pain and/or fatigue: a qualitative study. *Physiotherapy*, 102(3):300–307, 2016.

Arnoud Arntz, Michael Rauner, and Marcel Van den Hout. "if i feel anxious, there must be danger": Ex-consequentia reasoning in inferring danger in anxiety disorders. *Behaviour research and therapy*, 33(8):917–925, 1995.

Karl J Astrom et al. Optimal control of markov processes with incomplete state information. *Journal of mathematical analysis and applications*, 10(1):174–205, 1965.

Tali M Ball and Lisa A Gunaydin. Measuring maladaptive avoidance: from animal models to clinical anxiety. *Neuropsychopharmacology*, pages 1–9, 2022.

Kirsty Bannister and AH Dickenson. The plasticity of descending controls in pain: translational probing. *The Journal of physiology*, 595(13):4159–4166, 2017.

Andrew G Barto and Ozgür Simsek. Intrinsic motivation for reinforcement learning systems. In *Proceedings of the thirteenth yale workshop on adaptive and learning systems*, pages 113–118. Yale University Press, New Haven, CO., 2005.

Allan I Basbaum and Howard L Fields. Endogenous pain control mechanisms: review and hypothesis. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 4(5): 451–462, 1978.

Richard Bellman. The theory of dynamic programming. *Bulletin of the American Mathematical Society*, 60(6):503–515, 1954.

Robert C Bolles and Michael S Fanselow. A perceptual-defensive-recuperative model of fear and pain. *Behavioral and Brain Sciences*, 3(2):291–301, 1980.

Christian Büchel, Stephan Geuter, Christian Sprenger, and Falk Eippert. Placebo analgesia: a predictive coding perspective. *Neuron*, 81(6):1223–1239, 2014.

Nuttapong Chentanez, Andrew Barto, and Satinder Singh. Intrinsically motivated reinforcement learning. *Advances in neural information processing systems*, 17, 2004.

Geert Crombez, Christopher Eccleston, Stefaan Van Damme, Johan WS Vlaeyen, and Paul Karoly. Fear-avoidance model of chronic pain: the next generation. *The Clinical journal of pain*, 28(6):475–483, 2012.

P Dayan, JP Roiser, and E Viding. The first steps on long marches: The costs of active observation. In *Psychiatry Reborn: Biopsychosocial psychiatry in modern medicine*, pages 213–228. Oxford University Press, 2020.

Peter Dayan. "liking" as an early and editable draft of long-run affective value. *PLoS Biology*, 20(1):e3001476, 2022.

Peter Dayan and Nathaniel D Daw. Decision theory, reinforcement learning, and the brain. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4):429–453, 2008.

Dominique Debanne. Information processing in the axon. *Nature Reviews Neuroscience*, 5(4):304–316, 2004.

Alvin W Drake. *Observation of a Markov process through a noisy channel*. PhD thesis, Massachusetts Institute of Technology, 1962.

Mary-Ann Fitzcharles, Steven P Cohen, Daniel J Clauw, Geoffrey Littlejohn, Chie Usui, and Winfried Häuser. Nociplastic pain: towards an understanding of prevalent pain conditions. *The Lancet*, 397(10289):2098–2110, 2021.

Ida Landström Flink, Katja Boersma, and Steven J Linton. Pain catastrophizing as repetitive negative thinking: a development of the conceptualization. *Cognitive behaviour therapy*, 42(3):215–223, 2013.

Karl J Friston, A David Redish, and Joshua A Gordon. Computational nosology and precision psychiatry. *Computational Psychiatry (Cambridge, Mass.)*, 1:2, 2017.

Chris Gagne and Peter Dayan. Peril, prudence and planning as risk, avoidance and worry. *Journal of Mathematical Psychology*, 106:102617, 2022.

Alan Gordon and Alon Ziv. *The way out: A revolutionary, scientifically proven approach to healing chronic pain*. Penguin, 2021.

Milos Hauskrecht. Value-function approximations for partially observable markov decision processes. *Journal of artificial intelligence research*, 13:33–94, 2000.

Marion Höfle, Michael Hauck, Andreas K Engel, and Daniel Senkowski. Pain processing in multisensory environments. *e-Neuroforum*, 1:23–28, 2010.

Quentin JM Huys, Marc Guitart-Masip, Raymond J Dolan, and Peter Dayan. Decision-theoretic psychiatry. *Clinical Psychological Science*, 3(3):400–421, 2015.

Marieke Jepma, Leonie Koban, Johnny van Doorn, Matt Jones, and Tor D Wager. Behavioural and neural evidence for self-reinforcing expectancy effects on pain. *Nature human behaviour*, 2(11):838–855, 2018.

Leslie Pack Kaelbling, Michael L Littman, and Anthony R Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1-2):99–134, 1998.

Mehdi Keramati and Boris Gutkin. Homeostatic reinforcement learning for integrating reward collection and physiological stability. *Elife*, 3:e04811, 2014.

Joohan Kim, Jorge E Esteves, Francesco Cerritelli, and Karl Friston. An active inference account of touch and verbal communication in therapy. *Frontiers in Psychology*, 13, 2022.

Mykel J Kochenderfer, Tim A Wheeler, and Kyle H Wray. *Algorithms for decision making*. MIT press, 2022.

Kika Konstantinou, Kate M Dunn, Reuben Ogollah, Martyn Lewis, Danielle van der Windt, Elaine M Hay, ATLAS Study Team, et al. Prognosis of sciatica and back-related leg pain in primary care: the atlas cohort. *The Spine Journal*, 18 (6):1030–1040, 2018.

William S Lovejoy. Computationally feasible bounds for partially observed markov decision processes. *Operations research*, 39(1):162–175, 1991.

Pranav Mahajan, Shuangyi Tong, Sang Wan Lee, and Ben Seymour. Balancing safety and efficiency in human decision making. *bioRxiv*, pages 2024–01, 2024.

Flavia Mancini, Suyi Zhang, and Ben Seymour. Computational and neural mechanisms of statistical pain learning. *Nature Communications*, 13(1):6613, 2022.

Christoph Mathys et al. How could we get nosology from computation. *Comput Psychiatry New Perspect Ment Illn*, 20: 121–38, 2016.

Zoe McParlin, Francesco Cerritelli, Giacomo Rossettini, Karl J Friston, and Jorge E Esteves. Therapeutic alliance as active inference: The role of therapeutic touch and biobehavioural synchrony in musculoskeletal care. *Frontiers in Behavioral Neuroscience*, page 224, 2022.

Ann Meulders. From fear of movement-related pain and avoidance to chronic pain disability: a state-of-the-art review. *Current Opinion in Behavioral Sciences*, 26:130–136, 2019.

G Lorimer Moseley. A pain neuromatrix approach to patients with chronic pain. *Manual therapy*, 8(3):130–140, 2003.

Stéphane Ross, Joelle Pineau, Sébastien Paquet, and Brahim Chaib-Draa. Online planning algorithms for POMDPs. *Journal of Artificial Intelligence Research*, 32:663–704, 2008.

Ben Seymour. Pain: a precision signal for reinforcement learning and control. *Neuron*, 101(6):1029–1041, 2019.

Ben Seymour and Flavia Mancini. Hierarchical models of pain: Inference, information-seeking, and adaptive control. *NeuroImage*, 222:117212, 2020.

Ben Seymour, RJ Dolan, et al. Emotion, motivation, and pain. *Textbook of Pain*, pages 248–255, 2013.

Ben Seymour, Robyn J Crook, and Zhe Sage Chen. Post-injury pain and behaviour: a control theory perspective. *Nature Reviews Neuroscience*, pages 1–15, 2023.

Satinder Singh, Richard L Lewis, and Andrew G Barto. Where do rewards come from. In *Proceedings of the annual conference of the cognitive science society*, pages 2601–2606. Cognitive Science Society, 2009.

Richard D Smallwood and Edward J Sondik. The optimal control of partially observable markov processes over a finite horizon. *Operations research*, 21(5):1071–1088, 1973.

Edward J Sondik. The optimal control of partially observable markov processes over the infinite horizon: Discounted costs. *Operations research*, 26(2):282–304, 1978.

Edward Jay Sondik. *The optimal control of partially observable Markov processes*. Stanford University, 1971.

Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

Richard S Sutton, Joseph Modayil, Michael Delp, Thomas Degris, Patrick M Pilarski, Adam White, and Doina Precup. Horde: A scalable real-time architecture for learning knowledge from unsupervised sensorimotor interaction. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pages 761–768, 2011.

Stefaan Van Damme, Geert Crombez, and Chris Eccleston. Retarded disengagement from pain cues: the effects of pain catastrophizing and pain expectancy. *Pain*, 100(1-2):111–118, 2002.

Stefaan Van Damme, Geert Crombez, and Christopher Eccleston. Disengagement from pain: the role of catastrophic thinking about pain. *Pain*, 107(1):70–76, 2004.

Christine M van Vliet, Ann Meulders, Linda MG Vancleef, and Johan WS Vlaeyen. The opportunity to avoid pain may paradoxically increase fear. *The Journal of Pain*, 19(10):1222–1230, 2018.

Johan WS Vlaeyen and Steven J Linton. Fear-avoidance and its consequences in chronic musculoskeletal pain: a state of the art. *Pain*, 85(3):317–332, 2000.

Johan WS Vlaeyen, Geert Crombez, and Steven J Linton. The fear-avoidance model of pain. *Pain*, 157(8):1588–1589, 2016.

Patrick CAJ Vroomen, MCTFM De Krom, and JA Knottnerus. Predicting the outcome of sciatica at short-term follow-up. *British Journal of General Practice*, 52(475):119–123, 2002.

Patrick D Wall. Three phases of evil: the relation of injury to pain. *Brain and Mind*, 69:293, 1979.

Katja Wiech. Deconstructing the sensation of pain: The influence of cognitive processes on pain perception. *Science*, 354(6312):584–587, 2016.

Amanda C de C Williams. Persistence of pain in humans and other mammals. *Philosophical Transactions of the Royal Society B*, 374(1785):20190276, 2019.

Appendix

Our mathematical model complements the Fear-Avoidance model of pain chronification

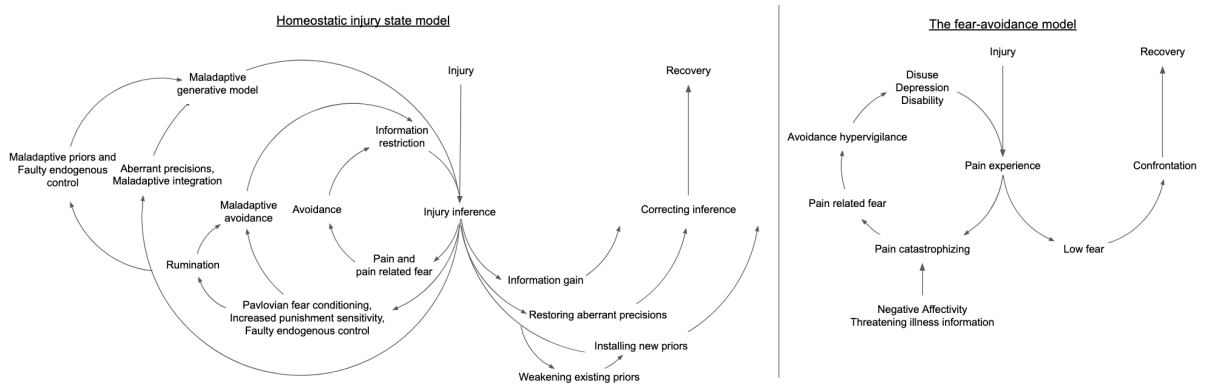


Figure 7: Comparison of predictions of our homeostatic injury state model to that of the Fear Avoidance model and how they play a complementary role in educating patients about their pain.