Altered Decision-Making Across Acute and Chronic Pain States: A Bias Towards Context-Dependent Valuations

Chad C. Williams[1,2*†], Lucy L.W. Owen[1,3†], Chloe Gunsilius[1], Matthew R. Nassar[1,4], Frederike H. Petzschner[1,5*]

[1*]Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, RI, USA.
[2]Department of Psychology, University of Victoria, Victoria, BC, Canada.
[3]Department of Computer Science, University of Montana, Missoula, MT, USA.
[4]Department of Neuroscience, Brown University, Providence, RI, USA.
[5]Department for Cognitive and Psychological Sciences, Brown University, Providence, RI, USA.

*Corresponding author(s). E-mail(s): chadwilliams@uvic.ca; frederike_petzschner@brown.edu;
†These authors contributed equally to this work.

Abstract

Chronic pain has been linked to changes in learning and decision-making, but the underlying cognitive mechanisms remain unclear. In this study, 239 participants with either chronic, acute, or no pain completed a reinforcement learning task designed to dissociate choices based on global expected value from those driven by context-specific reinforcement history. While learning performance was comparable across groups, individuals with chronic pain relied more on recent reinforcement history and less on integrated value representations during decision-making. Computational modeling confirmed that a mixture model best captured behavior across groups, with the chronic pain group assigning significantly less weight to global expected values. The acute pain group showed an intermediate pattern, suggesting a potential transitional phenotype. This shift in valuation strategy was associated with the duration of pain rather than its intensity, suggesting that prolonged exposure to pain may gradually alter

decision-making processes. These findings indicate a shift in valuation strategy associated with chronic pain, which may contribute to the persistence of maladaptive behaviors and offer a cognitive marker for pain chronification.

Keywords: Pain, Reinforcement Learning, Context, Basal Ganglia, Prefrontal Cortex, Actor-Critic, Q-Learning, Expected Value, Relative Value

Introduction

Pain is among the most powerful teaching signals available to an organism. It serves a critical adaptive function by signaling actual or potential tissue damage and promoting behaviors that safeguard the body. However, when pain persists beyond normal tissue healing time—typically more than three to six months—it becomes chronic and often maladaptive. Chronic pain now affects as many as 30% of adults globally and is a leading cause of long-term disability [1–3], yet the neurocognitive mechanisms that underlie the transition from acute to chronic pain remain incompletely understood.

Prominent psychological models, such as the fear-avoidance model, suggest that chronic pain is maintained through maladaptive reinforcement cycles [4, 5]. In this view, behaviors that reduce pain in the short term—such as inactivity, withdrawal, or hypervigilance—are reinforced during the acute pain state and may continue to persist even when no longer adaptive; for example, when movement is no longer harmful and would in fact support recovery and re-engagement with daily activities. Over time, this reinforcement structure can entrench avoidance patterns, reduce behavioral exploration, and contribute to functional disability and emotional distress [6, 7]. Although these models offer a compelling behavioral framework, they remain underspecified regarding the underlying learning mechanisms and neural processes that drive this overreliance on context-bound behaviors once the original pain state has changed.

Recent advances in neuroimaging have begun to illuminate how chronic pain might emerge from changes in the brain's reward and valuation systems. In particular, corticostriatal circuits—especially the functional connectivity between the nucleus accumbens and the ventromedial prefrontal cortex—have been shown to predict the transition from subacute to chronic pain in individuals with back pain [8–10]. These circuits are central to reward-based learning and value-based decision-making [11, 12], raising the possibility that pain chronification reflects a disruption in how the brain evaluates outcomes and makes decisions.

Within the framework of reinforcement learning, decision-making is often conceptualized as relying on at least two complementary computational processes [11–14]. The first centers on learning from prediction errors—discrepancies between expected and received outcomes—that drive adaptive changes in behavior based on local reinforcement history [13, 14]. This form of learning has been closely linked to phasic dopaminergic signaling and basal ganglia function, particularly in the striatum [11, 15, 16]. Whereby the striatum is thought to employ an actor-critic learning system, in which actions are evaluated relative to available alternatives, reinforcing those

that yield better-than-expected outcomes—even if their absolute expected value is low [17–20].

The second process involves the formation of expected value representations, which summarize the long-term consequences of actions by integrating across past outcomes [11, 15, 16]. These value representations have been observed in multiple brain regions, including the striatum, orbitofrontal cortex, and ventromedial prefrontal cortex, though their precise computational roles and timescales may differ [11, 12, 21, 22]. Importantly, these processes are not mutually exclusive and likely interact dynamically: prediction errors may update both local contextual representations and more stable value estimates over time. A balanced interplay between PE-driven, context-sensitive reinforcement and more stable value representations is thought to be essential for adaptive, goal-directed behavior.

Emerging evidence suggests that this balance may be disrupted in chronic pain. Specifically, individuals with chronic pain may become disproportionately reliant on the recent reinforcement history, driven by prediction errors—reinforcing actions that provide immediate relief—at the expense of forming stable value expectations that support long-term behavioral planning. This bias may help explain why behaviors acquired during acute pain, such as rest or avoidance, persist even after tissue healing. Supporting this view, Löffler et al. [10] found that individuals with a higher likelihood of pain chronification exhibited altered striatal responses to reward prediction errors and disrupted corticostriatal dynamics during a reinforcement learning task, suggesting a shift in learning strategy that favors short-term reinforcement over integrated value-based decision-making [10].

Here, we directly test this hypothesis by examining how individuals in acute and chronic pain states differ from pain-free individuals in reward-based decision-making. To this end, we employ computational modeling approaches in conjunction with an adapted probabilistic reinforcement learning task specifically designed to dissociate two choice strategies: one based on absolute expected value, and another based on context-dependent relative value [23–25]. Whereas absolute value reflects the overall expected utility of a stimulus across contexts, relative value in this task is implicitly learned through recent reinforcement history—specifically, the frequency of positive prediction errors, that is, how often a choice led to better-than-expected outcomes within its original context. Relative valuation is then a proxy for context-bound learning, shaped by local reinforcement dynamics rather than integrated global value.

During the learning phase, participants learned to associate specific stimuli with probabilistic outcomes in either a reward-seeking or punishment-avoidance context. In the subsequent transfer phase, participants were asked to choose between novel combinations of previously learned stimuli, presented without feedback. These novel pairings allowed us to probe whether decisions were guided by globally integrated value representations or, instead, by reinforcement signals tied to local context.

We hypothesized that individuals with chronic pain would rely more heavily on relative value, reflecting increased sensitivity to recent context and prediction errors. In contrast, individuals without pain would be more likely to base decisions on stable, absolute expected value representations. Such a shift in valuation strategy would support a reinforcement learning account of chronic pain and offer mechanistic insight

into how altered decision-making may contribute to the persistence of pain-related behaviors.

Results

To investigate how reward-based decision making varies among individuals in different pain states, we adapted a well-established probabilistic two-armed bandit task previously used to dissociate whether participants' choices reflect absolute expected value integration or are instead shaped by local context associations and reinforcement dynamics [24, 25]. In our study, a modified version of this task was administered to 239 individuals experiencing either chronic ($n = 78$), acute ($n = 53$), or no pain ($n = 108$), allowing us to systematically examine how pain states influence learning and value-based choice behavior (Figure 1a).

During the initial learning phase of the task, participants repeatedly selected between fixed pairs of stimuli presented in either a reward context or a punishment-avoidance context. Context was signalled in two ways: a background image indicating either a forest or a desert scene, and the unique pairing of two stimuli (Figure 1b,c). Within each context, one stimulus was objectively more advantageous than the other. Specifically, participants learned to choose between a stimulus with a high probability of reward (75%, high reward rate, HR) and a low probability of reward (25%, low reward rate, LR) in the reward seeking context, or between a low-probability punishment stimulus (25% loss, low punishment rate, LP) and a high-probability punishment stimulus (75% loss, high punishment rate, HP) in the punishment avoidance context. These four stimulus types differed both in their absolute expected value and their relative standing within their local context (Figure 1d). During learning, participants received outcome feedback after each choice for both the chosen and unchosen stimuli. However, only the chosen option resulted in actual gains or losses.

After learning, participants entered a transfer phase in which all stimuli were recombined into novel pairings and presented without feedback. Critically, this phase was designed to reveal which aspects of the learning phase shaped choice behavior. Participants were now required to make decisions between stimuli that had never been directly compared, for instance, a high-reward rate stimulus versus a low-punishment rate stimulus.

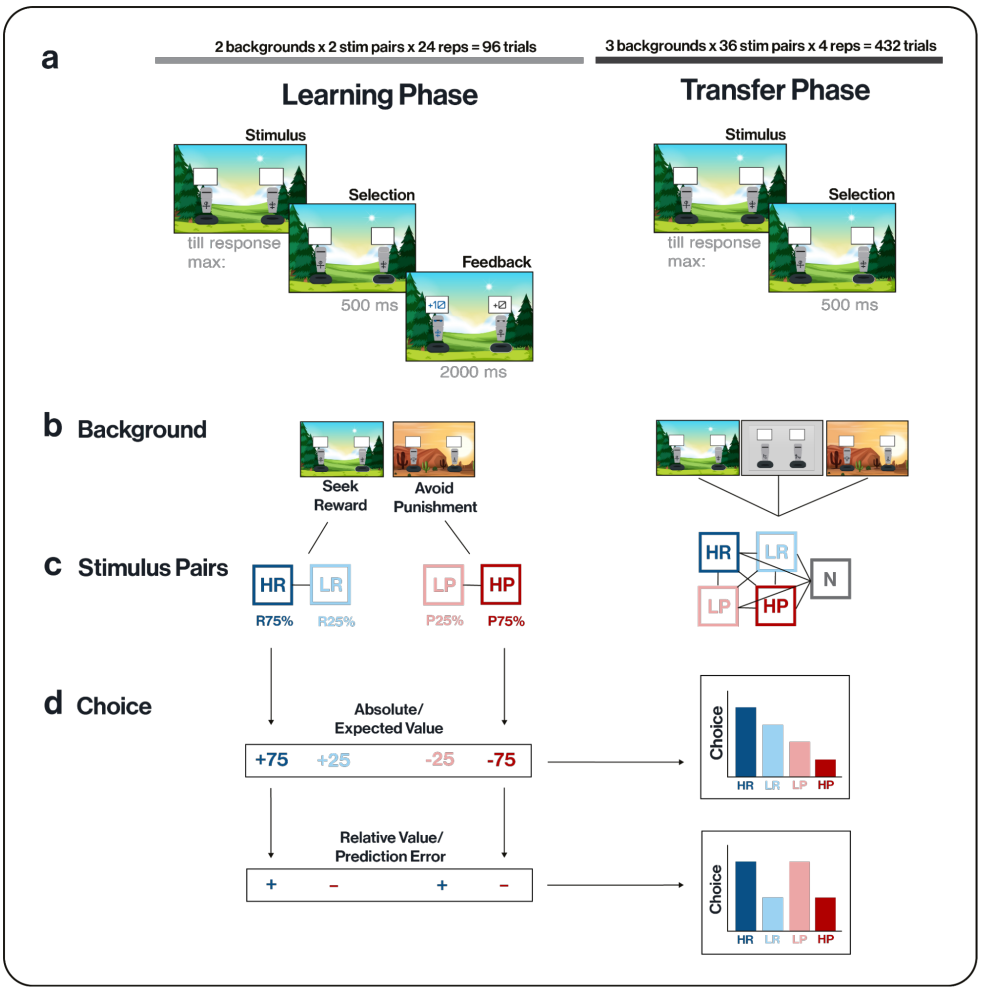

Fig. 1 Task structure. a. Trial Scheme: Each trial began with the presentation of a fixed stimulus pair (artificial knights dissociable by the symbol on their harness) and a contextual background image (forest or desert). Participants selected one of the two stimuli (knights) and received feedback indicating either reward or punishment, depending on the context. During the learning phase, feedback was provided after every choice for both the chosen and unchosen options, whereas in the transfer phase, participants made choices between novel pairings of previously learned stimuli without receiving feedback. b. Context Backgrounds: In the learning phase, each background image was uniquely associated with a stimulus pair type: i.e., the desert background signaled a reward-seeking context (2 unique pairs of low-reward rate [LR] and high-reward rate [HR] stimuli), while the forest background signaled a punishment-avoidance context (2 unique pairs of low-punishment rate [LP] and high-punishment rate [HP] stimuli). In the transfer phase, all possible pairwise combinations of the four stimulus types (HR, LR, LP, HP) and a never-seen novel stimulus (N) were presented across three background conditions: the original reward background, the original punishment background, and a neutral grey background. This design allowed us to assess whether contextual cues provided by visual background or by the original stimulus pairings continued to influence decision-making. c. Stimulus Pairs: During the learning phase, stimuli were paired exclusively within valence domains: HR and LR were always presented together in the reward context, and LP and HP were always paired in the punishment context (2 stimulus pairs per category). HR stimuli yielded rewards on 75% of trials, LR on 25%, and otherwise a neutral outcome. Conversely, HP stimuli resulted in losses on 75% of trials, while LP stimuli resulted in losses on 25% of trials. During the transfer phase, all previously experienced stimuli (2x HR, LR, LP, HP) and a previously unseen novel stimulus (N) were paired in combinations with one another, resulting in 36 possible unique stimulus pairings. d. Potential Choice Strategies: If participants based their choices on absolute expected values, preferences should follow the order HR > LR > LP > HP. However, if decision-making relied on context-specific relative value or the frequency of positive prediction errors (PEs) during learning, LP stimuli—although objectively worse than LR—might be preferred due to their relatively advantageous status within their original context and frequent positive prediction errors. The plot illustrates predicted choice probabilities for these competing strategies across critical stimulus comparisons. Abbreviations: HR – high reward rate (75% reward), LR – low reward rate (25% reward), LP – low punishment rate (25% loss), HP – high punishment rate (75% loss).

These novel pairings were constructed to dissociate two distinct decision-making strategies. One strategy involves selecting the option with the highest absolute expected value—a global representation learned across all contexts based on outcome probabilities. The alternative strategy relies on context-specific learning signals, such as the relative value of a stimulus within its original pair, i.e., based on the frequency of positive prediction errors it elicited during learning. For example, while the low-punishment rate (LP) stimulus had a negative absolute expected value, it was the relatively better option in its original punishment context and consistently produced positive prediction errors. While the low-reward rate stimulus (LR) had a positive absolute expected value, it was the relatively worse option in its original context and consistently produced negative prediction errors. As a result, whereas participants who rely on absolute values would prefer the LR stimulus over the LP stimulus, those whose decisions are guided more by relative value or prediction-error dynamics may instead favor the LP stimulus (Figure 1d).

Learning Performance Does Not Differ Across Pain Groups, but Chronic Pain is Associated with Slower Response Times

Participants across all three groups—chronic pain, acute pain, and no pain—demonstrated comparable learning performance during the initial phase of the task (Figure 2a). Accuracy improved with repeated exposure to stimulus pairs, and critically, no group differences emerged in the ability to learn optimal actions within either reward-seeking or punishment-avoidance contexts. A generalized linear mixed-effects model (GLMM) revealed a significant main effect of time ($p < .0001$), reflecting overall learning progression. However, there were no significant main effects of group ($p = .9192$) or context (reward seeking vs. punishment avoidance; $p = .1406$), nor evidence for interactions involving group and context ($ps > .14$), except for a modest group $\times$ time interaction ($p = .0037$), suggesting slight variation in learning curves over time (Supplementary Table S1). Planned comparisons confirmed the absence of group differences: neither the chronic pain group versus no pain ($p = .4730$, $d = -0.11$) nor chronic versus acute pain ($p = .7991$, $d = -0.05$) differed significantly. No effects were found when comparing reward and punishment contexts across groups ($ps > .05$; Supplementary Table S1).

By contrast, response times differed significantly across groups. Participants with chronic pain responded more slowly than both pain-free individuals and those with acute pain, despite equivalent accuracy (Supplementary Figure S1a). The GLMM revealed significant main effects of group ($p = .0441$), context ($p < .0001$), and time ($p < .0001$), as well as significant group $\times$ context ($p = .0056$) and group $\times$ time ($p = .0007$) interactions (Supplementary Table S2). Planned comparisons showed slower responses in the chronic pain group relative to both the no pain (chronic $>$ no pain: $p = .0020$, $d = 0.47$) and acute pain (chronic $>$ acute: $p = .0224$, $d = 0.41$) groups. These differences held across both reward ($p = .0011$, $d = 0.51$) and punishment ($p = .0083$, $d = 0.40$) contexts, although no differences were found in post hoc comparisons of reaction times between the acute and no pain groups ($p = .9047$, $d = -0.07$). In addition, overall reaction times were slower in the punishment than in the reward context and decreased across time.

Together, these findings support that basic learning performance is intact in individuals with both acute and chronic pain, and that the ability to learn to seek rewards or avoid punishments does not differ between pain groups. However, individuals with chronic pain exhibited significantly slower response times, suggesting alterations in decision speed rather than learning capacity.

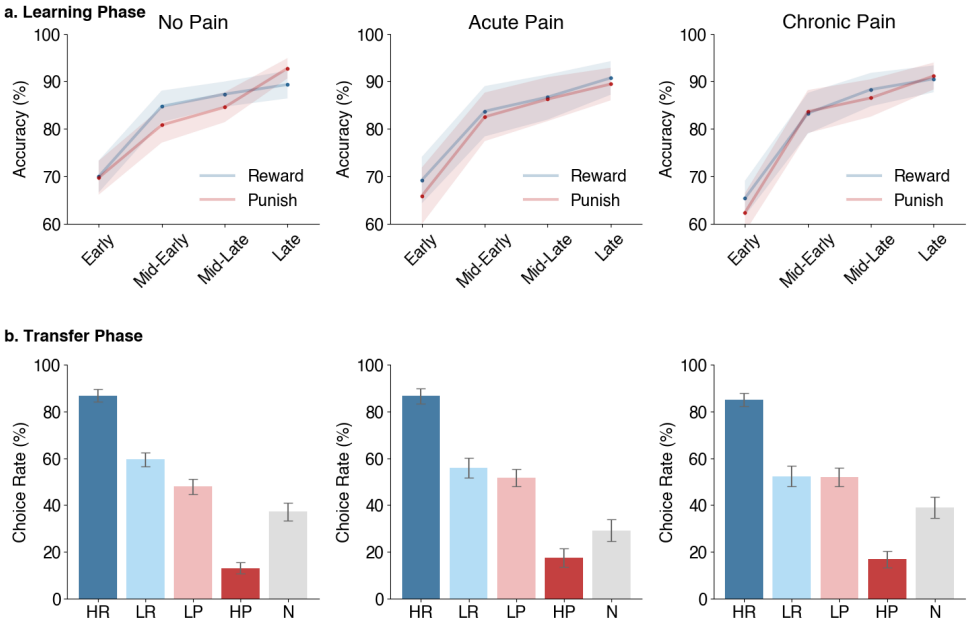

Fig. 2 Empirical findings of learning accuracy and transfer choice rates. a. Learning Phase: Behavioral performance across binned learning trials for the reward and punishment contexts for each group. Shaded regions represent 95% confidence intervals. b. Transfer Phase: Choice rates for each stimulus type during transfer trials for each group. Choice rate is computed as the percentage of times a stimulus type was chosen, given the number of times it was presented. Bar plots show the mean and 95% confidence intervals of the choice rate for each stimulus type across participants within each group. Abbreviations: HR – high reward rate (75% reward), LR – low reward rate (25% reward), LP – low punishment rate (25% loss), HP – high punishment rate (75% loss), N – novel stimulus.

Pain Groups Differ in Reward-Based Decision Making During the Transfer Phase

There are at least two distinct ways participants could have learned to make accurate choices during the learning phase: by encoding the relative values of options within each context (i.e., learning which option was better in a given pair), or by computing the overall expected value of individual stimuli across trials. The transfer phase was designed to disentangle these strategies. In this phase, previously learned stimuli—along with a novel stimulus—were presented in new pairings across various

background contexts, without feedback (Figure 1). Because options were now combined across contexts that differed in their overall reward structure, participants' choices revealed whether they had formed absolute value estimates (i.e. HR > LR > LP > HP) or relied on context-specific reinforcement (i.e. HR and LP > LR and HP) (Figure 1d). Thus the transfer phase allowed us to test our central hypothesis that valuation strategies differ across groups in different pain states.

Indeed, we find that participants with chronic pain relied more heavily on context-specific reinforcement history than on globally integrated expected value representations during the transfer phase. Choice behavior during the transfer phase varied across stimulus types, which significantly interacted with group identity. A generalized linear mixed-effects model (GLMM) revealed a robust main effect of stimulus type (HR, LR, LP, HP, Novel [N]; $p < .0001$), no main effect of group ($p = .1408$), but a significant group × stimulus type interaction ($p = .0002$) (Figures 2b and 3; Supplementary Figure S2; Table 1). Planned comparisons indicated that the chronic pain group differed significantly in their choice preferences from both the no pain group ($p = .0136$, $d = -0.37$) and the acute pain group ($p = .0420$, $d = -0.37$) (Table 1).

These effects remained robust when controlling for background context (i.e., the visual background image, Supplementary Table S3). This lack of influence from the visual context images suggests that the observed group differences in valuation strategies cannot be explained by differential incorporation of contextual features into the state representation and suggests that participants' decisions were driven by previously encountered stimulus evaluations rather than contextual cues. Moreover, performance differences were consistent across the first and second halves of the transfer phase, suggesting that valuation strategies remained stable over time and were not confounded by fatigue or task disengagement.

Reaction time patterns during the transfer phase paralleled the main effects in choice behavior. While there was no main group effect on overall reaction times ($p = .8893$), there was a significant main effect of stimulus type ($p < .0001$) and a group × stimulus type interaction ($p < .0001$; see Supplementary Figure S1b and Supplementary Table S4).

Overall, these results indicate that the chronic pain group adopted a different valuation strategy during the transfer phase—favoring context-specific reinforcement over integrated value representations—despite similar learning performance across groups.

Choices in Chronic Pain Are Biased Towards Context-Dependent Reinforcement History

To determine whether distinct valuation strategies drove the observed group differences during transfer, we focused on stimulus pairings that dissociate absolute expected value from context-specific reinforcement history. Specifically, we analyzed choices and reaction times for HR vs. LP and LR vs. LP stimulus pairs.

The HR vs. LP contrast compares two stimuli that were both the better option within their original learning contexts and thus share similar reinforcement histories (i.e., frequent positive prediction errors). However, they differ substantially in absolute value: HR carries a positive expected value, whereas LP has a negative one (Figure

1d). A strong preference for HR over LP reflects sensitivity to global expected value independent of reinforcement history.

Across groups, participants showed a clear preference for HR over LP, indicating a shared sensitivity to absolute expected value when reinforcement history was similar. Across groups, we found no significant differences in HR vs. LP choice rates, suggesting that all participants relied to some extent on absolute expected value when reinforcement history was similar. Specifically, planned comparisons of choice behavior did not differ between the no pain and acute pain groups (p = .7294, d = 0.06), the acute and chronic pain groups (p = .5114, d = 0.12), or the no pain and chronic pain groups (p = .2435, d = 0.17) (Table 1). Similarly, reaction time data of the HR vs. LP contrast revealed no group differences in choice speed between the no pain and acute pain groups (p = .3361, d = -0.18), the no pain and chronic pain groups (p = .2966, d = -0.17), and the acute and chronic pain groups (p = .9095, d = 0.02) (Supplementary Table S4).

The LR vs. LP comparison contrasts opposing reinforcement histories and value profiles. LR has a higher expected value but was the worst option within its original reward context, typically associated with negative prediction errors. In contrast, LP has a lower expected value but was contextually positively reinforced. Preference for LR over LP reflects sensitivity to global expected value; preference for LP over LR reflects reliance on context-specific reinforcement.

Here, group differences in choices were evident. Participants in the chronic pain group were significantly less likely to prefer LR over LP compared to the no pain group (chronic $<$ no pain: p = .0062, d = -0.41), indicating greater reliance on reinforcement history. The acute pain group did not differ significantly from either comparison group (acute vs. no pain: p = .1429, d = -0.25, acute vs. chronic pain: p = .3136, d = 0.18), but had choice rates that were numerically between the chronic pain and no pain group (Figure 3). Post hoc analyses showed that only the no pain group reliably preferred LR over LP ($p < .0001$, $d = 0.44$), while both pain groups treated the options similarly (acute: p = .3693, d = 0.22; chronic: p = 1.0000, d = 0.02). In contrast, there were no effects of reaction times when comparing the LR to LP stimulus types (Supplementary Table S4). Specifically, there was no difference between the no pain and the acute pain groups (p = .3052, d = -0.18), the no pain and chronic pain groups (p = .1400, d = -0.23), or the acute pain and chronic pain groups (p = .7528, d = -0.06).

Together, these findings indicate a progressive shift from expected value–driven decision-making in pain-free individuals toward a greater influence of reinforcement history–based valuation in individuals with chronic pain, with the acute pain group occupying a transitional position between the two. Furthermore, this shift occurred independently of response time differences, thus reflecting a genuine shift in valuation strategies rather than changes in processing speed.

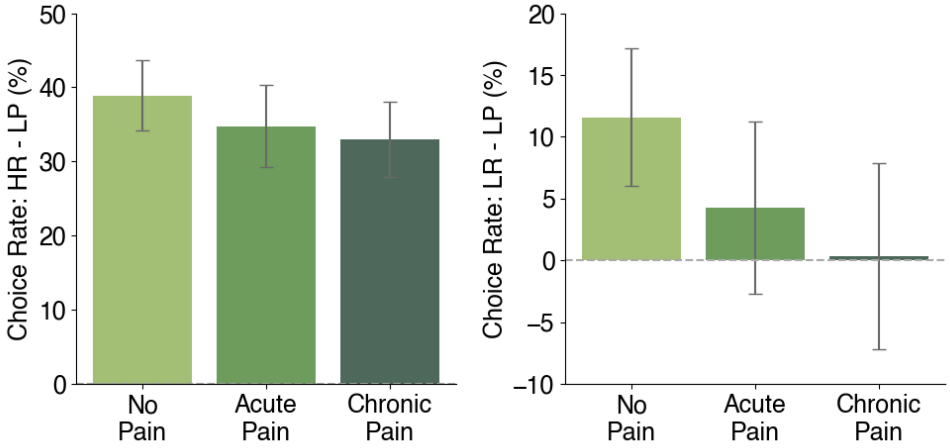

Fig. 3 Choice rate differences for critical analyses assessing valuation strategies. Difference in choice rates (%) comparing the high reward (HR) with the low punish (LP) stimulus type and the low reward (LR) with the low punish (LP) stimulus type across the groups. Choice rate is here computed as the number of times that the stimulus type was selected, given the number of times it was presented in its pair. This was computed for each stimulus in the pair (e.g., HR and LP for the left plot and LR and LP for the right plot), and the difference of these is plotted here. Bar plots show the mean and 95% confidence intervals of the difference in choice rates.

Table 1 Analysis of Choice Rates During the Transfer Phase. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for choice rates in the transfer phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. The planned comparisons of the group x stimulus type interaction use slightly different data than the GLMMs and the other planned comparisons. Specifically, these assess choice rates for each pair (e.g., HR vs LP) only when presented in that pair (rather than choice rates across all pairs). **Bold p-values** indicate statistically significant effects at $\alpha = .05$.

Effect	df	Test Value	p-value	Cohen's d
Group	2	3.92	.1408	
Chronic Pain Vs No Pain	184	-2.49	**.0136**	-0.37
Chronic Pain Vs Acute Pain	129	-2.05	**.0420**	-0.37
Stimulus Type	4	586.29	**<.0001**	
High Reward Vs Low Punish	238	24.19	**<.0001**	1.56
Low Reward Vs Low Punish	238	3.26	**.0013**	0.21
Group:Stimulus Type	8	29.80	**.0002**	
No Pain Vs Acute Pain: HR-LP	159	0.35	.7294	0.06
No Pain Vs Chronic Pain: HR-LP	184	1.17	.2435	0.17
Acute Pain Vs Chronic Pain: HR-LP	129	0.66	.5114	0.12
No Pain Vs Acute Pain: LR-LP	159	1.47	.1429	0.25
No Pain Vs Chronic Pain: LR-LP	184	2.77	**.0062**	0.41
Acute Pain Vs Chronic Pain: LR-LP	127	1.01	.3136	0.18

Decision-Making Across Pain States Is Best Explained by a Mixture Model of Expected Value and Reinforcement History

While behavioral analyses identified group-level differences in choice patterns, computational modeling provides a more granular account of the cognitive mechanisms that shape these decisions. To this end, we fit five reinforcement learning models to participants' trial-by-trial choices across both the learning and transfer phases. These models systematically varied in how they weighted global expected value versus local context-specific reinforcement history (see Methods for full specifications and Supplementary Information for model validations).

The Q-Learning model captures decision-making based on the integration of past outcomes into an absolute expected value for each option (pure expected value strategy) [17, 26]. By contrast, the Actor-Critic model distinguishes between state and action learning, emphasizing recent reinforcement signals without forming stable value representations (pure reinforcement history strategy) [17, 27]. The remaining three models implement mixture strategies that combine global and local learning mechanisms typically captured by a single model parameter (mixture models). The Hybrid model uses a weighted combination of Q-Learning and Actor-Critic mechanisms [23, 24]. The Relative model integrates Q-values with within-context value normalization, capturing the dynamic scaling of preferences based on local context [25]. Finally, the Advantage model, which we introduce here, simplifies this approach by incorporating both absolute Q-values and a context-dependent advantage term that reflects reward comparisons within each pair. Thus, our model set contained one model that learned global values (Q-Learning), one that learned relative values (Actor-Critic), as well as three that enabled a mixture of global and relative value learning (Hybrid, Relative, Advantage).

Model comparison using the Bayesian Information Criterion (BIC) revealed that the Advantage model performed best across all three groups (Figure 4, see Supplementary Table S5 for fitted values). This model dynamically integrates global expected value estimates with recent context-specific reinforcement, allowing it to flexibly capture shifts in valuation strategy. Posterior predictive checks indicate that the model replicates the key behavior findings for both the learning and transfer phases of the task across all groups (Figure 5a,b; Supplementary Tables S6 and S7). The following results focus on group differences in key parameters from the Advantage model (but see Supplementary Information for results from the Relative and Hybrid models).

Chronic Pain is Associated with Increased Weighting of Context-Dependent Relative Values in Decision-Making

To identify the mechanisms underlying group differences in choice behavior, we examined fitted parameters from the best-fitting Advantage model. Of all model parameters, only the weighting factor—which determines the relative influence of global expected value versus local context-specific reinforcement—differed significantly across groups ($p = .0270$) (Figure 5c, Supplemental Figure S3, and Table 2). Planned comparisons revealed that participants with chronic pain assigned significantly less weight to global expected value than those without pain ($p = .0088$, $d =$

–0.39), indicating a stronger reliance on relative context-dependent stimulus evaluations. The acute pain group did not differ significantly from either the no pain (p = .1394, d = –0.25) or chronic pain group (p = .6678, d = 0.15), again showing an intermediate pattern.

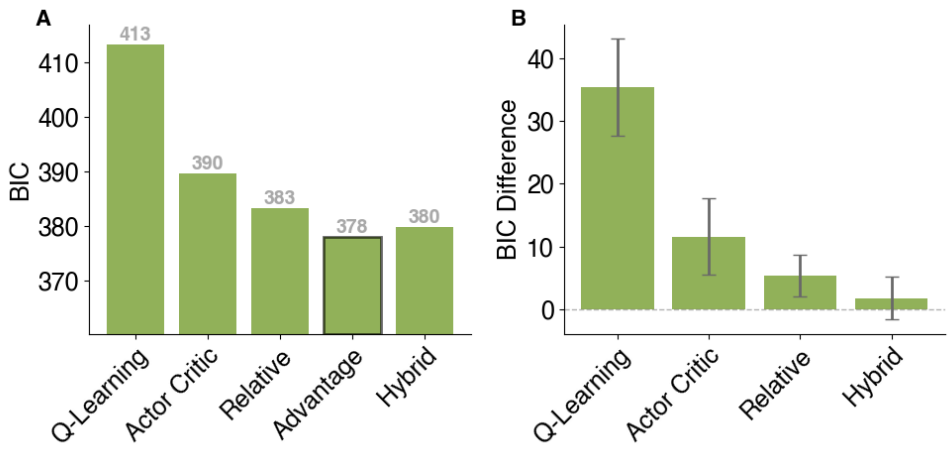

Fig. 4 Model selection using Bayesian information criterion (BIC). A: Grand averaged BIC metrics for each model across all participants. Lower values indicate better model fit. The black border indicates the best model. B: Difference in grand average BIC values for each model relative to the best model. Error bars represent 95% confidence intervals.

This computational result mirrors the empirical differences observed in LR vs. LP comparisons, suggesting a gradual shift from expected value–based to reinforcement history–driven valuation across the pain spectrum. The weighting factor from the Advantage model was highly correlated with analogous parameters in the other mixture models (Relative and Hybrid model), suggesting that this parameter reliably captures a generalizable decision strategy across model architectures (see Supplementary Information).

No other model parameters—including learning rates for chosen and unchosen options, softmax temperature, or valuation of the novel stimulus—differed significantly between groups ($ps > .05$, Table 2), underscoring the specificity of the weighting factor as the key computational mechanism underlying group differences in decision-making.

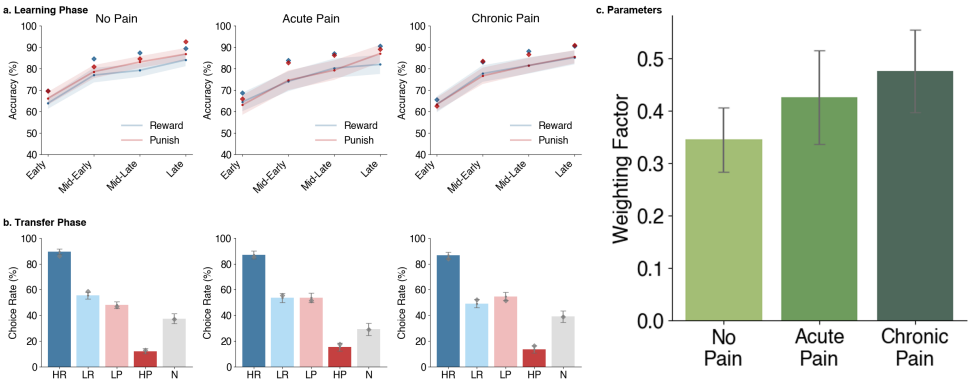

Fig. 5 Posterior Predictive checks of the best-fitted computational model and fitted parameters. Model simulations of the best-fitted model, the Advantage model. During the fitting procedure, each participant was fitted ten times, and the best of these runs was selected to determine the participant's fitted parameters. We then used each model to simulate a single dataset per participant using these fitted parameters. Thus, these data represent the same sample size as our empirical data. a. Learning Phase: Model performance across binned learning trials for the reward and punishment contexts for each group. Shaded regions represent 95% confidence intervals. Blue and red diamonds indicate empirical means of participant accuracy. b. Transfer Phase: Choice rates for each stimulus type during transfer trials for each group. Choice rate is computed as the percentage of times a stimulus type was chosen, given the number of times it was presented. Bar plots show the mean and 95% confidence intervals of the choice rate for each stimulus type across participants within each group. Grey diamonds indicate empirical means of participant choice rates. Abbreviations: HR – high reward rate (75% reward), LR – low reward rate (25% reward), LP – low punishment rate (25% loss), HP – high punishment rate (75% loss), N – novel stimulus. c. Parameters: Fitted values of the significant parameter from the Advantage model. The parameter is displayed as a bar plot, showing the mean and 95% confidence intervals across participants for each group.

Pain Duration, but not Severity, is Associated with Shifts in Valuation Strategy

The observed shift from expected value–based to reinforcement history–driven decision-making across pain groups appears to be primarily driven by the duration of pain rather than its subjective severity. Across participants, the weighting parameter from the best-fitting Advantage model—reflecting the relative reliance on global expected value versus local reinforcement history—was significantly associated with self-reported pain duration (ranging from 'I don't have pain' to 'pain lasting longer than 10 years'; $\rho = .17$, $p = .0086$) but not with pain severity metrics ($ps > .05$; for example, the weighting parameter was not significantly related to the composite pain score in the no pain [$r =$ -.12 $p = .2115$], chronic pain [$r =$ -.18 $p = .1893$], or chronic pain groups [$r =$ -.03 $p = .7972$], see Supplementary Table S8). Thus, prolonged exposure to pain, rather than its momentary severity, may gradually bias decision-making toward context-dependent reinforcement signals.

None of the other model parameters were related to pain duration (factual learning rate: $\rho =$ -.02, $p = .8175$; counterfactual learning rate: $\rho = .02$, $p = .7943$; temperature:

Table 2 Analysis of Group Differences in Model Parameters. Main effects of between-subject ANOVAs (**bold**) and planned comparisons (indented and *italicized*) for each parameter assessing group differences. Test value represents F-values for ANOVA effects and t-values for planned comparisons. **Bold p-values** indicate statistically significant effects at $\alpha = .05$.

Effect	df	Test Value	p-value	Cohen's d
Factual LR	2, 236	0.13	.8792	
No Pain Vs Acute Pain	159	0.50	.6143	0.08
No Pain Vs Chronic Pain	150	0.32	.7516	0.05
Counterfactual LR	2, 236	0.04	.9574	
No Pain Vs Acute Pain	159	0.01	.9897	0.00
No Pain Vs Chronic Pain	184	0.28	.7809	0.04
Temperature	2, 236	0.11	.8991	
No Pain Vs Acute Pain	159	-0.47	.6424	-0.08
No Pain Vs Chronic Pain	184	-0.05	.9590	-0.01
Weighting Factor	2, 236	3.67	**.0270**	
No Pain Vs Acute Pain	159	-1.49	.1394	-0.25
No Pain Vs Chronic Pain	184	-2.65	**.0088**	-0.39
Novel Value	2, 236	1.57	.2096	
No Pain Vs Acute Pain	159	1.57	.1192	0.26
No Pain Vs Chronic Pain	184	-0.27	.7848	-0.04

$\rho = .05$, $p = .4833$; novel value: $\rho = .01$, $p = .9133$). However, within the chronic pain group, higher levels of pain unpleasantness, interference, and composite symptom scores were positively correlated with both factual and counterfactual learning rates, as well as the softmax temperature parameter ($rs > .23$, all $ps < .05$; Supplementary Figure S4 and Supplementary Table S8). These associations were absent in the acute pain group (Supplementary Figure S5 and Supplementary Table S8), despite having the same average pain intensities as the chronic pain group (Supplementary Table S9).

Taken together, these findings suggest that while valuation strategy is modulated by the chronicity of pain, symptom severity in chronic pain may independently influence how strongly individuals update beliefs and how consistently they translate those beliefs into actions, contributing to more reactive and less stable decision-making profiles.

Discussion

Chronic pain systematically shifts reward-based decision-making. Compared to pain-free individuals, participants with chronic pain relied more on context-specific reinforcement history and less on globally integrated expected value in a probabilistic learning task. This behavioral pattern was reflected in choice preferences and supported by computational modeling. The best-fitting model, which dynamically integrates absolute and relative value signals, revealed that individuals with chronic pain assigned significantly less weight to global expected value representations. These findings suggest that chronic pain is accompanied by a shift in valuation strategy, favoring recent reinforcement signals over integrated value estimates.

This shift was not explained by current pain intensity or unpleasantness, but rather by the duration of time individuals had spent in pain, which was the defining factor distinguishing the acute and chronic pain groups. This effect was evident in the intermediate profile of the valuation bias in the acute pain group across both behavioural and modelling analyses, in the strong association between pain duration and the strength of this bias, and in the fact that other decision-making mechanisms—such as belief updating rates and response indecisiveness—were linked to pain severity only within the chronic pain group.

These findings suggest that chronic exposure, rather than momentary symptom severity, gradually biases learning and decision-making toward greater reliance on recent reinforcement history. While the acute pain group did not differ significantly from either the chronic pain or no pain groups, their pattern of performance suggests a potential transitional phenotype. Individuals in acute pain may already exhibit early signs of biased reinforcement-driven valuation, raising the possibility that changes in decision-making strategy could serve as a cognitive marker for risk of pain chronification.

Our findings are consistent with prior work linking chronic pain to disruptions in reward-based learning and corticostriatal circuit function. Löffler et al. [10] showed that individuals with subacute back pain who went on to develop chronic pain exhibited heightened prediction error signals in the nucleus accumbens (NAc) and increased functional connectivity between the NAc and the ventromedial prefrontal cortex (vmPFC) in response to negative prediction errors. These results build on earlier work demonstrating that vmPFC–NAc connectivity predicts the transition from acute to chronic pain [8, 28] and increased NAc dopamine levels in animal models of chronic pain [29], suggesting that corticostriatal circuits undergo plastic changes during this transition.

The shift in valuation strategies found with chronic pain aligns with neurocomputational models of decision-making that distinguish between two partially dissociable systems: a dopaminergic, striatum-based system that updates choices based on prediction errors (captured by Actor-Critic learning), and a prefrontal system, centered in the orbitofrontal cortex and vmPFC, that encodes global expected value (as in Q-Learning; [11, 12]). Computational mixture models, which best account for behavior across our groups, integrate these systems via different implementations of a dynamic weighting mechanism [23–25]. In these frameworks, individuals with chronic pain assigned less weight to prefrontal global value signals, suggesting a dominance of striatal relative value signals in guiding behavior.

Our findings also illustrate the utility of a computational psychiatry approach for dissociating valuation mechanisms in complex clinical states [30–34]. One mixture model best captured behavior across all three groups, wherein its weighting parameter—reflecting the relative influence of global versus recent reinforcement signals—systematically differentiated individuals with chronic, acute, and no pain. This parameter not only tracked group-level differences but also scaled with individual pain duration, suggesting a dimensional shift in valuation strategy that emerges with chronicity. Computational models have been increasingly used to isolate latent processes underlying altered learning and decision-making in psychiatric and neurological

disorders [35–40], offering interpretable, mechanistic insights that bridge behavior and brain function [41, 42]. By applying this framework to pain, our work highlights how computational signatures of altered valuation may not only reflect internal shifts in decision-making architecture but could also help explain the persistence of maladaptive behavioral patterns observed in chronic pain.

Mechanistically, such a shift in valuation strategy may contribute to the persistence of maladaptive behaviors in chronic pain, such as avoidance, withdrawal, or inactivity. In line with the fear-avoidance model of chronic pain [4, 5], these behaviors may initially be reinforced through short-term relief during acute pain, but persist beyond the healing phase, where they become decoupled from their negative long-term consequences. Specifically, if choices are dominated by prediction error–driven reinforcement over global value representations, actions that once provided immediate relief may continue to be selected even when they hinder recovery. Over time, this imbalance may entrench dysfunctional learning loops that resist change, even when avoidance is no longer adaptive, thereby exacerbating the chronic pain state.

Limitations

Several limitations of the present study warrant consideration. First, although our findings align with previous work on value-based learning, we did not replicate the exact behavioral patterns reported in earlier studies using a similar paradigm. For instance, compared to Palminteri et al. [25], we incorporated background images to signal context and used slightly modified task instructions, which may have influenced how participants interpreted feedback and their engagement with novel decision contexts. Prior research has demonstrated that valuation strategies are highly sensitive to framing, feedback structure, and instructional emphasis [43]. Subtle changes in task setup may shift the balance between absolute and relative value coding.

Second, although our study sampled individuals across distinct self-reported pain states (acute, chronic, and no pain), it was not designed to examine differences across specific chronic pain conditions. Chronic pain is a heterogeneous phenomenon with diverse underlying mechanisms—including inflammatory, neuropathic, and musculoskeletal subtypes—and our sample included individuals with varying etiologies [44–46]. Here, our primary goal was to test whether the duration of being in a state of pain, regardless of diagnosis, was associated with changes in decision-making. Future studies with more fine-grained phenotyping and larger diagnostic subgroups will be necessary to determine whether different types of chronic pain exhibit distinct reinforcement learning profiles.

Finally, our cross-sectional design does not allow us to determine whether changes in valuation strategy precede the development of chronic pain or emerge as a consequence of prolonged pain experience. The observed strong association between a shift in valuation strategy and the duration individuals have been in pain suggests that being in a prolonged state of pain may gradually reshape decision-making processes, potentially reinforcing maladaptive behaviors that increase the likelihood of continued pain. However, our findings cannot fully resolve the directionality of the observed effects. Longitudinal studies following individuals from acute injury through potential

pain chronification will be essential to test whether shifts in decision-making are predictive of pain persistence, or instead reflect an adaptive (or maladaptive) response to chronic pain itself.

Methods

Participants

Participants were recruited as part of a large-scale study on chronic pain. A total of 360 individuals enrolled via the online data acquisition platform Prolific. Given reports that approximately 30% of Prolific participants may be inattentive or non-human [47], we implemented strict performance-based exclusion criteria to ensure data quality. Specifically, participants were required to correctly answer several task-related comprehension questions following the instructions before proceeding to the experimental phase. In addition, participants were excluded if they failed to achieve an average accuracy of at least 70% during the final quarter of the learning phase (see Task section below), resulting in the exclusion of 57 participants (16%). A further 64 participants (18%) were excluded due to inconsistencies in their pain reports (see Pain Groups section below). The final sample comprised of 239 participants: 108 in the no pain group (age = 33.74 [SD: 11.34], female/male/undisclosed = 54/54/0), 53 in the acute pain group (age = 36.17 [SD: 13.64], f/m/u = 29/23/1), and 78 in the chronic pain group (age = 40.96 [SD: 12.69], f/m/u = 55/22/1). For behavioral analyses, we excluded trials in which participants responded in under 200 ms or over 5000 ms, resulting in the removal of 1.46% of all trials. Participants received monetary compensation for their participation. The study was approved by the local Ethics Committee (IRB No. 2022003301), and all participants provided informed consent in accordance with the Declaration of Helsinki (2013).

Demographics & Pain Groups

To assess the current pain status and severity, all participants completed a brief self-report survey modeled after the ICD-11 diagnostic guidance for chronic pain classification. The survey included three items, each rated on a 0–10 numerical scale, assessing (1) pain intensity ("Averaged over the past week, how intense is your pain?"), (2) pain unpleasantness ("Averaged over the past week, how unpleasant is your pain?"), and (3) pain interference ("How much has your pain interfered with your activities over the past week?"). These dimensions correspond to the ICD-11-recommended criteria for coding the severity of chronic pain, which emphasize the multidimensional nature of pain-related impairment, including physical intensity, emotional distress, and disruption of daily functioning [48, 49].

Participants also reported the duration of their current pain experience by selecting one of the following options: "I am not in pain," "< 2 weeks," "2–4 weeks," "1–3 months," "3–6 months," "6–12 months," "1–5 years," "> 5 years," or "> 10 years." Additionally, they indicated whether they had been diagnosed with any pain-related condition from a predefined list (e.g., scoliosis, fibromyalgia; Supplementary Table S10).

Based on their self-reported pain duration, participants were categorized into one of three pain groups: **no pain** ("I am not in pain"), **acute pain** (pain duration ≤ 6 months), and **chronic pain** (pain duration > 6 months). This duration-based definition of chronic pain follows standard clinical and research conventions, including those of the International Association for the Study of Pain (IASP [50]), which defines chronic pain as pain that persists beyond the expected period of tissue healing. This is typically considered three to six months after pain onset. Since some people still heal in the 3-6 months after pain started, we used a conservative cut-off duration of six months for defining the chronic pain group in this study [44, 45, 49].

To ensure internal consistency between pain duration and pain severity, we computed a composite pain score for each participant by averaging their ratings of intensity, unpleasantness, and interference. Participants in the no pain group with a composite score > 2 were excluded, as were participants in the acute or chronic pain groups with a composite score < 2. This procedure resulted in the exclusion of 64 participants from the final sample.

As per the definition, the pain groups differed significantly in their self-reported pain experiences. Robust group effects were observed for pain intensity, unpleasantness, interference, and the composite pain score ($ps < .0001$) (Supplementary Figure S6 and Supplementary Table S9). Post hoc comparisons revealed that participants in the chronic pain group reported significantly higher scores than those in the no pain group across all dimensions ($ps < .0001$; $ds = 3.45$ for intensity, 3.79 for unpleasantness, 3.22 for interference, and 3.80 for the composite score). No significant differences were found between the chronic and acute pain groups ($ps > .05$; $ds = 0.29$, 0.24, 0.08, and 0.22, respectively), indicating the same pain severity levels despite differences in chronicity.

In addition, there was a significant group difference in age ($p < .0001$), with participants in the chronic pain group being older than those in the no pain ($p < .0001$, $d = 0.61$) and acute pain ($p < .0001$, $d = 0.37$) groups. However, no age difference was found between the no pain and acute pain groups ($p = .4691$, $d = 0.20$).

To confirm that observed behavioral effects were not confounded by age or pain severity, we repeated all main behavioral analyses while including each of age and composite pain intensity as covariates. The results remained unchanged, indicating that the group effects observed in learning and decision-making were not attributable to these covariates, but rather to the pain state (acute versus chronic) itself (Supplementary Tables S3 and S11).

Task

Participants completed a contextual probabilistic two-armed bandit task with a transfer phase, adapted from prior reinforcement learning studies [23–25] (Figure 1a). The task was designed to dissociate two valuation strategies: decision-making based on absolute expected value versus context-specific reinforcement history. The task consisted of two phases: a learning phase and a transfer phase.

Learning Phase

During the learning phase, participants made choices between four fixed pairs of abstract visual stimuli ("knights"; i.e., A-B, C-D, E-F, G-H), each distinguishable by a unique chestplate symbol randomly selected from a predefined set of Agathodaimon characters. Each stimulus pair was consistently associated with one of two learning contexts: a reward-seeking context or a punishment-avoidance context. These contexts were signaled in two ways: (1) by the background image on which the pair appeared (either a forest or a desert scene, randomly assigned across participants), and (2) by the identity of the stimulus pair itself. Specifically, the A–B and E–F pairs always appeared in the reward context and involved potential gains, while the C–D and G–H pairs were always presented in the punishment context and involved potential losses. Thus, both the visual background and the stimulus pairing provided distinct cues about the current context.

Within each pair, one stimulus was consistently more advantageous than the other. In the reward context, the high reward rate stimulus (HR, e.g., A) delivered a gain (+10 points) on 75% of trials and zero otherwise, while the low reward rate stimulus (LR, e.g., B) delivered the same gain on only 25% of trials. In the punishment context, the high punishment rate stimulus (HP, e.g., C) resulted in a loss (–10 points) on 75% of trials and zero otherwise, whereas the low punishment rate stimulus (LP, e.g., D) produced the same loss on only 25% of trials. After each choice, participants received feedback on the outcomes of both options, though only the outcome of the chosen stimulus affected their earnings. This learning phase structure required participants to learn which option was relatively better within each pair (Figure 1b,c). Participants were informed that they would receive a monetary performance bonus at the end of the task based on their overall task performance.

Each of the four pairs was presented 24 times in a randomized order, yielding a total of 96 learning trials. On each trial, participants had up to 10 seconds to choose by pressing the 'F' key (for the left option) or the 'J' key (for the right). Upon selection, the chosen knight was briefly highlighted, followed by outcome feedback. Feedback was displayed above both options for 2000 ms, with outcome values color-coded to indicate valence: blue for gains, red for losses, and black for neutral outcomes. In addition to numeric feedback, visual cues on the knights (e.g., eye and symbol color) changed to match the outcome valence.

Transfer Phase

Following the learning phase, participants entered a transfer phase designed to assess how previously acquired values were generalized to novel decision contexts. In this phase, all eight stimuli from the learning phase (HR, LR, LP, HP from two pairs each) were recombined into every possible pairwise combination. In addition, a novel ninth stimulus—never seen during learning—was introduced and paired once with each of the original eight. This resulted in 36 unique stimulus pairs.

To test for contextual influences of the background, each of these 36 pairs was presented across three background conditions: (1) the background associated with the reward context, (2) the background associated with the punishment context, and

(3) a neutral grey background not previously used. This manipulation resulted in 108 unique trials (36 pairs × 3 contexts), each of which was repeated four times in randomized order, yielding a total of 432 transfer trials.

The sequence and timing of each transfer trial mirrored that of the learning phase: participants viewed a pair of stimuli and made a selection using the 'F' or 'J' key within a 10-second window. However, no feedback was provided during the transfer phase, ensuring that choices reflected previously learned value representations and choice preferences rather than ongoing learning.

Critically, this design allowed us to disentangle distinct reward-based decision-making strategies, either based on global expected values or context-specific reinforcement learning history. Global expected value-based choices reflect outcome probabilities aggregated across the task, independent of local context. For example, a stimulus with a 25% chance of gain (LR) is objectively more valuable than one with a 25% chance of loss (LP), even though the former was the worse choice, while the latter was the best choice, within their original contexts. Preference for the LR stimulus in this case indicates reliance on global expected value. Context-specific reinforcement history, by contrast, reflects local learning within a fixed stimulus pair and decision context. Stimuli that were consistently the better option within their original pair (i.e., HR and LP) were associated with more frequent positive prediction errors during learning. For example, while LP has a negative absolute expected value, it was the relatively better option in its context (a 75% chance of avoiding loss). It was thus often associated with better-than-expected outcomes. Preference for LP over LR—despite its lower expected value—indicates sensitivity to recent, context-bound reinforcement signals.

By systematically varying pairings and contexts in the absence of feedback, the transfer phase provided a precise behavioral assay of each participant's decision-making strategy.

Instructions and Comprehension Check

Before beginning the main experiment, participants completed a guided instruction phase that familiarized them with the task structure and response procedures (see Supplementary Section Task Instructions). This phase included ten practice trials for each context—reward and punishment—using a separate set of stimuli and background images not used in the main task. These trials demonstrated the core principles of context-dependent learning and probabilistic feedback.

After the practice trials, participants were required to answer four comprehension questions correctly in order to proceed. These questions assessed their understanding of: (1) which keys to use for making responses, (2) how rewards and losses were accrued, (3) how to determine which knight in a pair was better, and (4) how their performance bonus would be calculated. Participants had up to three attempts to answer all questions correctly. If they failed to do so, they were excluded from the study and did not continue to the learning or transfer phases.

Task Versions

The data used in this study were collected using two task versions: version A (n = 84) and version B (n = 155). Thus far, we have described version B; however, there were slight differences between version A and version B. In version A, participants received dollar feedback (i.e., +$2, $0, -$2) instead of point feedback, and they selected stimuli using the left and right arrow keys rather than the F and J keys. Furthermore, in version A, participants conducted the experiment as described twice, with a break in between; however, data only from the first implementation were used in our analyses. We confirmed the task version did not affect our findings by repeating all main behavioral analyses with task version as a covariate (Supplementary Tables S3 and S11).

Statistical Analyses

Learning Phase

Performance during the learning phase was quantified using two measures: accuracy and reaction time (RT). Accuracy was defined as the proportion of trials in which the objectively better stimulus was selected, specifically, the high reward (HR) over the low reward (LR) stimulus in the reward context, and the low punishment (LP) over the high punishment (HP) stimulus in the punishment context. Reaction times were recorded as the latency between stimulus onset and response. As RTs exhibited a positive skew, a log transformation was applied prior to analyses.

To examine the temporal dynamics of learning, the 24 presentations of each stimulus pair were divided into four trial bins (hereafter named time): early (trials 1–6), mid-early (7–12), mid-late (13–18), and late (19–24). This binning enabled the time-resolved assessment of changes in accuracy and RT across the learning period.

Statistical analyses were conducted using generalized linear mixed-effects models (GLMMs). For both accuracy and RT, fixed effects included group (no pain, acute pain, chronic pain), context (reward, punishment), and time (trial bins), as well as all two- and three-way interactions. Participant identity was entered as a random effect, allowing the intercept to vary across participants. Accuracy was modeled using a binomial distribution as per equation (1), and RTs were modeled using a gamma distribution with a log link function as per equation (2):

$$\text{accuracy} \sim \text{group} \times \text{context} \times \text{time} + (1 \mid \text{participant}), \quad \text{family: binomial} \quad (1)$$

$$\text{reaction time} \sim \text{group} \times \text{context} \times \text{time} + (1 \mid \text{participant}), \quad \text{family: gamma} \quad (2)$$

Planned comparisons were conducted to test specific hypotheses regarding group differences. For the main effect of group, two independent-samples t-tests were performed to compare the chronic pain group against the no pain and acute pain groups. To assess the group-by-context interaction, accuracy and RTs were compared between

the chronic and no pain groups within each context separately, as well as for the difference between reward and punishment contexts. All additional comparisons were evaluated using Tukey's Honestly Significant Difference (HSD) post hoc tests.

Transfer Phase

Performance during the transfer phase was quantified using two measures: choice rate and RT. Choice rate was defined as the proportion of times a given stimulus was selected when paired with another. Trials in which both stimuli had identical values (e.g., two high reward stimuli) were excluded from analysis, as they provided no informative contrast. RTs were defined as the time, in milliseconds, taken to select one of the two stimuli from their presentation, and again log-transformed.

Choice rates and reaction times were analyzed using GLMMs. Fixed effects included group (no pain, acute pain, chronic pain), stimulus type (HR, LR, LP, HP, Novel [N]), and their interaction. Participant was again modeled as a random intercept. Choice rates were modeled using a Gaussian distribution as per equation (3), and reaction times were modeled using a gamma distribution with a log link function as per equation (4):

$$\text{choice rate} \sim \text{group} \times \text{stimulus type} + (1 \mid \text{participant}), \quad \text{family: Gaussian} \tag{3}$$

$$\text{reaction time} \sim \text{group} \times \text{stimulus type} + (1 \mid \text{participant}), \quad \text{family: gamma} \tag{4}$$

Planned comparisons were conducted to test specific hypotheses regarding group differences in valuation strategy. For both outcome variables, the no pain and acute pain groups were compared to the chronic pain group to detect overall group effects. To investigate the influence of valuation strategy, two stimulus comparisons were of particular interest.

First, the contrast between HR and LP stimuli (HR-LP) was used to assess sensitivity to objective value. Second, the contrast between LR and LP stimuli (LR-LP) allowed us to examine situations in which objective and relative values diverged. Preference for LR would suggest reliance on global value representations, while preference for LP would reflect sensitivity to reinforcement history. Equal preference for both would imply equal integration of both dimensions.

For the group-by-stimulus interaction, planned comparisons were conducted to examine group differences in these two contrasts (HR-LP and LR-LP). Specifically, the no pain, acute pain, and chronic pain groups were compared pairwise to assess whether valuation strategies varied as a function of pain state. These planned comparisons functioned on choice rates specific to these pairs of data—for example, the HR-LP choice rates would be the number of times that the HR stimulus type was chosen when paired with the LP stimulus type minus the number of times that the LP stimulus type was chosen when paired with the HR stimulus type (which would be the complement of the first). All remaining comparisons were evaluated using Tukey's Honestly Significant Difference (HSD) post hoc tests.

Significance

Statistical significance was determined using a threshold of $\alpha < .05$. For all GLMMs, assumptions of linearity, homoskedasticity, and normality of residuals were assessed. Linearity was satisfied in all cases. Although violations of homoskedasticity and normality were observed, these assumptions have been shown to have a limited impact on the validity of GLMMs [51]; therefore, no corrective measures were applied. For independent-samples t-tests, assumptions of normality and homogeneity of variance were evaluated. Where necessary, variables exhibiting non-normal distributions were transformed. When the assumption of equal variances was violated, Welch's t-tests were used in place of standard t-tests.

Computational Modeling

Computational modeling was used to quantify the cognitive mechanisms underlying group differences in decision-making. As described earlier, reward-based decision-making can be informed by two factors: an objective dimension based on absolute expected value, and a relative dimension shaped by context-specific reinforcement history. To dissociate the influence of these factors, we fit a set of reinforcement learning models that captured each dimension independently or in combination: a Q-Learning model to represent pure expected value learning, an Actor-Critic model to capture sensitivity to local reinforcement [17], a Relative model [25], a Hybrid model [23, 24], and an Advantage model that all integrate both factors in different ways.

Q-Learning Model

To capture decision-making based on absolute expected value, we implemented a Q-Learning model with counterfactual feedback, which updates action values for both the chosen and unchosen option on each trial [26]. In this model, Q-values represent the expected value of each action in a given state and are updated according to a delta rule. For the chosen action α_c and unchosen action α_u, the updates are defined as equations (5) and (6):

$$Q(s, a_c, t{+}1) = Q(s, a_c, t) + \alpha_c \delta_c \quad (5)$$
$$Q(s, a_u, t{+}1) = Q(s, a_u, t) + \alpha_u \delta_u \quad (6)$$

where α_c and α_u are the learning rates for the chosen and unchosen actions, respectively. Prediction errors, δ_c and δ_u, are computed as equations (7) and (8):

$$\delta_c = R_c - Q(s, a_c, t) \quad (7)$$
$$\delta_u = R_u - Q(s, a_u, t) \quad (8)$$

where R_c and R_u are the outcome feedback for the chosen and unchosen action, respectively. To standardize across gain and loss contexts, outcomes were rescaled to -1 (loss), 0 (neutral), and +1 (gain) prior to model fitting.

Action selection was modeled using a softmax choice rule, as per equation (9):

$$P(s, a_1, t) = \frac{e^{Q(s,a_1,t)/\tau}}{e^{Q(s,a_1,t)/\tau} + e^{Q(s,a_2,t)/\tau}} \quad (9)$$

where τ is the temperature parameter controlling choice stochasticity. The complete model, therefore, includes three free parameters: α_c, α_u, and τ.

Actor-Critic

To model decision-making driven by context-specific reinforcement history, we implemented an Actor-Critic model with counterfactual feedback. This architecture reflects the functional organization of basal ganglia circuitry, where action selection is influenced by bottom-up prediction errors relative to the current state value [27].

In this framework, the critic component estimates the value of each state, which is updated based on the outcome of the chosen action, as per equation (10):

$$V(s, t+1) = V(s, t) + \alpha_c \delta_c \quad (10)$$

where α_c is the critic learning rate, and δ_c is the prediction error of the chosen action. Prediction errors are computed for the chosen and unchosen actions, which are displayed throughout the learning phase, as equations (11) and (12):

$$\delta_c = R_c - V(s, t) \quad (11)$$

$$\delta_u = R_u - V(s, t) \quad (12)$$

where R_c and R_u are the outcome feedback for the chosen and unchosen actions, respectively. Each action within a state has a w-value, which is updated as per equations (13) and (14):

$$w(s, a_c, t+1) = w(s, a_c, t) + \alpha_{ac} \delta_c \quad (13)$$

$$w(s, a_u, t+1) = w(s, a_u, t) + \alpha_{au} \delta_u \quad (14)$$

where α_{ac} is the actor learning rate for the chosen action, and α_{au} is the actor learning rate for the unchosen action. To prevent unbounded growth of these values, we normalize them as per equation (15):

$$w(s, a_1, t+1) = \frac{w(s, a_1, t+1)}{|w(s, a_1, t+1)| + |w(s, a_2, t+1)|} \quad (15)$$

Action selection was modeled using a softmax choice rule as per equation (16):

$$P(s, a_1, t) = \frac{e^{w(s, a_1, t)/\tau}}{e^{w(s, a_1, t)/\tau} + e^{w(s, a_2, t)/\tau}} \quad (16)$$

where τ is the temperature parameter. This model results in a total of four parameters: α_c, α_{ac}, α_{au}, and τ.

Hybrid Model

The Hybrid model captures the combined influence of absolute expected value and relative reinforcement history on decision-making by integrating Q-Learning and Actor-Critic components [23, 24], by computing a weighted mixture of Q-value and Actor-Critic weights, indicated in equation (17):

$$H(s, a_i, t) = cQ(s, a_i, t) + (1 - c)w(s, a_i, t) \quad (17)$$

where $Q(s, a_i, t)$ reflects the absolute expected value of action i in state s, learned through the Q-Learning component, and $w(s, a_i, t)$ reflects the action preference learned via the Actor-Critic component. The mixing parameter c determines the relative contribution of each system: when $c = 1$, the model behaves as a pure Q-Learner; when $c = 0$, it reduces to an Actor-Critic model.

Actions are selected using a softmax decision rule based on H-values, as per equation (18):

$$P(s, a_1, t) = \frac{e^{H(s,a_1,t)/\tau}}{e^{H(s,a_1,t)/\tau} + e^{H(s,a_2,t)/\tau}} \quad (18)$$

The Hybrid model was initially introduced by Gold et al. [24] and later extended by Geana et al. [23] to include noise and decay parameters. In the present study, we employed a streamlined version of the original Hybrid model, modified to incorporate counterfactual learning rates for both the Q-Learning and Actor-Critic components, and excluding the valence bias parameter. Although we focus here on this specific variant, we systematically evaluated a set of ten Hybrid model variations, including both original formulations and modified versions with or without counterfactual learning, valence bias, and decay terms.

The Hybrid model used in the present study includes seven free parameters: two specific to the Q-Learning component, α_{qc} and α_{qu}, three specific to the Actor-Critic component, α_{wc}, α_{wac}, and α_{wau}, and two additional parameters governing decision dynamics—the softmax temperature τ and the mixture weight c, which determines the relative contribution of each component to the choice behavior.

Relative Model

The relative model is a mixture model that learns combined objective-relative values of actions [25]. It is a modified Q-Learning model with the addition of a learned state value, which serves as a reference point when computing prediction errors (see also [43, 52, 53]. As with the Q-Learning model, Q-values are updated as per equations (5) and (6), which include α_c as the learning rate for the chosen action and α_u as the learning rate for the unchosen action. However, unlike the Q-Learning model, the prediction errors are computed as per equations (19) and (20):

$$\delta_c = R_c - V(s, t) - Q(s, a_c, t) \quad (19)$$

$$\delta_u = R_u - V(s, t) - Q(s, a_u, t) \quad (20)$$

where R_c and R_u are the outcome feedback for the chosen and unchosen actions, respectively. $V(s,t)$ is the learned state value, which is updated as in equation (21):

$$V(s,t+1) = V(s,t) + \alpha_v \delta_v \quad (21)$$

where α_v is the state learning rate. δ_v is the state prediction error, computed as in equation (22):

$$\delta_v = R(s,t) - V(s,t) \quad (22)$$

where $R(s,t)$ is the average reward of each action computed as per equation (23):

$$R(s,t) = \frac{R_c + R_u}{2} \quad (23)$$

As with the Q-Learning model, action selection uses the softmax function depicted in equation (9), which includes τ as the temperature. This model results in a total of four parameters: α_v, α_c, α_u, and τ.

Advantage Model

Here, we introduce the Advantage model, which is a simplified version of the Relative model [25]. The Advantage model includes the averaged reward of the current state within the prediction error directly, rather than as a learned process. This may be reflected in findings suggesting that striatal dopamine fluctuations encode the integration of a reward prediction error with a counterfactual prediction error, reflecting how much better or worse the current outcome could have been [54]. The prediction errors in the Advantage model are calculated as in equations (24) and (25):

$$\delta_c = R_c - mR(s,t) - Q(s,a_c,t) \quad (24)$$

$$\delta_u = R_u - mR(s,t) - Q(s,a_u,t) \quad (25)$$

where m is a weighting factor determining the influence of reward on the prediction error, and $R(s,t)$ is the average of the rewards available for the chosen and unchosen option, as described in equation (23). Q-value updates remain the same as the Q-Learning and Relative models, as depicted in equations (5) and (6), and so does the action selection function, as per equation (9). This model results in a total of four parameters: m, α_c, α_u, and τ.

Novelty Parameter

During the transfer phase, participants encountered a novel stimulus that had not appeared during the learning phase and was therefore unassociated with any past reinforcement history. In principle, such a stimulus could be treated as neutral, with no expected value. However, behavioral data revealed systematic preferences or aversions toward the novel option, suggesting that participants assigned it a subjective value (see Figure 2). To account for this, all models included an additional free parameter,

η, which captures the estimated value of the novel stimulus during transfer-based decisions. This parameter allows the model to flexibly accommodate individual tendencies toward novelty seeking (exploration bias) or novelty avoidance.

Fitting Procedure

Models were fit to individual participant data by minimizing the negative log-likelihood of observed choices across both the learning and transfer phases. For each participant and model, fitting was repeated ten times with pseudo-randomized initial parameter values to avoid local minima. Initial values were drawn from normal distributions centered on prior estimates derived from the original publications of each model (see Supplementary Table S12), with standard deviations set to one-fifteenth of the parameter's bounded range. Values falling outside these bounds were clipped to the nearest allowable limit.

Model comparison was conducted using the average Bayesian Information Criterion (BIC) based on each participant n and model m, calculated as in equation (26):

$$BIC_m = \frac{1}{N} \sum_{n=1}^{N} (2\, NLL_{n,m} + t_{\log} k_m) \quad (26)$$

where $NLL_{n,m}$ is the negative log-likelihood for the corresponding participant n and model m, $t_{\log}$ is the log-transformed number of trials for that participant (which is equal across participants), k_m is the number of parameters for the corresponding model, and N is the total number of participants.

Statistical Analyses of Parameter Estimates

Statistical analyses were performed on the best-fitting model, as determined by BIC. For each model parameter, group differences were assessed using one-way ANOVAs with group (no pain, acute pain, chronic pain) as the between-subject factor. Assumptions of normality and homogeneity of variance were tested; when normality was violated, log-transformations were applied. Homogeneity of variance was generally upheld.

Planned comparisons focused on differences between the no pain group and both the acute and chronic pain groups. For each comparison, independent-samples t-tests were conducted. Welch's t-tests were used when the assumption of equal variances was violated. Tukey HSD post hoc tests were applied to evaluate all remaining group differences.

We also conducted a series of correlational analyses comparing the different fitted parameters with pain metrics. First, for each of the acute and chronic pain groups, we conducted correlational analyses between the fitted parameters of the winning model (Advantage model: factual learning rate, counterfactual learning rate, temperature, weighting factor, novel value) and the pain metrics (intensity, unpleasantness, interference, composite) using Pearson r correlations. These analyses were not conducted for the no pain group because their range of pain scores was too narrow. Second, we conducted ordinal correlation analyses between the winning model parameter fits and the task duration responses using Spearman Rho correlations. Pain duration responses

were coded ordinally as follows: I am not in pain = 0, < 2 weeks = 1, 2–4 weeks = 2, 1–3 months = 3, 3–6 months = 4, 6–12 months = 5, 1–5 years = 6, > 5 years = 7, and > 10 years = 8.

Data and Code Availability

Data and analysis code for analyses are available within the GitHub repository, avoid learning analysis, while processed data and modeling code for computational modeling are available within the GitHub repository, avoid learning rl models.

Acknowledgements

This work was supported by the Brainstorm Program at the Robert J. & Nancy D. Carney Institute for Brain Science (FHP) and the COBRE Center for Nervous System Function (NIGMS 5P20GM103645).

References

[1] Cohen, S.P., Vase, L., Hooten, W.M.: Chronic pain: an update on burden, best practices, and new advances. The Lancet **397**(10289), 2082–2097 (2021)

[2] Nahin, R.L., Feinberg, T., Kapos, F.P., Terman, G.W.: Estimated rates of incident and persistent chronic pain among us adults, 2019-2020. JAMA Network Open **6**(5), 2313563–2313563 (2023)

[3] Yong, R.J., Mullins, P.M., Bhattacharyya, N.: Prevalence of chronic pain among adults in the united states. Pain **163**(2), 328–332 (2022)

[4] Lethem, J., Slade, P.D., Troup, J.D.G., Bentley, G.: Outline of a fear-avoidance model of exaggerated pain perception—i. Behaviour Research and Therapy **21**(4), 401–408 (1983)

[5] Vlaeyen, J.W., Crombez, G., Linton, S.J.: The fear-avoidance model of pain. Pain **157**(8), 1588–1589 (2016)

[6] Crombez, G., Eccleston, C., Van Damme, S., Vlaeyen, J.W., Karoly, P.: Fear-avoidance model of chronic pain: the next generation. The Clinical Journal of Pain **28**(6), 475–483 (2012)

[7] Zale, E.L., Maisto, S.A., Ditre, J.W.: Interrelations between pain and alcohol: An integrative review. Clinical Psychology Review **37**, 57–71 (2015) https://doi.org/10.1016/j.cpr.2015.02.004

[8] Baliki, M.N., Petre, B., Torbey, S., Herrmann, K.M., Huang, L., Schnitzer, T.J., Apkarian, A.V.: Corticostriatal functional connectivity predicts transition to chronic back pain. Nature Neuroscience **15**(8), 1117–1119 (2012)

[9] Woo, C.W., Roy, M., Buhle, J.T., Wager, T.D.: Distinct brain systems mediate the effects of nociceptive input and self-regulation on pain. PLoS Biology **13**(1), 1002036 (2015)

[10] Löffler, M., Levine, S.M., Usai, K., Desch, S., Kandić, M., Nees, F., Flor, H.: Corticostriatal circuits in the transition to chronic back pain: the predictive role of reward learning. Cell Reports Medicine **3**(7) (2022)

[11] Averbeck, B., O'Doherty, J.P.: Reinforcement-learning in fronto-striatal circuits. Neuropsychopharmacology **47**(1), 147–162 (2022) https://doi.org/10.1038/s41386-021-01108-9

[12] O'Doherty, J.P., Cockburn, J., Pauli, W.M.: Learning, reward, and decision making. Annual Review of Psychology **68**(1), 73–100 (2017) https://doi.org/10.1146/annurev-psych-010416-044216

[13] Montague, P.R., Dayan, P., Sejnowski, T.J.: A framework for mesencephalic

dopamine systems based on predictive hebbian learning. The Journal of Neuroscience **16**(5), 1936–1947 (1996) https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996

[14] Schultz, W., Dayan, P., Montague, P.R.: A neural substrate of prediction and reward. Science **275**(5306), 1593–1599 (1997) https://doi.org/10.1126/science.275.5306.1593

[15] Masset, P., Gershman, S.J.: Reinforcement learning with dopamine: A convergence of natural and artificial intelligence. In: The Handbook of Dopamine. Elsevier, ??? (2024)

[16] Niv, Y.: Reinforcement learning in the brain. Journal of Mathematical Psychology **53**(3), 139–154 (2009) https://doi.org/10.1016/j.jmp.2008.12.005

[17] Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction vol. 1, pp. 9–11. MIT Press, Cambridge, MA (1998)

[18] Hazy, T.E., Frank, M.J., O'Reilly, R.C.: Neural mechanisms of acquired phasic dopamine responses in learning. Neuroscience and Biobehavioral Reviews **34**, 701–720 (2010) https://doi.org/10.1016/j.neubiorev.2009.11.019

[19] O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., Dolan, R.J.: Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science **304**, 452–454 (2004)

[20] Roesch, M.R., Calu, D.J., Schoenbaum, G.: Dopamine neurons encode the better option in rats deciding between differently delayed or sized rewards. Nature Neuroscience **10**, 1615–1624 (2007) https://doi.org/10.1038/nn2013

[21] Bartra, O., McGuire, J.T., Kable, J.W.: The valuation system: a coordinate-based meta-analysis of bold fmri experiments examining neural correlates of subjective value. Neuroimage **76**, 412–427 (2013)

[22] Wiehler, A., Petzschner, F.H., Stephan, K.E., Peters, J.: Episodic tags enhance striatal valuation signals during temporal discounting in pathological gamblers. ENeuro **4**(3), 0159–172017 (2017)

[23] Geana, A., Barch, D.M., Gold, J.M., Carter, C.S., MacDonald, A.W.I., Ragland, J.D., Frank, M.J.: Using computational modeling to capture schizophrenia-specific reinforcement learning differences and their implications on patient classification. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **7**(10), 1035–1046 (2022) https://doi.org/10.1016/j.bpsc.2022.01.008

[24] Gold, J.M., Waltz, J.A., Matveeva, T.M., Kasanova, Z., Strauss, G.P., Herbener, E.S., Frank, M.J.: Negative symptoms and the failure to represent the

expected reward value of actions: Behavioral and computational modeling evidence. Archives of General Psychiatry **69**(2), 129–138 (2012) https://doi.org/10.1001/archgenpsychiatry.2011.1269

[25] Palminteri, S., Khamassi, M., Joffily, M., Coricelli, G.: Contextual modulation of value signals in reward and punishment learning. Nature Communications **6**(1), 8096 (2015) https://doi.org/10.1038/ncommS8096

[26] Watkins, C.J.C.H., Dayan, P.: Q-learning. Machine Learning **8**, 279–292 (1992) https://doi.org/10.1007/BF00992698

[27] Konda, V., Tsitsiklis, J.: Actor-critic algorithms. In: Advances in Neural Information Processing Systems, vol. 12 (1999)

[28] Apkarian, A.V., Baliki, M.N., Farmer, M.A.: Predicting transition to chronic pain. Current Opinion in Neurology **26**, 360–367 (2013) https://doi.org/10.1097/wco.0b013e32836336ad

[29] Kato, T., Ide, S., Minami, M.: Pain relief induces dopamine release in the rat nucleus accumbens during the early but not late phase of neuropathic pain. Neuroscience Letters **629**, 73–78 (2016)

[30] Huys, Q.J.M., Maia, T.V., Frank, M.J.: Computational psychiatry as a bridge between neuroscience and clinical applications. Nature Neuroscience **19**, 404–413 (2016)

[31] Maia, T.V., Frank, M.J.: From reinforcement learning models to psychiatric and neurological disorders. Nature Neuroscience **14**, 154–162 (2011)

[32] Montague, R.P., Dolan, R.J., Friston, K.J., Dayan, P.: Computational psychiatry. Trends in Cognitive Sciences **16**(1), 72–80 (2012)

[33] Petzschner, F.H., Weber, L.A.E., Gard, T., Stephan, K.E.: Computational psychosomatics and computational psychiatry: Toward a joint framework for differential diagnosis. Biological Psychiatry **82**, 421–430 (2017)

[34] Redish, A.D., Gordon, J.A.: Computational Psychiatry: New Perspectives on Mental Illness. MIT Press, ??? (2016)

[35] Wiecki, T.V., Poland, J., Frank, M.J.: Model-based cognitive neuroscience approaches to computational psychiatry: Clustering and classification. Clinical Psychological Science (2015) https://doi.org/10.1177/2167702614565359

[36] Wiecki, T.V., al.: A computational cognitive biomarker for early-stage huntington's disease. PLoS ONE **11**, 0148409 (2016)

[37] Conceição, V.A., Petzschner, F.H., Cole, D.M., Wellstein, K.V., Müller, D., Raman, S., Maia, T.V.: Serotonin reduces belief stickiness. bioRxiv (2023) https:

//doi.org/10.1101/2023.12.08.570769

[38] Schneebeli, M., Haker, H., Rüesch, A., Zahnd, N., Marino, S., Paolini, G., Iglesias, S., Petzschner, F.H., Stephan, K.E.: Disentangling "Bayesian brain" theories of autism spectrum disorder. medRxiv (2022). https://doi.org/10.1101/2022.02.07.22270242

[39] Paliwal, S., Petzschner, F.H., Schmitz, A.K., Tittgemeyer, M., Stephan, K.E.: A model-based analysis of impulsivity using a slot-machine gambling paradigm. Frontiers in Human Neuroscience **8**, 428 (2014)

[40] Vandendriessche, H., al.: Contextual influence on reinforcement learning performance in depression: evidence for a negativity bias? Psychological Medicine **53**, 4696–4706 (2023)

[41] Huys, Q.J.M., Maia, T.V., Paulus, M.P.: Computational psychiatry: From mechanistic insights to the development of new treatments. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **1**, 382–385 (2016)

[42] Paulus, M.P., Huys, Q.J.M., Maia, T.V.: A roadmap for the development of applied computational psychiatry. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging **1**, 386–392 (2016)

[43] Juechems, K., Altun, T., Hira, R., Jarvstad, A.: Human value learning and representation reflect rational adaptation to task demands. Nature Human Behaviour **6**(9), 1268–1279 (2022)

[44] Perrot, S., Cohen, M., Barke, A., Korwisi, B., Rief, W., Treede, R.D.: The iasp classification of chronic pain for icd-11: chronic secondary musculoskeletal pain. Pain **160**(1), 77–82 (2019)

[45] Nicholas, M., Vlaeyen, J.W., Rief, W., Barke, A., Aziz, Q., Benoliel, R., Treede, R.D.: The iasp classification of chronic pain for icd-11: chronic primary pain. Pain **160**(1), 28–37 (2019)

[46] Vardeh, D., Mannion, R.J., Woolf, C.J.: Toward a mechanism-based approach to pain diagnosis. The Journal of Pain **17**(9), 50–69 (2016)

[47] Douglas, B.D., Ewell, P.J., Brauer, M.: Data quality in online human-subjects research: Comparisons between mturk, prolific, cloudresearch, qualtrics, and sona. PLOS ONE **18**(3), 0279720 (2023) https://doi.org/10.1371/journal.pone.0279720

[48] Gunsilius Zimmerman, C., Heffner, J., Bruinsma, S., Corinha, M., Cortinez, M., Dalton, H., Duong, E., Lu, J., Omar, A., Owen, L., Roarr, B.N., Tang, K., Petzschner, F.H.: Somascience: A novel platform for multidimensional, longitudinal pain assessment. JMIR mHealth and uHealth **12**(52), 47177 (2024)

[49] Treede, R.D., Rief, W., Barke, A., Aziz, Q., Bennett, M.I., Benoliel, R., Wang, S.J.: Chronic pain as a symptom or a disease: the iasp classification of chronic pain for the international classification of diseases (icd-11). Pain **160**(1), 19–27 (2019)

[50] Merskey, H., Bogduk, N.: Classification of Chronic Pain vol. 2. IASP Press, Seattle (1994)

[51] Schielzeth, H., Dingemanse, N.J., Nakagawa, S., Westneat, D.F., Allegue, H., Teplitsky, C., Araya-Ajoy, Y.G.: Robustness of linear mixed-effects models to violations of distributional assumptions. Methods in Ecology and Evolution **11**(9), 1141–1152 (2020) https://doi.org/10.1111/2041-210X.13434

[52] Bavard, S., Palminteri, S.: The functional form of value normalization in human reinforcement learning. Elife **12**, 83891 (2023)

[53] Louie, K.: Asymmetric and adaptive reward coding via normalized reinforcement learning. PLoS Computational Biology **18**(7), 1010350 (2022)

[54] Kishida, K.T., Saez, I., Lohrenz, T., Witcher, M.R., Laxton, A.W., Tatter, S.B., Montague, P.R.: Subsecond dopamine fluctuations in human striatum encode superposed error signals about actual and counterfactual reward. Proceedings of the National Academy of Sciences **113**(1), 200–205 (2016)

[55] Wilson, R.C., Collins, A.G.: Ten simple rules for the computational modeling of behavioral data. Elife **8**, 49547 (2019)

Supplementary Information

Supplementary Methods

Model Validation

To validate the robustness of our computational models and fitting procedures, we conducted a series of standard model checks, following best practices in reinforcement learning research [55]. In addition to the posterior predictive checks, we performed both parameter recovery and model recovery analyses.

For both validations, we generated simulated datasets for the given task structure for each of our 239 participants. Specifically, for each model, we sampled parameter values independently from uniform distributions bounded by each parameter's plausible range (Supplementary Table S12). Note that high values of temperature would make the models act randomly, making all other parameters obsolete and thus unrecoverable. Due to this, we constrained our model recovery to include temperature bounds of 0.01 to 0.20. These sampled parameters were then used to simulate participant behavior, matching the sequence of contexts, stimulus pairs, and outcomes encountered by each real participant. This procedure resulted in 239 synthetic datasets per model.

These simulated datasets were then submitted to the same model fitting and comparison procedures used in the main analyses, allowing us to assess whether model parameters and model identities could be accurately recovered (Supplementary Figures S7 and S8).

Parameter Recovery. To assess whether the models could reliably recover their underlying parameters, we performed parameter recovery analyses on the simulated datasets. For each model, parameters were re-estimated from its corresponding simulated data using the same fitting procedures as in the main manuscript, with two modifications. First, initial parameter values were drawn from uniform distributions constrained to the full range of each parameter, rather than from priors. Second, to improve identifiability, the upper bound for the softmax temperature parameter was reduced from 1.00 to 0.20 during recovery.

Recovered parameters were then compared to the ground-truth values used to generate the simulations, and Pearson correlations were computed as a measure of recovery fidelity. All models demonstrated successful recovery, though to varying degrees (Supplementary Figure S7, Supplementary Table S13). Recovery was highest for the best-fitting model, Advantage, with an average correlation of $r = .93$ across parameters. The Hybrid model showed the weakest recovery overall ($r = .57$), likely due to its greater parameter complexity and non-linear structure. However, even in the Hybrid model, the key parameter indexing the weighting between value systems—the mixing factor—was well recovered ($r = .89$). This suggests that while caution may be warranted when interpreting individual parameters in more complex models, the core mechanisms of interest were robustly estimated.

Model Recovery. To verify that each model was selectively identifiable, we conducted model recovery analyses. Specifically, we fit all models to every simulated dataset, using the same fitting procedure as in the parameter recovery analysis. For

each simulated dataset, model recovery was assessed using the Bayesian Information Criterion (BIC), with successful recovery defined as the generating model having the lowest BIC among all candidate models. All models showed successful recovery (Supplementary Figure S8).

Supplemental Results

Relative and Hybrid Model Results

As described in the main manuscript, the Advantage model provided the best fit to the empirical data based on model comparison metrics. Accordingly, we focused our main analyses on this model and identified the weighting factor as the key parameter capturing group-level differences in valuation strategy. However, two other mixture models—the Relative and Hybrid models—also showed good overall performance and are widely used in the literature [23–25]. For completeness and comparison, we report the results of these models here in the Supplementary Information. Specifically, we examine model fits, posterior simulations, and group differences in key parameters, and compare these outcomes to those observed in the Advantage model.

Relative Model Results. Posterior predictive checks for the Relative model closely replicated the empirical patterns observed in both the learning and transfer phases (Supplementary Figure S9a,b; Supplementary Tables S14, S15). Consistent with the Advantage model, only one parameter—the contextual learning rate—significantly differed across groups (p = .0216) (Supplementary Figure S9c). Planned comparisons revealed that this effect was driven by higher contextual learning rates in the chronic pain group compared to the no pain group (p = .0062, d = 0.41). No significant difference was found between the chronic and acute pain groups (p = .1053, d = 0.27), and post hoc comparisons indicated no reliable difference between the no pain and acute pain groups (p = .2609, d = –0.03). These results reinforce the interpretation that chronic pain is associated with a greater reliance on context-dependent relative values.

Hybrid Model Results. Posterior predictive checks for the Hybrid model closely reproduced the empirical choice behavior observed across pain groups (Supplementary Figure S10a,b; Supplementary Tables S16, S17). However, unlike the Advantage and Relative models, none of the Hybrid model parameters significantly distinguished between groups (Supplementary Figure S10c). Of particular interest was the mixing parameter, which plays a role analogous to the weighting factor in the Advantage model and the contextual learning rate in the Relative model. This parameter did not differ significantly across groups (p = .1766). Planned comparisons revealed no difference between the chronic pain and no pain groups (p = .0729, d = 0.27), nor between the acute and no pain groups (p = .3419, d = 0.16).

Although these effects did not reach significance, the downward trend in the mixing parameter from no pain to chronic pain mirrors the group differences captured by the other two models. Given the Hybrid model's greater complexity and reduced parameter recoverability, these null effects may reflect reduced sensitivity rather than a genuine absence of group differences.

Mixture Parameters Capture a Common Latent Mechanism

All three mixture models—the Advantage, Relative, and Hybrid models—include a parameter that modulates the influence of contextual information on value-based learning. In the Advantage model, this is the weighting factor; in the Relative model, it is the contextual learning rate; and in the Hybrid model, it is the mixing parameter. Although these parameters differ in their specific computational implementation, they share a common cognitive function: determining the balance between global expected value and context-specific reinforcement.

The interpretation of higher values differs across models. In the Advantage and Relative models, larger values indicate greater reliance on contextual or locally normalized value representations. In contrast, in the Hybrid model, higher mixing values correspond to increased reliance on global expected value, with reduced contextual influence. To assess whether these parameters capture a shared latent process, we examined their correlations across participants (Supplementary Figure S11). Consistent with their theoretical overlap, we observed strong correlations between all three: Advantage and Relative ($r = .85$, $p < .0001$), Advantage and Hybrid ($r = -0.59$, $p < .0001$), and Relative and Hybrid ($r = -0.56$, $p < .0001$). These findings support the interpretation that all three parameters capture a common valuation mechanism, albeit expressed differently across model architectures.

Task Instructions

Preamble

In this task, you are picking a team of knights. The knights will look like the ones below. Each knight will have a unique symbol on its chestplate. This symbol will help you identify each knight. You'll also pick your team of knights from different places, either the desert or the forest. On every turn, you will choose a knight for your team. When you select a knight, it may give you: +$2, +0 dollars, or –$2. Once you've selected your knight, their platform and visor will light up to indicate your choice. To help you learn, we will also show you the dollars you could have earned if you had chosen the other knight. Note, you will earn dollars only for the knight you chose. Some knights are better than others. Some will gain dollars while others will avoid losing dollars. Try to earn as many dollars as you can.

Learning Phase

During the task, there will be many different knights to choose from. Remember to pay close attention to their symbols on their chestplates. Your job is to try to select the best knight in each pair. Even though you will learn the outcomes for both knights, you will only earn dollars for the knight you choose. Hint: The knights may not always give you dollars, but some knights are better at gaining dollars while other knights are better at avoiding losing dollars. You should try to earn as many dollars as you can, even if it's not possible to win dollars or avoid losing dollars on every round. At the end of the task, the total number of dollars you've earned will be converted

into a performance percentage. You can earn up to an additional $4 based on your performance.

Transfer Phase

In this next part, you will see the same knights as before, but they will be shown in new pair combinations. Again, your job will be to select the knight you would like to join your team. As you make your choices, you will not receive any feedback after your choice. You should still choose the knight you think is better on each trial. Your choices will still contribute to your performance bonus.

Supplementary Figures

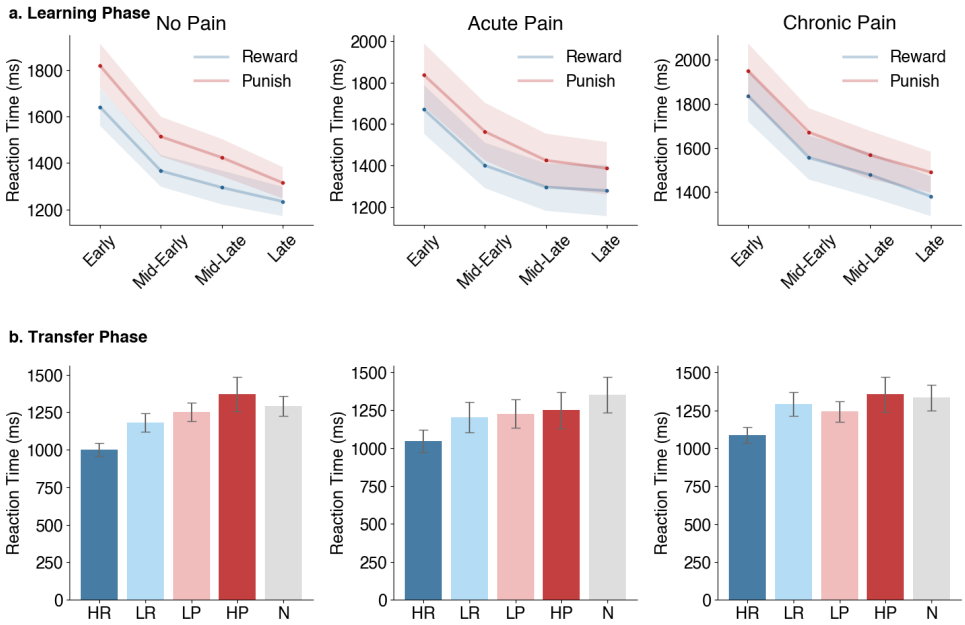

Fig. S1 Empirical findings of learning and transfer reaction times. a. Learning Phase: Behavioral reaction times across binned learning trials for the reward and punishment contexts for each group. Shaded regions represent 95% confidence intervals. b. Transfer Phase: Reaction times for each stimulus type during transfer trials for each group. Bar plots show the mean and 95% confidence intervals of the choice rate for each stimulus type across participants within each group. Abbreviations: HR – high reward rate (75% reward), LR – low reward rate (25% reward), LP – low punishment rate (25% loss), HP – high punishment rate (75% loss), N – novel stimulus.

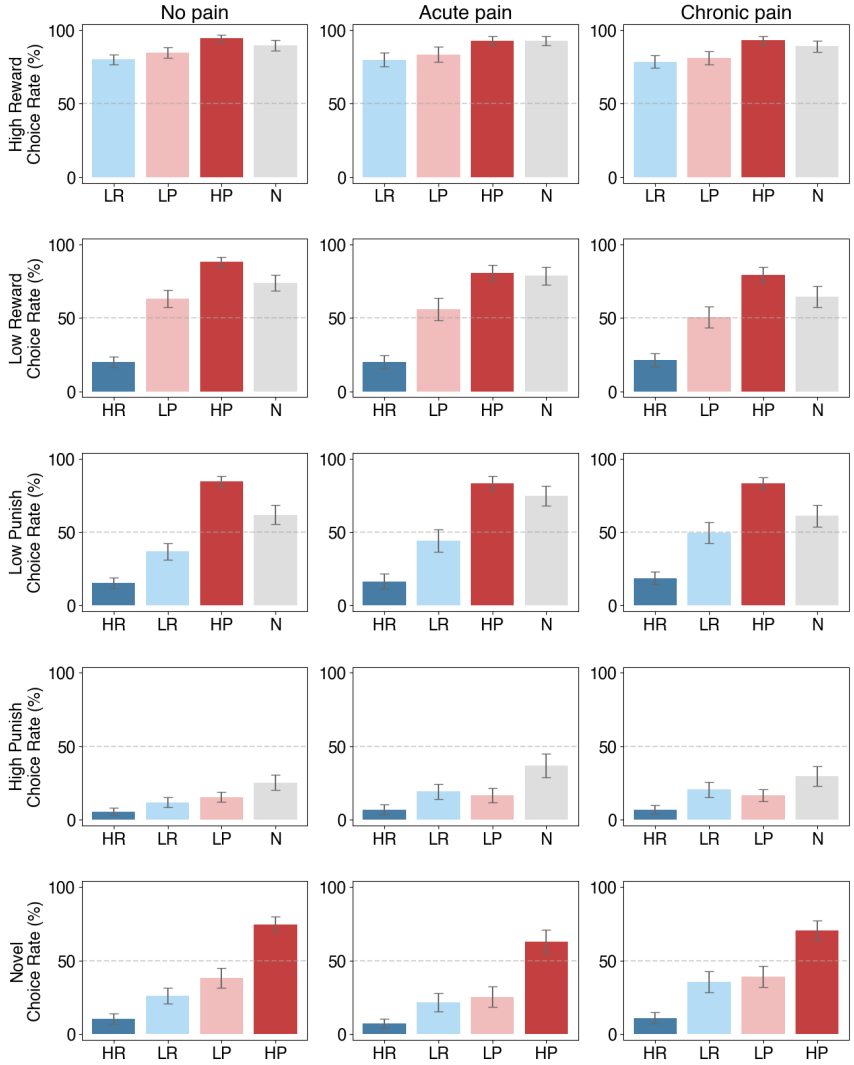

Fig. S2 Empirical findings of choice rates. Choice rates for each stimulus type during transfer trials for each group. Data are split where each row represents choice rates anchored to that stimulus type. For example, the first row indicates choice rates of the high reward stimulus type relative to all other stimulus types. This is simply an alternative viewing of the data presented in Figure 2b. Bar plots show the mean and 95% confidence intervals of the choice rate for each stimulus type across participants within each group. Abbreviations: HR – high reward rate (75% reward), LR – low reward rate (25% reward), LP – low punishment rate (25% loss), HP – high punishment rate (75% loss), N – novel stimulus.

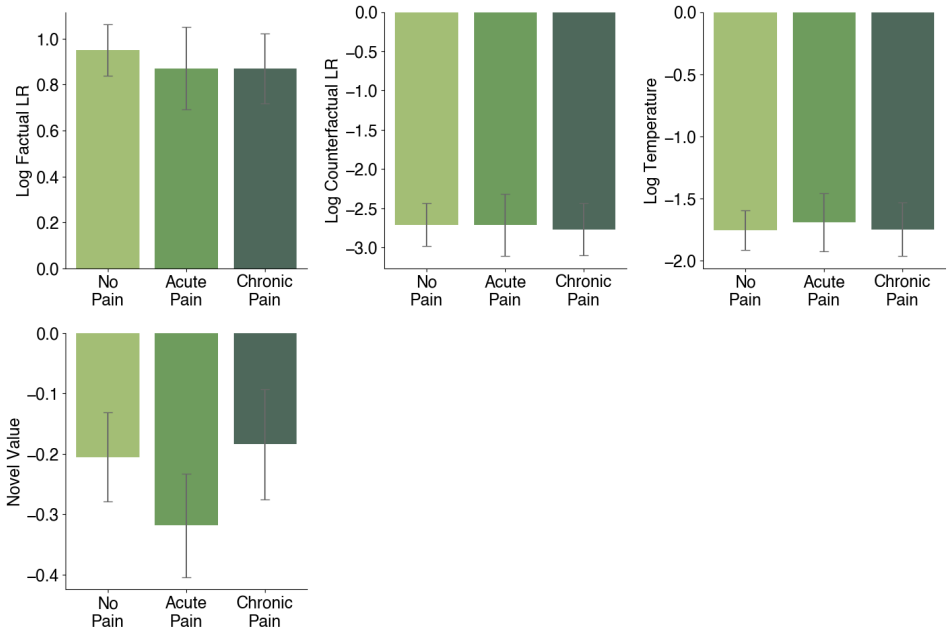

Fig. S3 Group effect in the parameters of the Advantage model. Fitted values of the non-significant parameters from the Advantage model. The parameter is displayed as a bar plot, showing the mean and 95% confidence intervals across participants for each group.

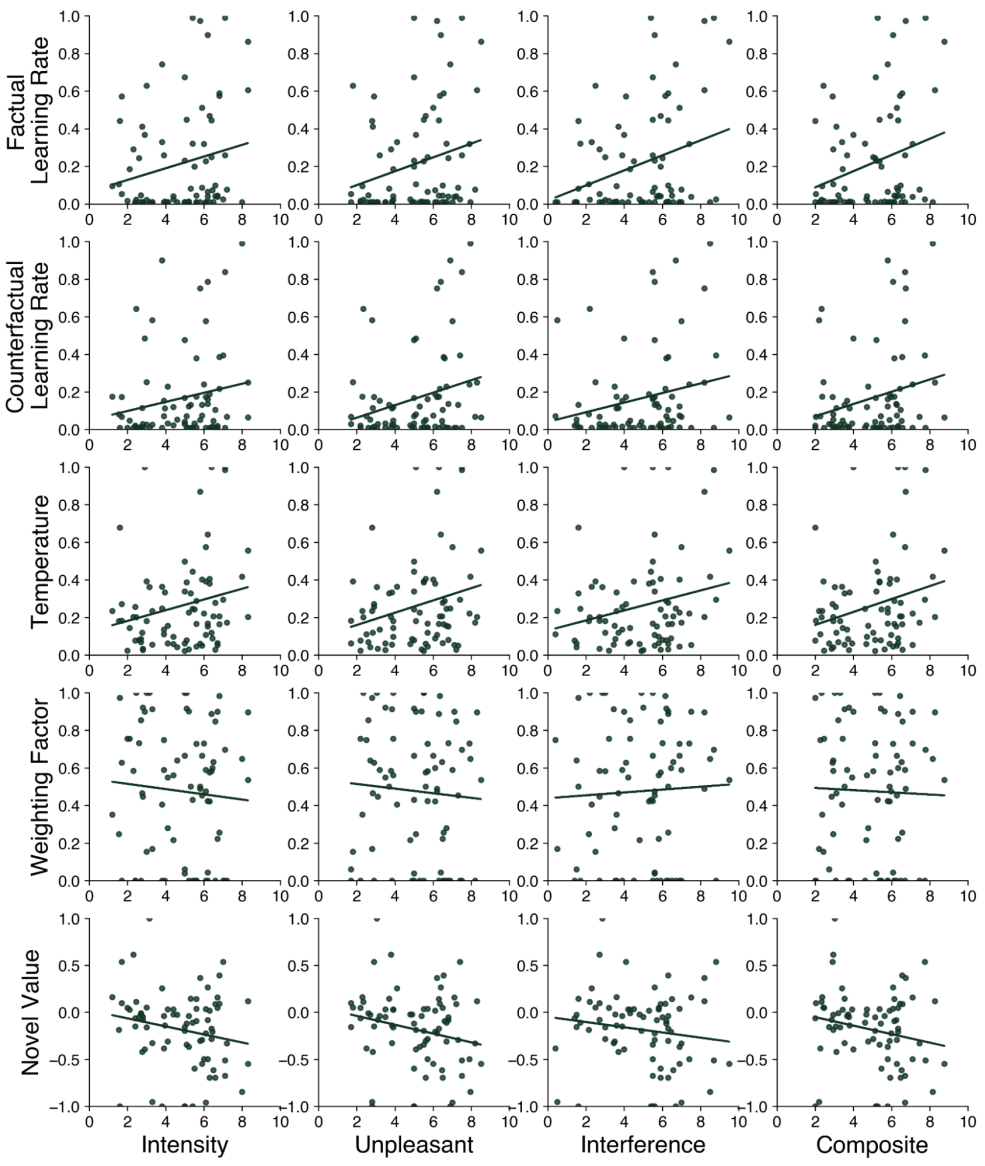

Fig. S4 Associations between model parameters and pain severity metrics for chronic pain. Scatterplots with regression lines linking the winning model's parameter fits with pain severity metrics for the chronic pain group. Associated Pearson r correlations can be found in Supplementary Table S8.

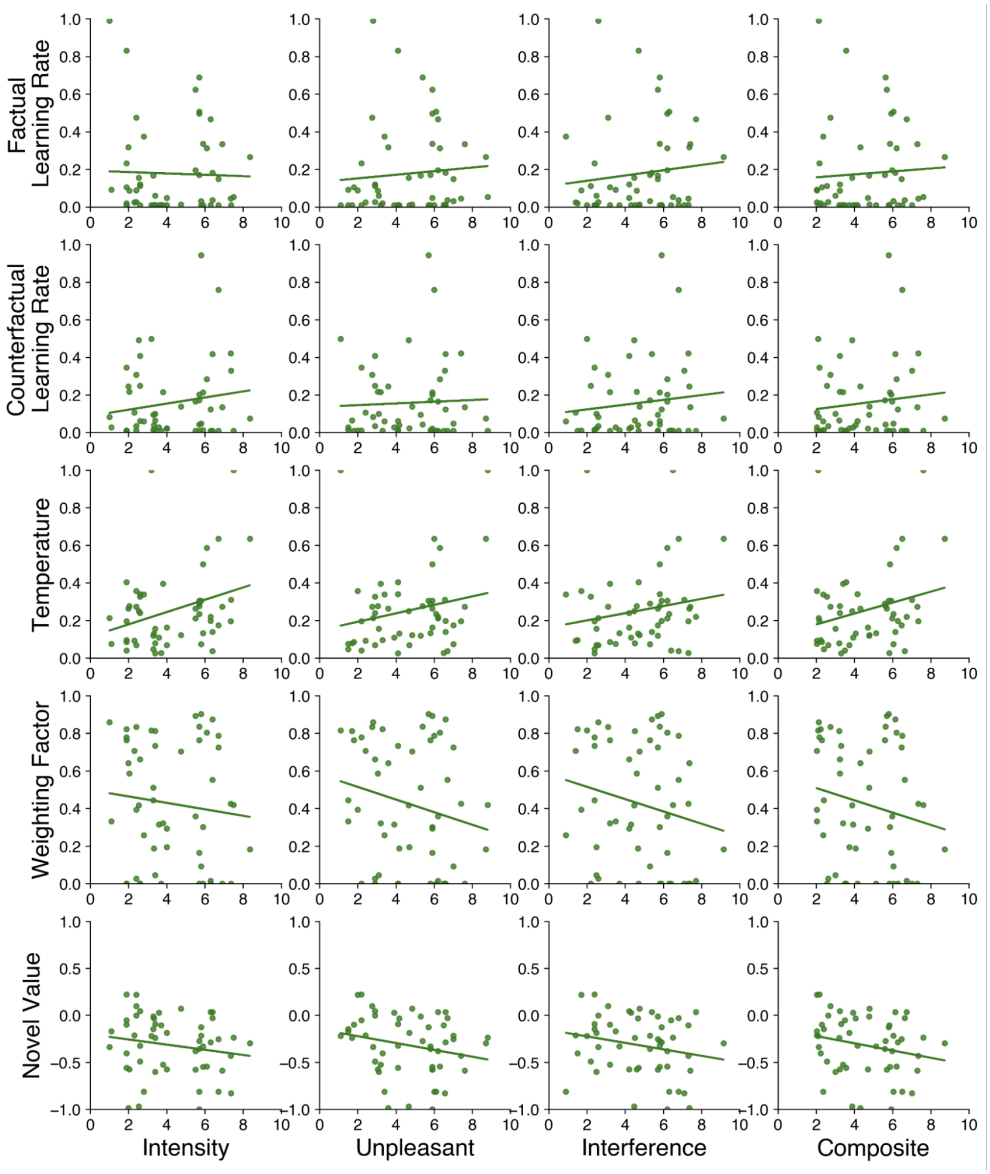

Fig. S5 Associations between model parameters and pain severity metrics for acute pain. Scatterplots with regression lines linking the winning model's parameter fits with pain severity metrics for the acute pain group. Associated Pearson r correlations can be found in Supplementary Table S8.

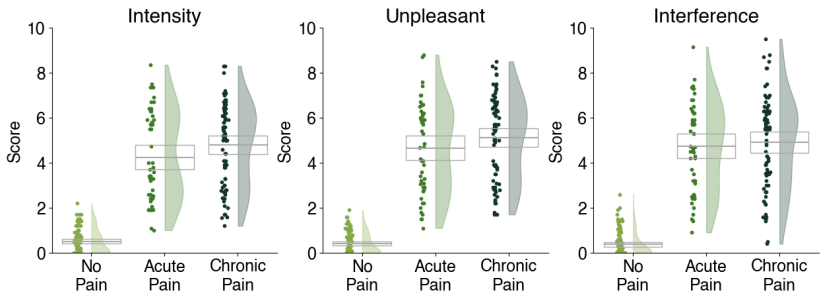

Fig. S6 Pain metrics for each group. Boxplots show the mean and 95% confidence intervals of the corresponding metric for each group. Half-violin plots show the distribution of the scores of the corresponding metric for each group. Scatter points show the scores of the corresponding metric for each participant within each group.

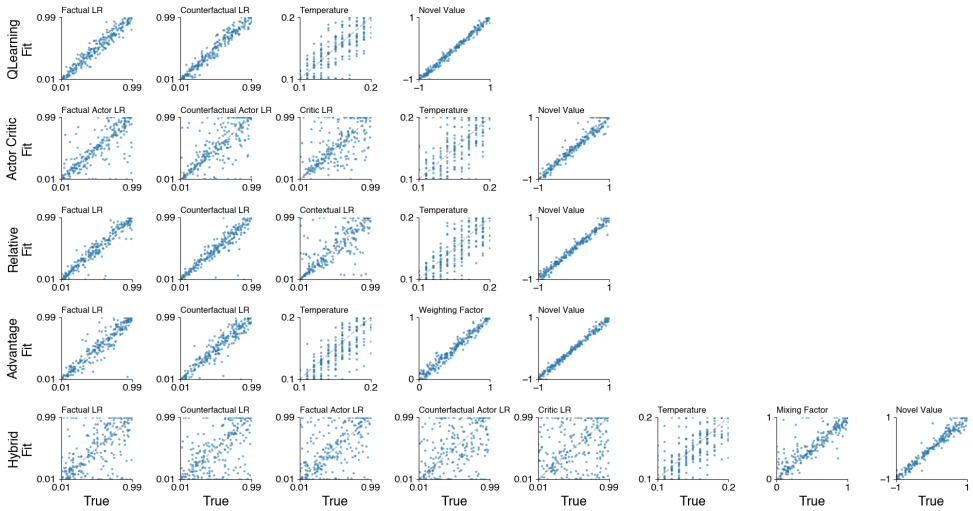

Fig. S7 Parameter recovery results. Parameter recovery for each parameter within each model. Data were generated for each model using randomly determined parameters (true values) and then fitted by that model (fit values) to assess the model's ability to recover parameters. Pearson's r correlations for each parameter determine the degree to which the parameter was recoverable. These values are presented in Supplementary Table S13. Grey dashed lines indicate a perfect recovery.

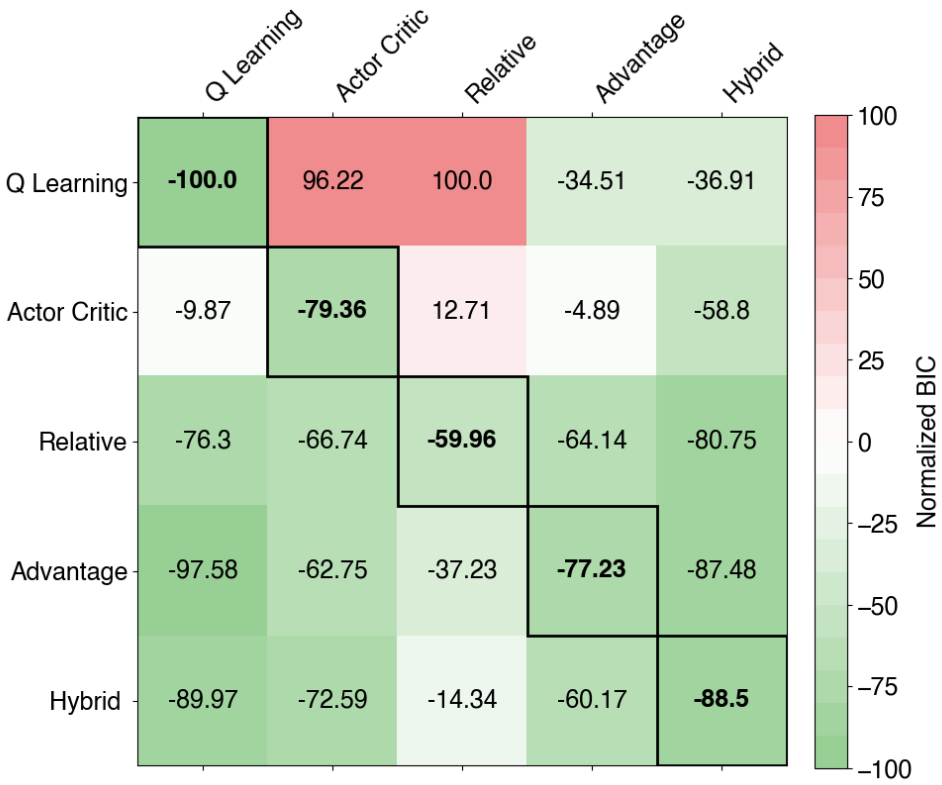

Fig. S8 Model recovery results. Model recovery for each model using normalized BIC metrics. Data were generated for each model using randomly determined parameters and then fitted by all models to assess each model's ability to fit the corresponding model's data. BIC values are normalized to the range of [0,1], scaled by 200, and then shifted by 100 to center around 0. A BIC of -100 indicates the best fit overall, and a value of +100 indicates the worst fit overall. Within each column, one model fit is highlighted (bolded and bordered), indicating the best-fitted model to explain the generated data. Successful model recovery is when the model best fits itself (as we see here). The values printed are the normalized BIC values.

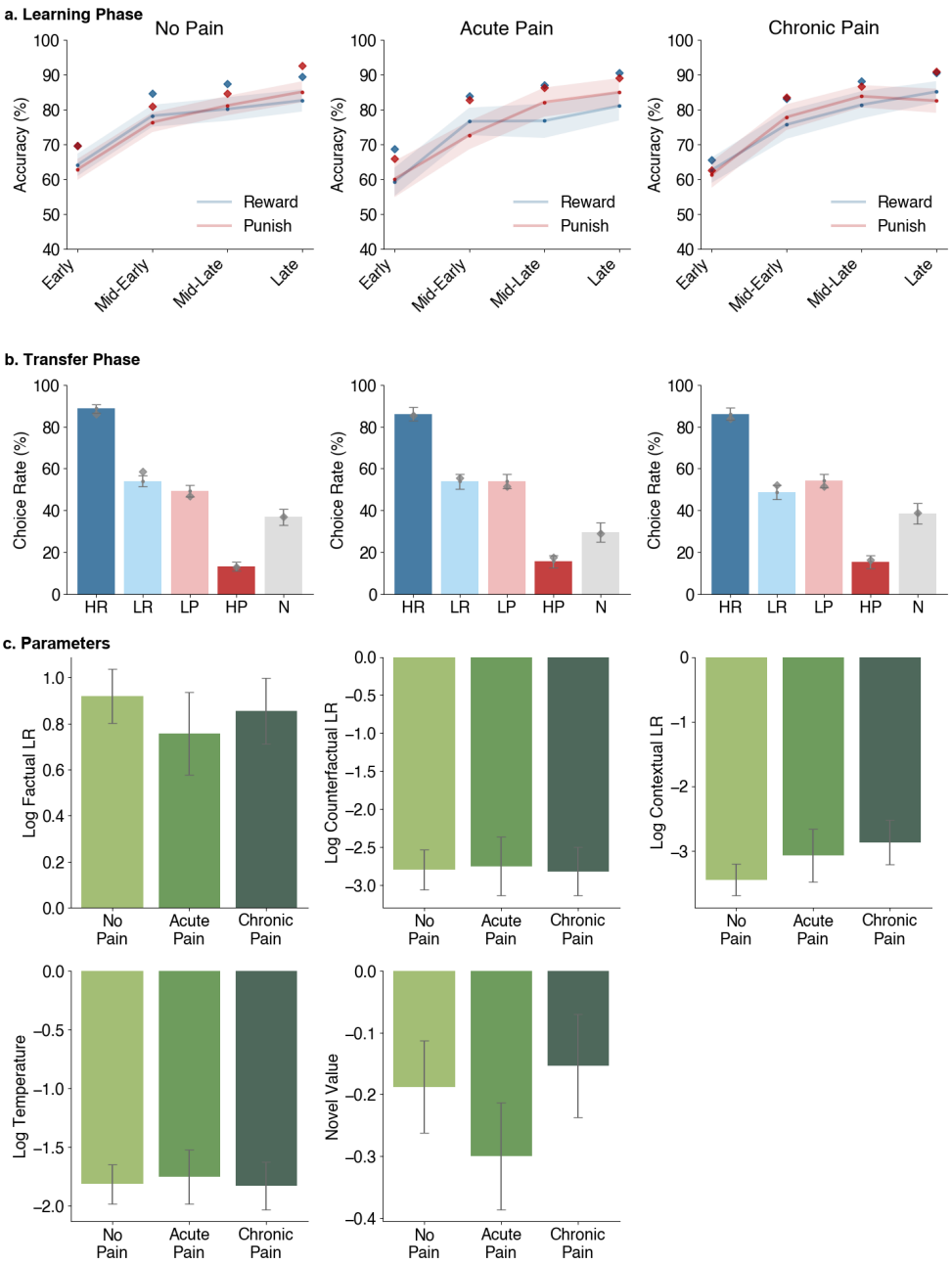

Fig. S9 Posterior Predictive checks for the runner-up computational model and fitted parameters. Model simulations of the runner-up model, the Relative model. During the fitting procedure, each participant was fitted ten times, and the best of these runs was selected to determine the participant's fitted parameters. We then used each model to simulate a single dataset per participant using these fitted parameters. Thus, these data represent the same sample size as our empirical data. a. Learning Phase: Model performance across binned learning trials for the reward and punishment contexts for each group. Shaded regions represent 95% confidence intervals. Blue and red diamonds indicate empirical means of participant accuracy. b. Transfer Phase: Choice rates for each stimulus type during transfer trials for each group. Choice rate is computed as the percentage of times a stimulus type was chosen, given the number of times it was presented. Bar plots show the mean and 95% confidence intervals of the choice rate for each stimulus type across participants within each group. Grey diamonds indicate empirical means of participant choice rates. Abbreviations: HR – high reward rate (75% reward), LR – low reward rate (25% reward), LP – low punishment rate (25% loss), HP – high punishment rate (75% loss), N – novel stimulus. c. Parameters: Fitted parameter values are displayed as a bar plot, showing the mean and 95% confidence intervals across participants for each group.

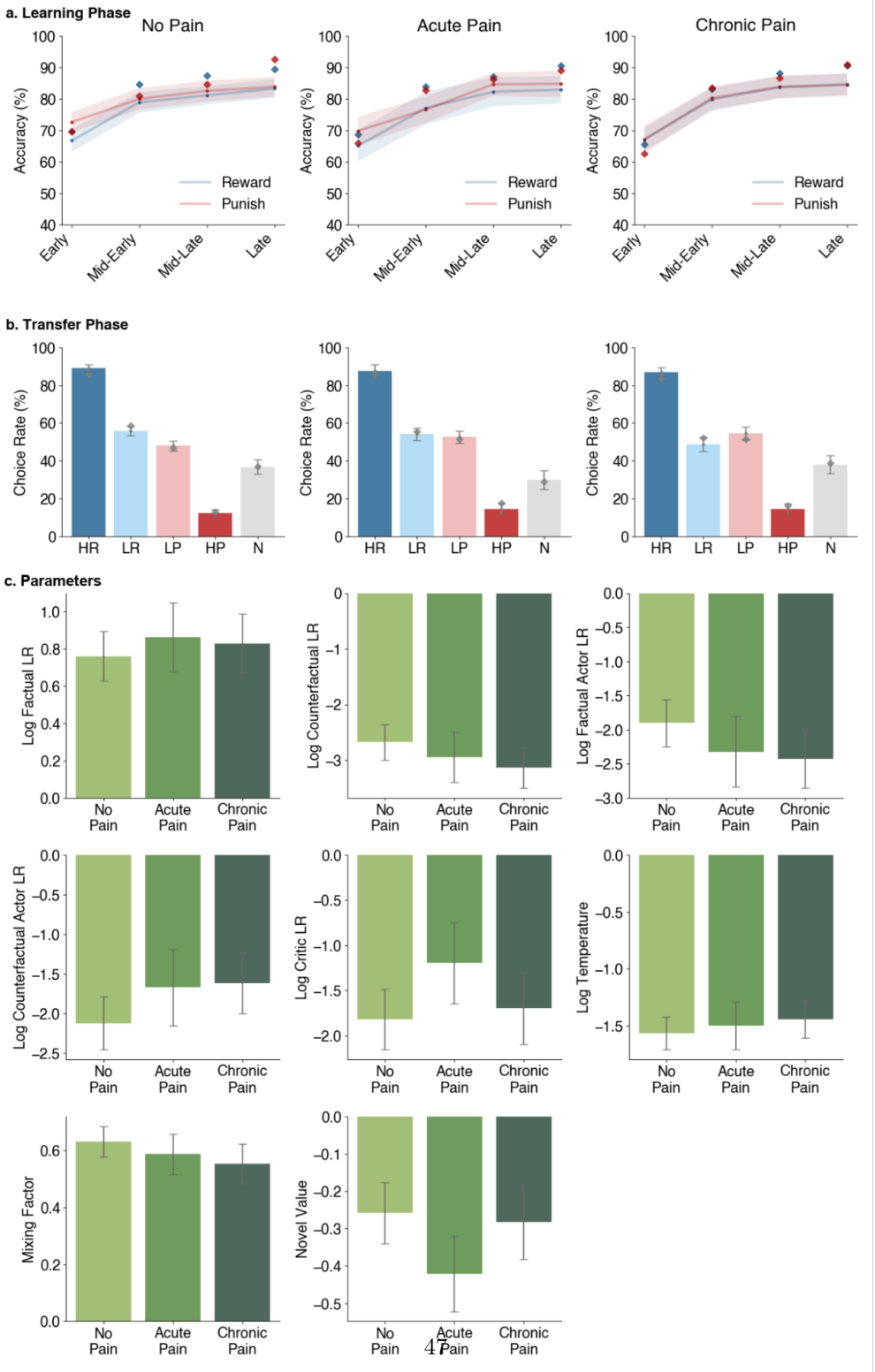

Fig. S10 Posterior Predictive checks for the second runner-up computational model and fitted parameters. Model simulations for the second runner-up model, the Hybrid model. During the fitting procedure, each participant was fitted ten times, and the best of these runs was selected to determine the participant's fitted parameters. We then used each model to simulate a single dataset per participant using these fitted parameters. Thus, these data represent the same sample size as our empirical data. a. Learning Phase: Model performance across binned learning trials for the reward and punishment contexts for each group. Shaded regions represent 95% confidence intervals. Blue and red diamonds indicate empirical means of participant accuracy. b. Transfer Phase: Choice rates for each stimulus type during transfer trials for each group. Choice rate is computed as the percentage of times a stimulus type was chosen, given the number of times it was presented. Bar plots show the mean and 95% confidence intervals of the choice rate for each stimulus type across participants within each group. Grey diamonds indicate empirical means of participant choice rates. Abbreviations: HR – high reward rate (75% reward), LR – low reward rate (25% reward), LP – low punishment rate (25% loss), HP – high punishment rate (75% loss), N – novel stimulus. c. Parameters: Fitted parameter values are displayed as a bar plot, showing the mean and 95% confidence intervals across participants for each group.

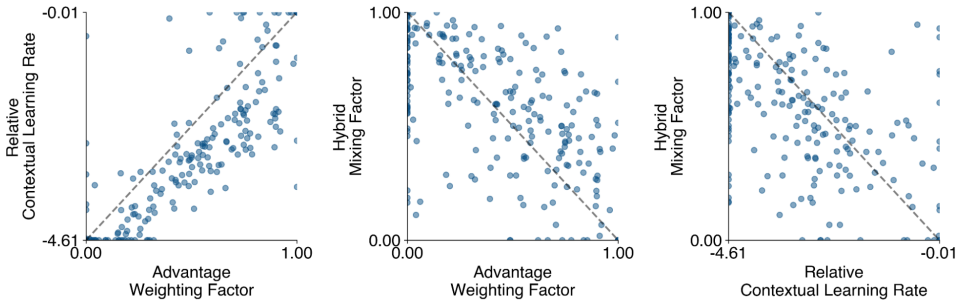

Fig. S11 Associations between mixing factors across the mixed computational models. Each of the Advantage, Relative, and Hybrid models includes a parameter that mixes absolute expected values with context-dependent relative values, albeit in different ways. The fit of these values is conceptually related, and these relationships and quantified here. Greater values in the Advantage and Relative models indicate a larger influence of context-dependent relative values, whereas greater values in the Hybrid model indicate a larger influence of absolute expected values. These parameters are highly correlated, suggesting a common underlying mechanism balancing these two valuation strategies.

Supplementary Tables

Table S1 Analysis of accuracy rates during the learning phase. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for accuracy rates in the learning phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. **Bold p-values** indicate statistically significant effects at $\alpha = .05$.

Effect	df	Test Value	p-value	Cohen's d
Group	2	0.17	.9192	
Chronic vs No Pain	184	-0.72	.4730	-0.11
Chronic vs Acute	129	-0.26	.7991	-0.05
Context	1	2.17	.1406	
Time	3	1117.24	**<.0001**	
Group:Context	2	0.69	.7082	
Chronic vs No Pain: Reward	184	-0.55	.5807	-0.08
Chronic vs No Pain: Punish	184	-0.57	.5679	-0.09
Chronic vs No Pain: Reward–Punish	184	0.03	.9735	0.00
Group:Time	6	19.32	**.0037**	
Context:Time	3	5.46	.1409	
Group:Context:Time	6	12.26	.0565	

Table S2 Analysis of reaction times during the learning phase. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for reaction times in the learning phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. **Bold p-values** indicate statistically significant effects at $\alpha = .05$.

Effect	df	Test Value	p-value	Cohen's d
Group	2	6.24	**.0441**	
Chronic Pain Vs No Pain	184	-3.14	**.0020**	-0.47
Chronic Pain Vs Acute Pain	129	-2.31	**.0224**	-0.41
Context	1	270.38	**<.0001**	
Time	3	1793.61	**<.0001**	
Group:Context	2	10.37	**.0056**	
Chronic Pain Vs No Pain: Reward	145	3.34	**.0011**	0.51
Chronic Pain Vs No Pain: Punish	184	2.67	**.0083**	0.40
Chronic Pain Vs No Pain: Reward–Punish	184	0.81	.4176	0.12
Group:Time	6	23.18	**.0007**	
Context:Time	3	3.68	.2988	
Group:Context:Time	6	3.51	.7423	

Table S3 Analysis of covariates on choice rate during the transfer phase. Age, pain severity, context, and task version were independently added as covariates to the GLMMs assessing choice rates in the transfer phase. These assessments confirm that each covariate does not explain the main findings of our study. Values here represent p-values for each main effect and interaction, as well as the p-value for the corresponding covariate. **Bold p-values** indicate statistically significant effects at $\alpha = .05$.

Effect	**Age**	**Pain**	**Context**	**Task**
Group	.1500	.2176	**.0042**	.1504
Stimulus Type	**<.0001**	**<.0001**	**<.0001**	**<.0001**
Group:Stimulus Type	**.0002**	**.0002**	**<.0001**	**.0002**
Age	.7767	—	—	—
Pain	—	.8222	—	—
Context	—	—	.9873	—
Task	—	—	—	.7279

Table S4 Analysis of reaction times during the transfer phase. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for reaction times in the transfer phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. The planned comparisons of the group x stimulus type interaction use slightly different data than the GLMMs and the other planned comparisons. Specifically, these assess reaction times for each pair (e.g., HR vs LP) only when presented in that pair (rather than reaction times across all pairs). **Bold p-values** indicate statistically significant effects at $\alpha = .05$.

Effect	**df**	**Test Value**	**p-value**	**Cohen's d**
Group	2	0.23	.8893	
Chronic Pain vs No Pain	184	-0.98	.3294	-0.15
Chronic Pain vs Acute Pain	129	-1.10	.2720	-0.20
Stimulus Type	4	374.44	**<.0001**	
High Reward vs Low Punish	238	-15.08	**<.0001**	-0.98
Low Reward vs Low Punish	238	-1.37	.1732	-0.09
Group:Stimulus Type	8	37.89	**<.0001**	
No Pain vs Acute Pain: HR-LP	159	-0.97	.3361	-0.18
No Pain vs Chronic Pain: HR-LP	184	-1.05	.2966	-0.17
Acute Pain vs Chronic Pain: HR-LP	129	0.11	.9095	0.02
No Pain vs Acute Pain: LR-LP	159	-1.03	.3052	-0.18
No Pain vs Chronic Pain: LR-LP	184	-1.48	.1400	-0.23
Acute Pain vs Chronic Pain: LR-LP	129	-0.32	.7528	-0.06

Table S5 Fitted parameters for each model parameter for each group. Values represent the averaged fitted values, with standard deviations in brackets, for each parameter within each model for each group.

Model	Parameter	No Pain	Acute Pain	Chronic Pain
QLearning	Factual Learning Rate	0.24 (0.32)	0.21 (0.26)	0.30 (0.35)
	Counterfactual Learning Rate	0.23 (0.32)	0.22 (0.30)	0.24 (0.29)
	Temperature	0.36 (0.25)	0.42 (0.26)	0.46 (0.28)
	Novel Value	-0.27 (0.46)	-0.46 (0.43)	-0.26 (0.54)
ActorCritic	Factual Actor Learning Rate	0.30 (0.30)	0.28 (0.29)	0.33 (0.30)
	Counterfactual Actor Learning Rate	0.27 (0.29)	0.38 (0.36)	0.39 (0.35)
	Critic Learning Rate	0.21 (0.35)	0.31 (0.40)	0.20 (0.31)
	Temperature	0.38 (0.21)	0.43 (0.21)	0.43 (0.24)
	Novel Value	-0.34 (0.50)	-0.56 (0.40)	-0.34 (0.49)
Relative	Factual Learning Rate	0.18 (0.23)	0.14 (0.20)	0.16 (0.21)
	Counterfactual Learning Rate	0.15 (0.22)	0.15 (0.19)	0.15 (0.21)
	Contextual Learning Rate	0.10 (0.21)	0.15 (0.27)	0.18 (0.29)
	Temperature	0.23 (0.21)	0.24 (0.21)	0.23 (0.22)
	Novel Value	-0.19 (0.39)	-0.30 (0.32)	-0.15 (0.37)
Advantage	Factual Learning Rate	0.18 (0.24)	0.18 (0.23)	0.21 (0.28)
	Counterfactual Learning Rate	0.17 (0.24)	0.16 (0.20)	0.17 (0.23)
	Temperature	0.24 (0.22)	0.25 (0.21)	0.26 (0.24)
	Weighting Factor	0.35 (0.32)	0.43 (0.32)	0.48 (0.35)
	Novel Value	-0.21 (0.39)	-0.32 (0.31)	-0.18 (0.40)
Hybrid	Factual Learning Rate	0.20 (0.29)	0.22 (0.30)	0.24 (0.34)
	Counterfactual Learning Rate	0.23 (0.34)	0.18 (0.28)	0.16 (0.28)
	Factual Actor Learning Rate	0.41 (0.38)	0.32 (0.35)	0.31 (0.35)
	Counterfactual Actor Learning Rate	0.35 (0.38)	0.46 (0.39)	0.46 (0.39)
	Critic Learning Rate	0.42 (0.39)	0.60 (0.41)	0.47 (0.42)
	Temperature	0.28 (0.22)	0.29 (0.22)	0.31 (0.24)
	Mixing Factor	0.63 (0.28)	0.59 (0.25)	0.55 (0.31)
	Novel Value	-0.26 (0.43)	-0.42 (0.37)	-0.28 (0.45)

Table S6 Analysis of model-simulated accuracy rates during the learning phase. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for the best-fitting Advantage model simulated accuracy rates in the learning phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. **Bold p-values** indicate statistically significant effects at $\alpha = .05$. Rep = replication and indicates whether the model simulations resulted in the same statistical outcome (significant or not significant) as the empirical parallels.

Effect	df	Test Value	p-value	Rep
Group	2	2.76	.2517	Yes
Chronic Pain Vs No Pain	184	-0.57	.5716	Yes
Chronic Pain Vs Acute Pain	129	0.89	.3760	Yes
Context	1	0.01	.9273	Yes
Time	3	816.70	**<.0001**	Yes
Group:Context	2	2.73	.2552	Yes
Chronic Pain Vs No Pain: Reward	184	-0.47	.6411	Yes
Chronic Pain Vs No Pain: Punish	184	-0.51	.6106	Yes
Chronic Pain Vs No Pain: Reward-Punish	184	0.06	.9547	Yes
Group:Time	6	3.24	.7781	No
Context:Time	3	5.84	.1196	Yes
Group:Context:Time	6	16.17	**.0128**	No

Table S7 Analysis of model-simulated choice rates during the transfer phase. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for the best-fitting Advantage model simulated choice rates in the transfer phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. Just as with empirical findings, the planned comparisons of the group x stimulus type interaction use slightly different data than the GLMMs and the other planned comparisons. Specifically, these assess choice rates for each pair (e.g., HR vs LP) only when presented in that pair (rather than choice rates across all pairs). **Bold p-values** indicate statistically significant effects at $\alpha = .05$. Rep = replication and indicates whether the model simulations resulted in the same statistical outcome (significant or not significant) as the empirical parallels.

Effect	df	Test Value	p-value	Rep
Group	2	1.28	.5272	Yes
Chronic Pain Vs No Pain	184	0.75	.4570	No
Chronic Pain Vs Acute Pain	129	2.82	**.0055**	Yes
Stimulus Type	4	791.45	**<.0001**	Yes
High Reward Vs Low Punish	238	29.03	**<.0001**	Yes
Low Reward Vs Low Punish	238	0.76	.4491	No
Group:Stimulus Type	8	39.41	**<.0001**	Yes
No Pain Vs Acute Pain: HR-LP	159	-2.47	**.0145**	No
No Pain Vs Chronic Pain: HR-LP	184	-3.46	**.0007**	No
Acute Pain Vs Chronic Pain: HR-LP	129	-0.55	.5799	Yes
No Pain Vs Acute Pain: LR-LP	159	1.25	.2141	Yes
No Pain Vs Chronic Pain: LR-LP	184	3.04	**.0027**	Yes
Acute Pain Vs Chronic Pain: LR-LP	129	1.48	.1406	Yes

Table S8 Associations between model parameters and pain severity metrics. Results of Pearson r correlational analyses linking the winning model's parameter fits with pain severity metrics. Values indicate Pearson r correlations with p-values in brackets. **Bold p-values** indicate statistically significant effects at $\alpha = .05$. LR = Learning Rate.

Group	Parameter	Intensity	Unpleasant	Interference	Composite
No Pain					
	Factual LR	-.02 (.8558)	-.01 (.9072)	-.12 (.2158)	-.06 (.5396)
	Counterfactual LR	.04 (.6821)	.12 (.2355)	.16 (.1016)	.12 (.2132)
	Temperature	.02 (.8657)	.04 (.6570)	.08 (.4185)	.05 (.5811)
	Weighting Factor	-.10 (.3253)	-.18 (.0674)	-.05 (.6094)	-.12 (.2115)
	Novel Value	.03 (.7238)	.09 (.3403)	.00 (.9840)	.05 (.6259)
Acute Pain					
	Factual LR	-.03 (.8254)	.08 (.5549)	.12 (.3937)	.06 (.6593)
	Counterfactual LR	.16 (.2485)	.05 (.7446)	.13 (.3694)	.12 (.3920)
	Temperature	**.31 (.0251)**	.21 (.1241)	.18 (.2001)	.25 (.0677)
	Weighting Factor	-.10 (.4600)	-.21 (.1401)	-.20 (.1561)	-.18 (.1893)
	Novel Value	-.17 (.2156)	-.23 (.0950)	-.22 (.1197)	-.22 (.1062)
Chronic Pain					
	Factual LR	.20 (.0740)	**.24 (.0318)**	**.30 (.0077)**	**.27 (.0156)**
	Counterfactual LR	.19 (.1000)	**.26 (.0212)**	**.23 (.0450)**	**.25 (.0306)**
	Temperature	.21 (.0591)	**.25 (.0282)**	**.23 (.0433)**	**.25 (.0267)**
	Weighting Factor	-.07 (.5265)	-.07 (.5714)	.05 (.6904)	-.03 (.7972)
	Novel Value	-.19 (.0923)	-.21 (.0617)	-.14 (.2123)	-.20 (.0849)

Table S9 Analysis of demographic and pain metrics across groups. Main effects of between-subject ANOVAs (**bold**) and planned comparisons (indented and *italicized*) for each metric assessing group differences. Test value represents F-values for ANOVA effects and t-values for planned comparisons. **Bold p-values** indicate statistically significant effects at $\alpha = .05$.

Effect	df	Test Value	p-value	Cohen's d
Age	2, 239	7.82	< **.0001**	
Chronic vs No Pain	154	4.00	< **.0001**	0.61
Chronic vs Acute	129	2.06	< **.0001**	0.37
Pain Intensity	2, 239	240.50	< **.0001**	
Chronic vs No Pain	87	20.25	< **.0001**	3.45
Chronic vs Acute	129	1.65	.1013	0.29
Pain Unpleasantness	2, 239	293.46	< **.0001**	
Chronic vs No Pain	84	22.11	< **.0001**	3.79
Chronic vs Acute	129	1.34	.1834	0.24
Pain Interference	2, 239	246.91	< **.0001**	
Chronic vs No Pain	84	18.78	< **.0001**	3.22
Chronic vs Acute	129	0.46	.6466	0.08
Pain Composite	2, 239	308.00	< **.0001**	
Chronic vs No Pain	84	22.15	< **.0001**	3.80
Chronic vs Acute	129	1.23	.2215	0.22

Table S10 Counts of body regions where pain originates. Self-reported locations of pain origin in participants from task version B ($n = 155$). These data were not collected for participants from task version A. Participants could indicate multiple regions. The no pain group was excluded as this question is irrelevant without pain. Twenty participants did not respond to this question.

Region	**Acute Pain**	**Chronic Pain**
Back or neck pain	5	20
Scoliosis	4	6
Osteoarthritis	1	5
Headache	6	4
Migraine	6	4
Irritable bowel syndrome (IBS)	2	3
Temporomandibular joint disorder (TMJ)	0	3
Knee	1	3
Herniation	0	2
Chronic musculoskeletal pain	0	2
Multiple Sclerosis	0	1
Shingles	0	1
Degenerative Disc Disease	0	1
Rheumatoid/osteoarthritis	0	1
Rheumatoid arthritis	1	1
Fibromyalgia	0	1
Post-stroke pain	0	1
Back	0	1
Hip	0	1
Shoulder	0	1
Ankle	2	0
Hand	1	0
Foot	1	0
Other	5	6

Table S11 Analysis of covariates on accuracy during the learning phase. Age, pain severity, and task version were independently added as covariates to the GLMMs assessing accuracy in the learning phase. These assessments confirm that each covariate does not explain the main findings of our study. Values represent p-values for each main effect and interaction, as well as the p-value for the corresponding covariate. **Bold p-values** indicate statistically significant effects at $\alpha = .05$.

Effect	**Age**	**Pain**	**Task**
Group	.9035	.9171	.9266
Context	.1407	.1407	.1406
Time	**<.0001**	**<.0001**	**<.0001**
Group:Context	.7084	.7084	.7082
Group:Time	**.0037**	**.0037**	**.0037**
Context:Time	.1409	.1409	.1409
Group:Context:Time	.0565	.0565	.0565
Age	.7635	—	—
Pain	—	.8340	—
Task	—	—	.9524

Table S12 Priors and Bounds for Each Model and Parameter. The priors and bounds used during model fitting of empirical data for each parameter within each model. Priors represent the mean of the normal distribution used as starting points, and bounds constrain each parameter's valid range.

Model	**Parameter**	**Prior**	**Bounds**
Q-Learning	Factual Learning Rate	0.28	(0.01, 0.99)
	Counterfactual Learning Rate	0.18	(0.01, 0.99)
	Temperature	0.06	(0.01, 1.00)
	Novel Value	0.00	(-1.00, 1.00)
Actor-Critic	Factual Actor Learning Rate	0.33	(0.01, 0.99)
	Counterfactual Actor Learning Rate	0.33	(0.01, 0.99)
	Critic Learning Rate	0.48	(0.01, 0.99)
	Temperature	0.06	(0.01, 1.00)
	Novel Value	0.00	(-1.00, 1.00)
Relative	Factual Learning Rate	0.19	(0.01, 0.99)
	Counterfactual Learning Rate	0.15	(0.01, 0.99)
	Contextual Learning Rate	0.33	(0.01, 0.99)
	Temperature	0.05	(0.01, 1.00)
	Novel Value	0.00	(-1.00, 1.00)
Advantage	Factual Learning Rate	0.28	(0.01, 0.99)
	Counterfactual Learning Rate	0.18	(0.01, 0.99)
	Temperature	0.06	(0.01, 1.00)
	Weighting Factor	0.50	(0.00, 1.00)
	Novel Value	0.00	(-1.00, 1.00)
Hybrid	Factual Learning Rate	0.49	(0.01, 0.99)
	Counterfactual Learning Rate	0.49	(0.01, 0.99)
	Factual Actor Learning Rate	0.33	(0.01, 0.99)
	Counterfactual Actor Learning Rate	0.33	(0.01, 0.99)
	Critic Learning Rate	0.48	(0.01, 0.99)
	Temperature	0.06	(0.01, 1.00)
	Mixing Factor	0.70	(0.00, 1.00)
	Novel Value	0.00	(-1.00, 1.00)

Table S13 Parameter recovery results. Pearson r correlations determining the degree to which each parameter within each model was recoverable. Data were generated for each model using randomly determined parameters and then fitted by that model to assess the model's ability to recover parameters. Aggregated values indicate the averaged Pearson r across parameters within each model. Higher values indicate better recovery.

Model	**Parameter**	**Pearson** r
Q-Learning	Factual Learning Rate (α_c)	0.97
	Counterfactual Learning Rate (α_u)	0.97
	Temperature (τ)	0.77
	Novel Value (η)	0.99
	Aggregated	**0.93**
Actor-Critic	Factual Actor Learning Rate (α_c)	0.80
	Counterfactual Actor Learning Rate (α_{ac})	0.73
	Critic Learning Rate (α_{au})	0.63
	Temperature (τ)	0.62
	Novel Value (η)	0.97
	Aggregated	**0.75**
Relative	Factual Learning Rate (α_c)	0.95
	Counterfactual Learning Rate (α_u)	0.92
	Contextual Learning Rate (α_v)	0.73
	Temperature (τ)	0.81
	Novel Value (η)	0.99
	Aggregated	**0.88**
Advantage	Factual Learning Rate (α_c)	0.93
	Counterfactual Learning Rate (α_u)	0.92
	Temperature (τ)	0.82
	Weighting Factor (m)	0.98
	Novel Value (η)	0.99
	Aggregated	**0.93**
Hybrid	Factual Learning Rate (α_{qc})	0.59
	Counterfactual Learning Rate (α_{qu})	0.56
	Factual Actor Learning Rate (α_{wac})	0.56
	Counterfactual Actor Learning Rate (α_{wau})	0.42
	Critic Learning Rate (α_{wc})	0.22
	Temperature (τ)	0.70
	Mixing Factor (c)	0.87
	Novel Value (η)	0.98
	Aggregated	**0.61**

Table S14 Analysis of model-simulated accuracy rates during the learning phase for the Relative model. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for the runner-up Relative model simulated accuracy rates in the learning phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. **Bold p-values** indicate statistically significant effects at $\alpha = .05$. *Rep* = replication and indicates whether the model simulations resulted in the same statistical outcome (significant or not significant) as the empirical parallels.

Effect	df	Test Value	p-value	Rep
Group	2	2.76	.2516	Yes
Chronic Pain Vs No Pain	184	-0.87	.3860	Yes
Chronic Pain Vs Acute Pain	129	0.85	.3956	Yes
Context	1	0.01	.9203	Yes
Time	3	816.7	< **.0001**	Yes
Group:Context	2	2.73	.2554	Yes
Chronic Pain Vs No Pain: Reward	145	-1.51	.1341	Yes
Chronic Pain Vs No Pain: Punish	184	0.11	.9119	Yes
Chronic Pain Vs No Pain: Reward-Punish	135	-1.64	.1040	Yes
Group:Time	6	3.24	.7782	No
Context:Time	3	5.84	.1200	Yes
Group:Context:Time	6	16.17	**.0130**	Yes

Table S15 Analysis of model-simulated choice rates during the transfer phase for the Relative model. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for the runner-up Relative model simulated choice rates in the transfer phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. The planned comparisons of the group x stimulus type interaction use slightly different data than the GLMMs and the other planned comparisons; specifically, these assess choice rates for each pair (e.g., HR vs LP) only when presented in that pair. **Bold p-values** indicate statistically significant effects at $\alpha = .05$. *Rep* indicates whether the model simulations resulted in the same statistical outcome as the empirical parallels.

Effect	df	Test Value	p-value	Rep
Group	2	1.28	.5270	Yes
Chronic Pain Vs No Pain	184	0.30	.7648	No
Chronic Pain Vs Acute Pain	126	2.99	**.0033**	Yes
Stimulus Type	4	791.45	< **.0001**	Yes
High Reward Vs Low Punish	238	24.47	< **.0001**	Yes
Low Reward Vs Low Punish	238	-0.29	.7739	No
Group:Stimulus Type	8	39.41	< **.0001**	Yes
No Pain Vs Acute Pain: HR-LP	159	-2.09	**.0384**	No
No Pain Vs Chronic Pain: HR-LP	184	-2.12	**.0354**	No
Acute Pain Vs Chronic Pain: HR-LP	129	0.12	.9027	Yes
No Pain Vs Acute Pain: LR-LP	159	1.17	.2454	Yes
No Pain Vs Chronic Pain: LR-LP	184	2.78	**.0060**	Yes
Acute Pain Vs Chronic Pain: LR-LP	129	1.23	.2191	Yes

Table S16 Analysis of model-simulated accuracy rates during the learning phase for the Hybrid model. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for the second runner-up Hybrid model simulated accuracy rates in the learning phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. **Bold p-values** indicate statistically significant effects at $\alpha = .05$. *Rep* indicates whether the model simulations resulted in the same statistical outcome as the empirical parallels.

Effect	df	Test Value	p-value	Rep
Group	2	2.76	.2516	Yes
Chronic Pain Vs No Pain	184	-1.37	.1734	Yes
Chronic Pain Vs Acute Pain	129	0.65	.5147	Yes
Context	1	0.01	.9203	Yes
Time	3	816.7	< **.0001**	Yes
Group:Context	2	2.73	.2554	Yes
Chronic Pain Vs No Pain: Reward	184	-0.01	.9921	Yes
Chronic Pain Vs No Pain: Punish	142	-2.17	**.0316**	No
Chronic Pain Vs No Pain: Reward-Punish	184	2.04	**.0427**	No
Group:Time	6	3.24	.7782	No
Context:Time	3	5.84	.1197	Yes
Group:Context:Time	6	16.17	**.0129**	No

Table S17 Analysis of model-simulated choice rates during the transfer phase for the Hybrid model. Statistical outcomes of GLMMs (**bold effects**) and planned comparisons (indented and *italicized effects*) for the second runner-up Hybrid model simulated choice rates in the transfer phase. Test value represents χ^2 for GLMM effects and t-values for planned comparisons. Just as with empirical findings, the planned comparisons of the group x stimulus type interaction use slightly different data than the GLMMs and the other planned comparisons; specifically, these assess choice rates for each pair (e.g., HR vs LP) only when presented in that pair (rather than choice rates across all pairs). **Bold p-values** indicate statistically significant effects at $\alpha = .05$. *Rep* = replication and indicates whether the model simulations resulted in the same statistical outcome (significant or not significant) as the empirical parallels.

Effect	df	Test Value	p-value	Rep
Group	2	1.28	.5270	Yes
Chronic Pain Vs No Pain	184	0.62	.5377	No
Chronic Pain Vs Acute Pain	129	2.86	**.0049**	Yes
Stimulus Type	4	791.45	< **.0001**	Yes
High Reward Vs Low Punish	238	28.62	< **.0001**	Yes
Low Reward Vs Low Punish	238	1.45	.1475	No
Group:Stimulus Type	8	39.41	< **.0001**	Yes
No Pain Vs Acute Pain: HR-LP	159	-1.8	.0731	No
No Pain Vs Chronic Pain: HR-LP	184	-2.17	**.0314**	No
Acute Pain Vs Chronic Pain: HR-LP	129	-0.12	.9073	Yes
No Pain Vs Acute Pain: LR-LP	159	1.44	.1525	Yes
No Pain Vs Chronic Pain: LR-LP	148	2.75	**.0067**	Yes
Acute Pain Vs Chronic Pain: LR-LP	127	1.19	.2370	Yes