Neurocomputational impairments in disambiguation of context as a key determinant of post-traumatic psychopathology

Adam Linson
School of Computing & Communications, Faculty of STEM, Open University, Edinburgh, UK, adam.linson@open.ac.uk

Heidi Anderson
Shields & Stripes, North Carolina, USA

Sarah Markham
Institute of Psychiatry, Psychology and Neuroscience, King's College London, UK

Karl Friston
Department of Imaging Neuroscience, Queen Square Institute of Neurology, University College London, London, UK

Michael Moutoussis
Department of Imaging Neuroscience, Queen Square Institute of Neurology, University College London, London, UK

(1-Aug-2025)

Abstract

We combine theoretical biology and systems neuroscience to relate the mechanisms of post-traumatic stress disorder (PTSD) models to biopsychosocial complexity. Moving beyond fear conditioning, extinction and habits formed by associative learning, we consider conserved neuromodulatory mechanisms related to harm expectation, defensive mobilisation, and context (in)sensitivity. Our account of harm-mitigating, neurobehavioural responses furnishes a mechanistic understanding of maladaptive behaviour and impaired belief-updating, which are often underspecified in conceptual analyses. Grounded in neurobiological evidence, we specify PTSD-related impairment in terms of Bayesian belief-updating and establish an adaptive/maladaptive boundary using active inference and simulation studies. The ensuing model plausibly illustrates PTSD pathogenesis, which remains understudied relative to aetiology. We locate pathogenesis in maladaptive precision following traumatic learning, characterised by aberrant Bayesian credit assignment. Future scientific and clinical studies can test the validity of these insights, which could underwrite our understanding of the phenomenology, diagnostics and ultimately treatment intervention in post-traumatic disorders.

During traumatic experiences, the imperative to mitigate against severe harm requires an intense mobilisation of biological subsystems, an 'emergency response mode'.[1,2] However, multiple factors involved in the traumatic experience, its traumatic memory and the experience-dependent plasticity of the traumatised person can lead this emergency response mode to become persistent, that is, maladaptive. Here, we focus on modelling post-traumatic pathogenesis as the process by which a trauma schema (or 'mode') becomes persistent and, in turn, causes a wide range of post-traumatic stress syndromes (PTSS), including those related to mental imagery (e.g., 'flashbacks'). The modelling also illustrates the process of returning to a pre-traumatic response mode, to suggest how PTSS treatment interventions could target the root cause of symptoms. In contrast to reductionist accounts that can lead to over-medicalisation, this computational approach is consistent with a biopsychosocial-environmental understanding of post-traumatic disorders.

The present work models conserved neuromodulatory and sensorimotor mechanisms related to harm expectation, defensive mobilisation,[3,4] and the associative learning of causes and consequences (model-based learning).[5,6] The modelling remains compatible with multiple accounts of fear and extinction,[7–10] without relying on these constructs, by shifting the focus to the general construct of harm. Perceptibly harmful administrations of pain, electric shock, restraint, etc., are common to experimental paradigms of fear conditioning and extinction in (non-human) animals. A capacity for harm perception, expectation, and mitigation is necessarily evolutionarily conserved, as an abundance of harm negatively impacts the survival of individuals and species.

The notion of expectation has been readily operationalised using a range of comparable theories including Hebbian learning, spike-timing dependent plasticity (STDP), predictive coding, and embodied cognition frameworks broadly known under 'Bayesian brain' or 'predictive processing' descriptions, sometimes specified as active inference (applied here). These theoretical approaches relate to methodologies for understanding and measuring correspondences between neural substrates and behaviour. Computational psychiatry commonly incorporates these methodologies,[11] found in an emerging body of research on understanding traumatic experience and PTSS.[1,12–14]

Neurobehavioural response patterns necessary for survival form the basis of our PTSD modelling. The sequence of agent-environment interactions engaging these neural mechanisms are: (1) learning about the sources of harm in multiple contexts; (2) learning when harm is probable; (3) making a (fallible) judgement about the likelihood of imminent harm; and (4) responding defensively only when harm is likely. Grounded in the anticipation of harm under varied conditions, the model remains consistent with experimental studies of fear conditioning and extinction in animal models,[9] in addition to observational and case studies of human PTSD and related clinical trials.[15] Nevertheless, we argue that probabilistic learning about contexts is central to understanding post-traumatic impairments,[5,16] beyond associative (model-free) mechanisms often used to describe traumatic learning.

There are, therefore, extensive commonalities between non-traumatic and traumatic experiences. By drawing on the sciences of visual attention, multimodal perception, memory, and sleep and combining them with neuromodulatory systems neuroscience, we articulate a general neuropsychological and neurobehavioral account that illustrates a putative pathogenesis of disorders related to sequelae of traumatic stress including PTSD and CPTSD. We then invite Experts by Lived Experience (ELEs) of trauma to consult and critique this account. This novel neurocomputational model thus captures biopsychosocial-environmental dynamics, relevant to the understanding and future treatment of these disorders.

Overview

A brief formulation of our analysis is linked below to modelling (see Results) and background literature (see Online Methods). Throughout, we underscore contiguities between realistic experiential narratives and common experimental paradigms (e.g., psychophysics). This approach extends our theoretical account beyond neurobiology and facilitates future empirical testing.[17]

From the standpoint of an individual in a new situation, there may be ambiguity as to whether the surrounding context is one in which harm is likely (harmful context) or unlikely (harmless context). Ordinarily, determinations of context (harmful vs. harmless) are made by gathering sensory evidence to reach a decision. The belief that I am in one context and not the other should ideally persist to some extent as a (contextual) prior belief, but should also decay, in order to strike a balance between (a) maintaining the belief that I am still in the same context and (b) re-assessing whether the context has changed.

Our theory describes PTSD pathogenesis as a process by which the balance between harmless and harmful contextual priors is lost. Here, the prior belief in a harmful context grows (in precision) beyond a critical tipping point. Once reached, the aforementioned decay rate is insufficient for rebalancing with contextual priors for harmless context.

Following traumatic exposure to a harmful context, the imbalance occurs depending on resilience or vulnerability. For a resilient phenotype, the learning rate for harmful context is sufficient to preclude an imbalance between contextual priors. For a vulnerable phenotype, if prior beliefs about the presence of a harmful context become unduly strong, then opportunities for re-assessing context will be continually foregone. The foregoing of contextual re-assessment will persist for harmless contexts, due to the (false) belief they are harmful. Moreover, without re-assessment, a (false) conclusion that the context is harmful will in turn be (mis)counted as new (false) evidence that the context is indeed harmful – thus continually strengthening the false contextual belief in a positive feedback loop, a 'vicious circle'. This miscounting can be described as part of a confirmation bias in perceptual decision-making modelled by hierarchical approximate inference.[18]

If PTSD pathogenesis occurs, a possible treatment intervention relates to the provision of evidence that I am in a harmless context, rather than a harmful one as believed *a priori*. For the treatment to succeed, I must accumulate enough evidence to update my contextual priors. Only then can I seek out novel situations and determine their true context, rather than giving undue weight to the harmful prior.

Below, we introduce a neuromodulatory circuit for a mechanistic understanding of PTSD-related phenomena. We then illustrate PTSD pathogenesis and treatment in model-based simulation studies (see Results). To draw on the literature that supports our account, we also present an interlinked cognitive psychological framing of processes and mechanisms related to context recognition (see Online Methods). Under this framing, we identify contextual interactions with what is widely understood as Visual Working Memory (VWM). We also cover contextual interactions between long-term memory (LTM) and task-related generalisation. Interested readers may wish to have this general cognitive psychological description in place first as a foundation.

Neuromodulatory circuits and PTSD pathogenesis

We first focus on putative functional neuroanatomy, to hypothesize the neural pathways that subserve the psychological, cognitive, and physiological processes involved in our account. As an initial step, we draw connections between PTSS and a pontine neuromodulatory subsystem. We then provide a schema that brings this subsystem into correspondence with

non-pathological and pathological neurobehavioral dynamics and relate these to psychiatric diagnostic criteria. Finally, we elaborate a long-range circuit relevant to pathogenesis and treatment of PTSD.

The reticular activating system (RAS) in the brainstem, implicated in sleep-wake cycle regulation, is a neurobiological locus of a range of sleep disorders connected to psychiatric conditions.[19–21] Under typical functioning, excitatory cholinergic projections from the pedunculopontine tegmental nucleus (PPtg) to the locus coeruleus (LC) stimulate an inhibitory noradrenergic projection from the LC to the PPtg. In this PPtg-LC (excitation-inhibition) interaction, acetylcholine (ACh) and noradrenaline (NA) are mutually stabilising, resulting in a balance characteristic of attentive wakefulness. At the initiation of sleep, the PPtg continues to emit ACh but temporarily ceases excitation of the LC, resulting in a marked rise in ACh and diminishment of NA. If, however, this pontine ACh/NA subsystem is at an (aberrant) elevated tonic level, one would expect sleep to involve vivid trauma replay nightmares (elevated ACh) and gross body movements (elevated NA), both of which are characteristic of PTSD.[21–25]

The elevated tonic level may be held over from an adaptive (i.e., situationally apt) phasic elevation during a wakeful traumatic experience. During such an event, phasic elevation would aid harm mitigation efforts by increasing ACh for selective sensory gain and increasing NA for rapid behavioural response. Known PTSS are depicted in Fig. 1, in relation to ACh and NA levels.

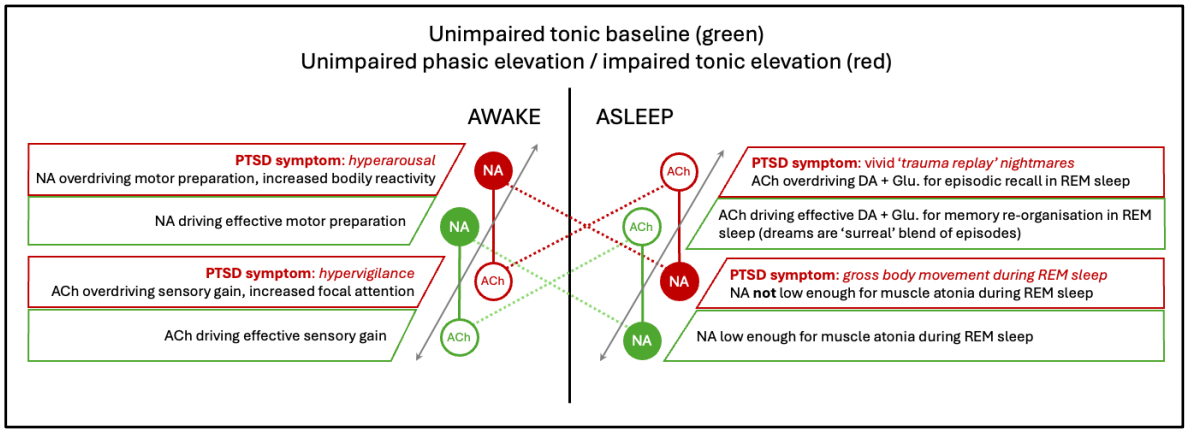


FIGURE 1. Across the central axis, the difference in relative ACh and NA levels in awake and asleep states is depicted (diagonal dotted lines). Green is used to indicate unimpaired tonic levels that raise to phasic levels during a traumatic experience and return to unimpaired tonic levels (green) for a resilient individual (movement along the grey arrows). A vulnerable individual may remain at elevated phasic levels (red) such that they become tonic. The text boxes indicate corresponding consequences of ACh and NA when unelevated or elevated, including PTSD symptoms under impaired tonic elevation (awake and asleep).

Following such an event, phasic levels should fall ahead of healthy sleep, such that memory can be re-encoded and re-organised in relation to event boundaries. Future reactions to trauma cues in new contexts could deploy attention to the sensory periphery to update beliefs about the veridical context (see Fig. 8, boxes A and 4A). However, if the phasic increase has not subsided prior to sleep, then disrupted sleep can hinder memory re-encoding and re-organisation, resulting in stronger subsequent activation during traumatic recall. This overly strong reactivation can provide sufficient evidence for a false context such that belief-updating about veridical conditions is hindered by peripheral sensory attenuation (see Fig. 8, boxes B_1, B_2, and 4B). If the phasic elevation endures, it can become effectively a new elevated tonic baseline, with wakeful hypervigilance and hyperarousal. Fig. 2 illustrates how phasic elevation can become tonic elevation, and relates this to PTSD diagnostic criteria.

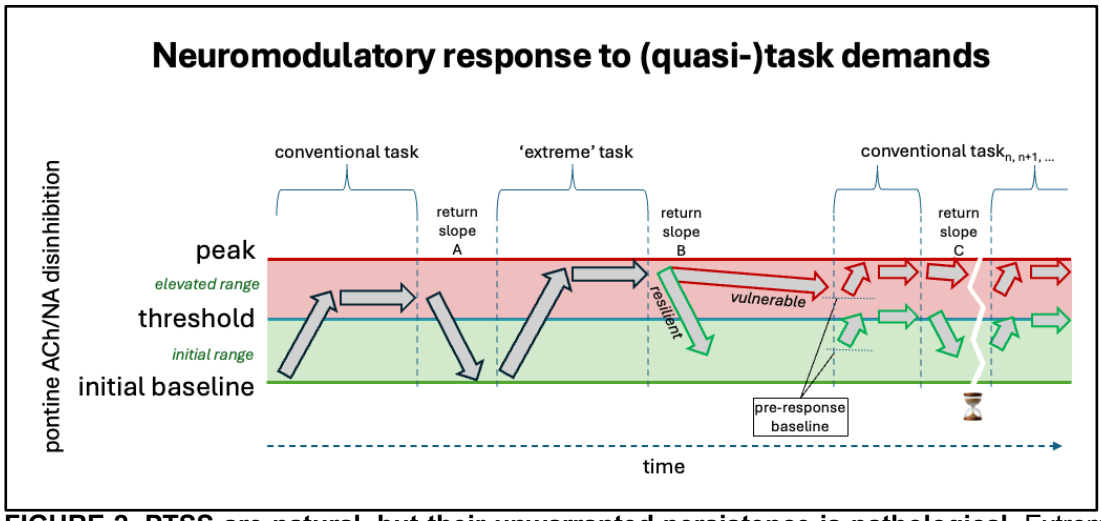


FIGURE 2. PTSS are natural, but their unwarranted persistence is pathological. Extreme task demands, as when life is at stake, elicit intense physiological responses, with physiological indicators peaking at high values (a time-course of sympathetic nervous system activation: increased heart rate (HR) and HR variability, blood pressure, breathing rate, adrenaline levels, vigorous motor preparation, etc.). How long it takes a defensive mobilisation to fully subside will depend on the elicited response (i) strength and (ii) the rate of response decay following a transition from a high to low probability of harm. An additional factor is (iii) the baseline immediately preceding the response, a local rather than global minimum ('pre-response baseline'). If defences are already partly mobilised, as when on high alert, the distance between the pre-response baseline and peak mobilisation will be reduced (upper right arrows).

For a diagnosis of post-traumatic stress disorder (PTSD), the prototypical disorder of its kind, symptoms must last "longer than a month" according to the DSM-5 but "at least several weeks" according to the ICD-11. From a computational point of view, the optimal decay rate should be determined by the persistence of high probability of harm. If, say, serious harm is likely to occur within three weeks, high alert should not stop within one week. Hence, we must not pathologize the initial occurrence of PTSD symptoms, but focus on how their nature increases the risk of unwarranted, maladaptive persistence.

We postulate a critical relationship between a 'return slope' (B) and its intercepting a threshold separating extreme from ordinary arousal. This intercept characterises the transition out of PTSS – the time axis would be set differently in different diagnostic classifications and related to different neurobiological indicators (e.g., neurochemical and connectome properties[26]). Resilience can then be related to the ability to clearly infer that one no longer faces an increased likelihood of harm, and rapidly return to the psychophysiological baseline if so (ability to 'bounce back'). A vulnerability to PTSS can be characterised with an inappropriately sluggish decay of arousal. Vulnerability factors may include genetic predispositions and lifespan events such as earlier psychological trauma or physical traumatic brain injury (TBI).

The wide arrows indicate neuromodulatory responses as temporal increases and decreases in pontine acetylcholine (ACh) and noradrenaline (NA) levels (see Fig. 1 for corresponding symptoms and Fig. 3 for the circuit that controls these changes). Behaviourally, the responses relate to a naturalistic account of tasks (the term 'task' is used to link to experimental studies), divided into conventional tasks (e.g., securing food to satisfy hunger) and 'extreme' tasks (e.g., protecting oneself from serious harm). Following an extreme task (e.g., surviving a life-threatening encounter), defences may persist in a mobilised state (red, 'elevated range'). Subsequent conventional tasks that would not have elicited a peak response when initiated at a low pre-response baseline instead elicit a peak response, as they are continually initiated at a high pre-response baseline (upper right arrows). The high pre-response baseline will be maintained continuously given a slow decay rate ('vulnerable') and subsequent encounters with relevant sensory cues (described below). Thus, responses in conventional situations may frequently reach peak defensive mobilisation, constituting PTSS.

One 'upstream' reason the pontine ACh-NA subsystem may be disordered in wake and sleep is reduced activation of the serotonergic (5HT) inhibitory 'brake' from the Dorsal Raphe Nucleus (DRN). The inhibition and disinhibition emerges through a complex interplay of GABA and glutamate (Glu.) from the prefrontal cortex (PFC), limbic systems (LS), and basal ganglia (BG).[27] Notably, altered GABA and Glu. levels and ratios have been found in PTSD impairment in contrast to unimpaired controls.[28] Taken together, these findings suggest a long-range circuit, by which forebrain interactions across the anterior cingulate cortex (ACC) can dampen DRN activity, thereby reducing the inhibitory influence of 5HT on pontine subsystems including the RAS, and especially the PPtg and LC. The PPtg and LC neuronal populations are the primary emitters of ACh and NA, respectively, with innervation throughout the brain in addition to reciprocal projections.

Excessive strength of the initial encoding of a traumatic experience is suggested by the increase in cholinergic stimulation from the PPtg to the Ventral Tegmental Area (VTA) and Substantia Nigra (SN). This ACh stimulation in turn stimulates a dopaminergic (DA) pathway to the forebrain (PFC, LS, and other parts of the BG). A DA surge is known to increase memory plasticity and encoding strength, especially during emotionally significant experience.[29–31] The full circuit is depicted in Fig. 3.

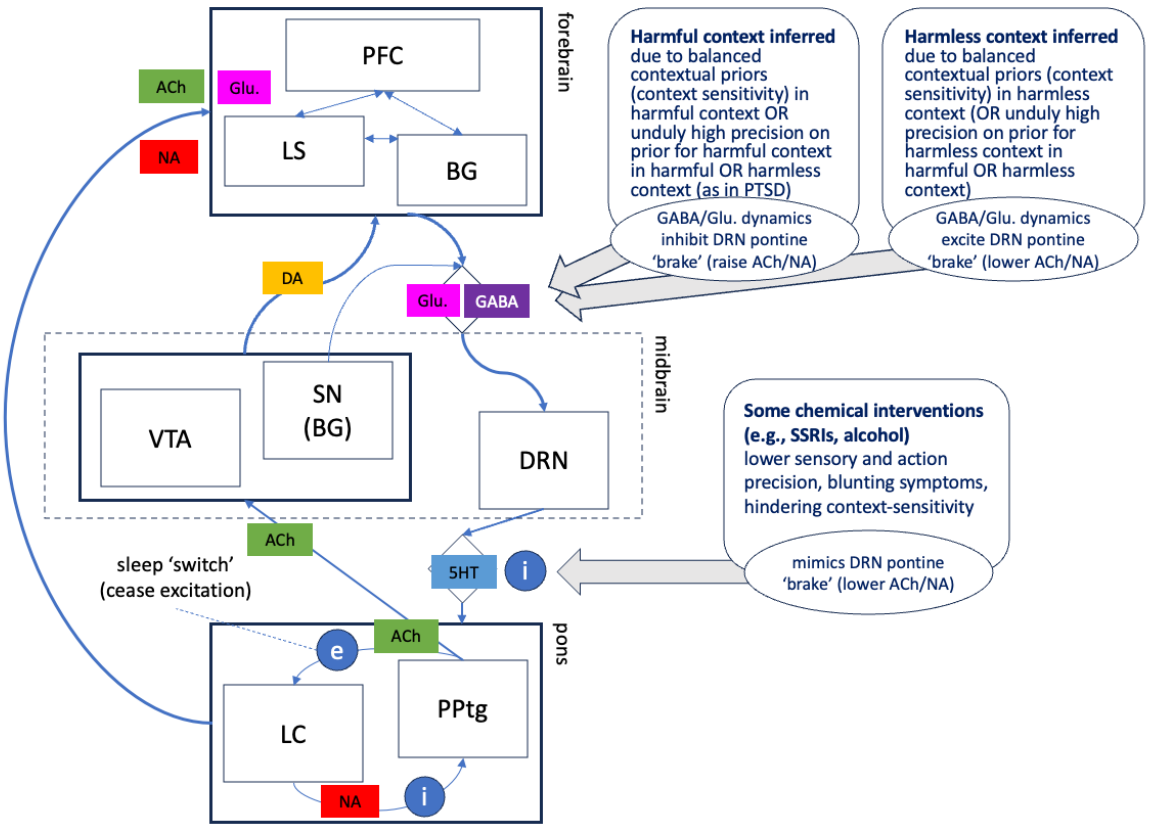


FIGURE 3. Neuromodulatory circuit. The neuromodulators noradrenaline (NA) and acetylcholine (Ach) are emitted in the pons. Cholinergic neuronal pathways from the pedunculopontine tegmental nucleus (PPtg) produce ACh and also have an excitatory connection to the locus coeruleus (LC). LC excitation produces NA that in turn has an inhibitory effect on the PPtg. This reciprocal relationship in the pons stabilises the ratio of NA and ACh perfusing the forebrain while awake. When asleep, the PPtg ceases excitation of the LC, allowing NA levels to drop and ACh to rise ('sleep 'switch'').

With a stabilised ratio of NA to ACh, their baseline levels rise and fall jointly. The joint levels are

modulated by the serotonergic projection from the midbrain dorsal raphe nucleus (DRN) to the pons. Increased serotonin (5HT) inhibits the pontine subsystem activity, thereby reducing its NA and ACh output. Conversely, decreased 5HT from the DRN disinhibits the pontine subsystem activity, allowing NA and ACh levels to rise.

The DRN is excited or inhibited by the interaction between glutamate (Glu.) and gamma-aminobutyric acid (GABA), via glutamatergic and GABAergic projections from a richly interconnected network situated primarily in the forebrain. This forebrain network (with some outer nodes) is comprised of the prefrontal cortex (PFC), limbic system (LS), and the basal ganglia (BG), and also includes the anterior cingulate cortex (ACC) (unlabelled). When a harmless context is inferred via the forebrain network, the DRN 'brake' is applied; its serotonergic projections increase the 5HT input to the pons, inhibiting NA and ACh levels. When a harmful context is inferred, the DRN is inhibited (the 'brake' is released). In turn, the decreased 5HT input disinhibits the pontine subsystem, leading NA and ACh levels to rise.

PTSD pathogenesis can be understood in relation to the circuit diagram as follows. A current or recalled traumatic experience releases the DRN 'brake' to raise ACh/NA for enhanced harm-mitigation (defensive response). A current harmful context will be strongly encoded with the potential to be easily cued. Later on, however, a resilient individual may experience cued traumatic recall, but with insufficient strength to resolve uncertainty about the present. This leads to peripheral sensory sampling, belief-updating about the veridical present, and subsequent applying of the DRN 'brake' to lower ACh/NA (see Fig. 2).

In a vulnerable individual, the overly strong recall of a harmful context may cascade through elevated ACh/NA, stimulated forebrain Glu. perfusion, and increased top-down excitement of VWM. Blending present visual sensation and recalled imagery in VWM (as described in Online Methods) can lead to false contextual inference with rapidly evinced behaviours that an agent (falsely) believes are contextually apt. Thus, instead of an exploratory policy to resolve uncertainty about context, an exploitation policy is selected; namely, peripheral sensory attenuation and vigorous defence (e.g., fight or flight). This is consistent with a role for ACh enhancing attention to significant cues,[32,33] amounting to hypervigilance (readily cuing harmful context), while NA simultaneously enhances motor responsivity,[34] amounting to hyperarousal (see Fig. 1).

Results

Neurocomputational simulations

We developed a neurocomputational model of interacting systems perspectives schematised here (see above and Online Methods): the *neurocognition of trauma associations* including learning, memory, and context; the *phenomenology and symptoms of PTSD* and their relation to treatment intervention; and the *neurobehavioural dynamics of* and a *neuromodulatory circuit for* defensive mobilisation. We show that our neurocomputational implementation explains (i) natural waning of PTSS in resilient individuals under conditions of ordinary stress, (ii) persistence of such symptoms in predisposed individuals in similar conditions, (iii) recovery of context disambiguation through a learning-based intervention, and (iv) the limited role of symptom reduction through monoaminergic therapy on its own.

We simulated PTSD pathogenesis and treatment intervention based on a generative model that refines, extends, and operationalises previous work.[1,11,12] The simulation studies use a generative model (Fig. 4) to reproduce the pathogenesis of PTSD following a traumatic encounter and a successful and unsuccessful intervention. The learning and behavioural dynamics of the system are driven by uncertainty resolution (implemented as free energy minimisation).

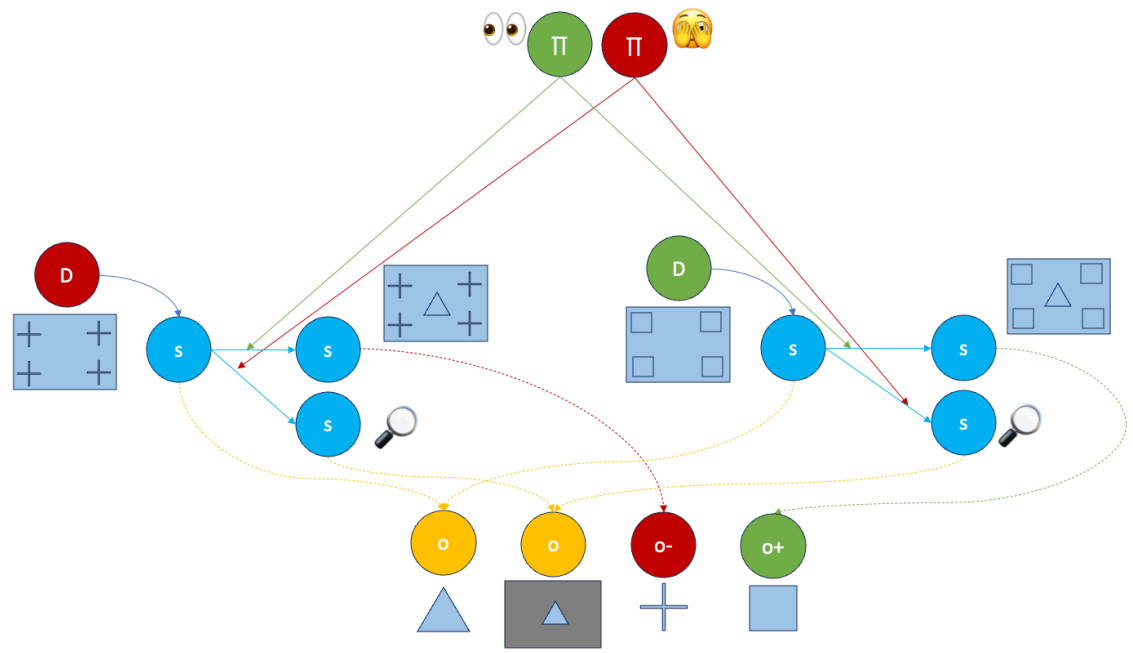


FIGURE 4. See main text for description.

In Fig. 4, the contextual priors (labelled **D**) condition the initial state, which generates a contextually ambiguous observation (triangle). During the initial traumatic experience, the system advances along the horizontal arrow from the initial state to the second state. The state sequence is controlled by an exploratory scene sampling policy ('two eyes' pictogram) to resolve uncertainty about the context. This second state generates adverse (surprising) observations that can include a range of 'harmful' sensations, among them pain (e.g., being struck by shrapnel) and visual scenes that inflict moral injury (e.g., bearing witness to extreme suffering). The observation allows for a veridical contextual inference, updating the prior probability (red **D**) that one is in the harmful context.

Following the harmful traumatic experience, in a new harmless context, one begins in the initial state conditioned by the green **D**. This start state also generates a contextually ambiguous observation (triangle). In other words, the initial observation (triangle) is insufficient to disambiguate between the two alternative start states. The exploratory policy would progress the state sequence along the horizontal line to generate context-specific observations (squares). This disambiguates the two contexts, leading to a veridical inference of the harmless context.

However, if one begins in the 'squares' initial state (conditioned by the green **D**), and the contextual prior of the past (traumatic) context is unduly precise, it is sufficient to resolve the uncertainty about the context, such that the exploratory policy is not selected (it is falsely believed to be uninformative due to overly informative prior beliefs). This results in false contextual inference and selection of the (peripheral) sensory attenuation policy ('partly covered eyes' pictogram). We emphasise that this form of sensory attenuation is consistent with hypervigilance, amounting to rapid shifting of foreground focus without resolving contextual uncertainty.

From a first-person perspective, there is enough information that one is in the harmful context, even though from a third-person perspective, one is in the harmless context. Moreover, under the sensory attenuation policy, the foreground focus state ('magnifying glass' pictogram) generates the same observation irrespective of context. Like the initial observation, the focal observation is insufficient to disambiguate context, so the agent never learns they are in the

harmless context. As long as the sensory attenuation policy is selected, no amount of observations will render prior beliefs obsolete, as in typical Bayesian learning. In past work, we proposed that in vulnerable brains, Bayesian learning 'breaks' because of an impairment in the ability to accumulate evidence that a context is benign (counterfactual blocking). Our simulation studies suggest that the specific weights of learning and forgetting, through which the memories of encounters with harmful and harmless contexts accumulate, interact with learning history to stabilise the inappropriate beliefs (see below and Fig. 6).

A second aspect of the generative model (Fig. 5) does not drive the simulation dynamics but is included to illustrate a secondary process important for re-experiencing symptoms. It emerges from policy selection under the core model definition (Fig. 4). In this secondary process, high precision on a contextual prior can lead to a blend of sensation and imagery in VWM indicative of flashback phenomenology. A trauma-associated cue (triangle) can elicit multiple contexts (harmful and harmless) without sufficient information to disambiguate them. Under relatively balanced priors, the exploratory policy is selected (see Fig. 8 boxes 4A and A). High sensory precision can be allocated to the periphery, even when there is sparse contextual information (see Fig. 8 box 3). This allows for 'bottom up' sensory evidence accumulation about the present harmless context. Under highly imbalanced priors, with undue precision on the harmful context, the sensory attenuation policy is selected. This places low sensory precision on the periphery, allowing for VWM to be partly excited by 'top down' recalled imagery, i.e., from the past harmful context (see Fig. 8 boxes 4B and $B_{1\text{-}2}$). In other words, a focal visual sensation would be congruent with focal mental imagery; this congruency combined with peripheral sensory attenuation and strong contextual imagery could meet a subjective 'reality threshold', amenable to experimental study in a reality testing paradigm.[35]

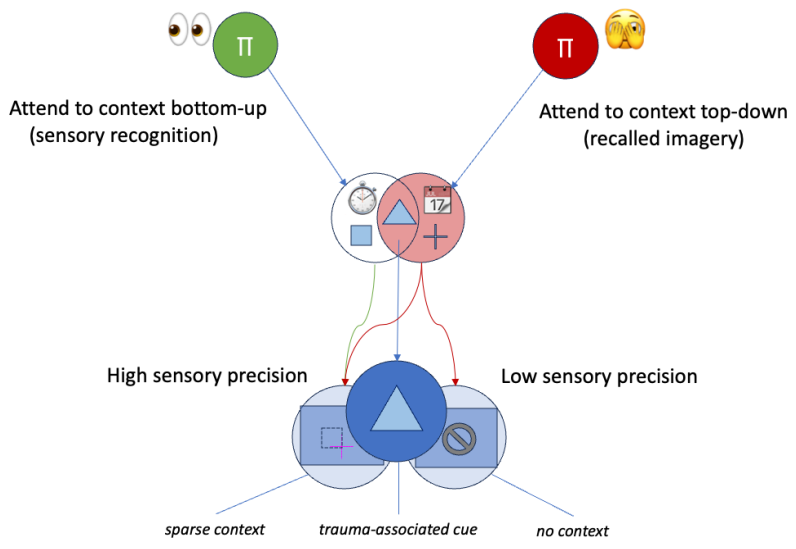


FIGURE 5. See main text for description.

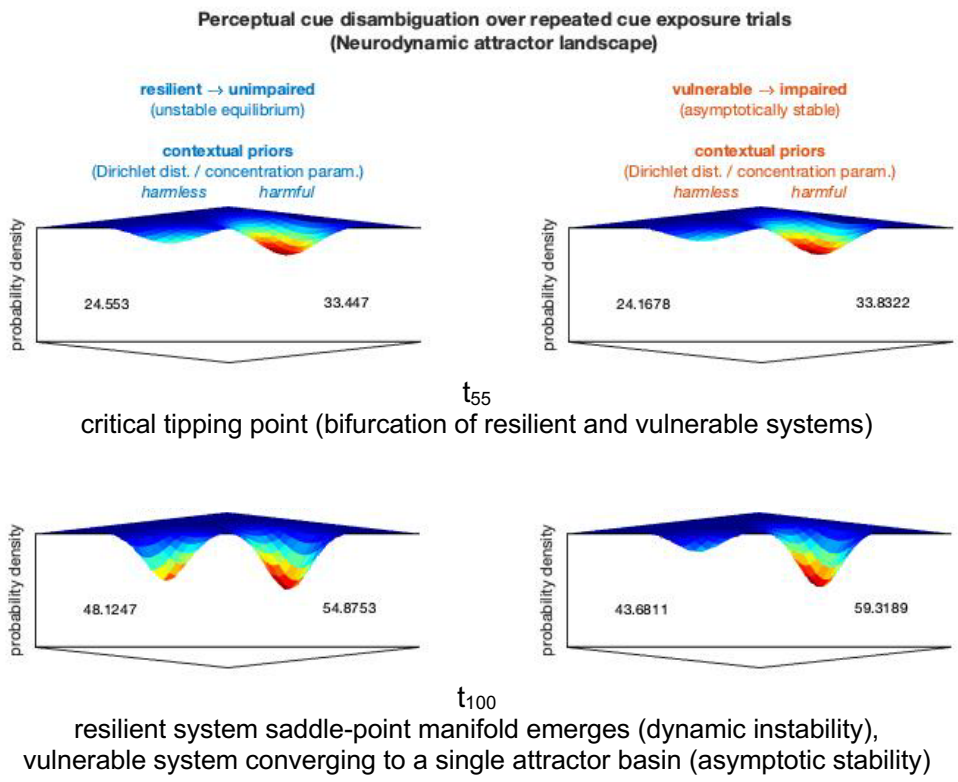
In Fig. 6, we see the evolution of contextual priors in a resilient (left) and vulnerable (right) system. Each system starts with two attractor basins that represent the initial priors **D**, with harmless (left) and harmful (right). Following a short sequence of harmful trials (that represent a traumatic experience) – followed by an extended sequence of randomly allocated harmful and harmless trials (50/50) – we see how the resilient system is able to garner veridical contextual information and update its priors. This leads to balanced priors that bias the system towards the exploratory policy for learning the true state of the world. In the vulnerable system, we see what amounts to a 'vicious circle' that characterises PTSD and related stress disorders: the sensory attenuation policy prevents learning about the harmless context, which

biases the system towards the sensory attenuation policy, and so on, in a self-reinforcing loop.

The resulting landscapes at the conclusion of both the resilient and vulnerable processes are consistent with recent neural recordings of macaque perceptual decision-making.[36] In the macaque study, neurons accumulate evidence for two alternative visual fixation target options preceding a saccade-to-target action. When there is ambiguity between the targets, decision time is longer, as sensory evidence accumulates to resolve the ambiguity. When there is bias towards one option, the decision time to reach a commitment (action) is faster, and the bias functions as greater 'pre-accumulated', or notional, evidence.

Our case is not about alternative perceptual targets, but the interpretation is similar.[37] Following a cue (triangle) common to two contexts (squares and crosses), if there is insufficient evidence to resolve which context is present, then more evidence accumulation is required to resolve the ambiguity (higher decision time to commit). In our model, this is achieved by an additional timestep of peripheral sensory sampling, to accumulate contextual evidence for either squares or crosses (see Fig. 8, boxes 4A and A). However, if there is a strong bias for one context (crosses), the contextual inference for crosses is made immediately, even if this amounts to committing to the incorrect decision when the true context is otherwise, i.e., false inference (see Fig. 8, boxes 4B and $B_{1\text{-}2}$).

In Fig. 6, the basin shape represents the contextual priors. The decision trajectory for resolving ambiguity must either continue to sensory sampling given the decision manifold under insufficient prior evidence (unstable equilibrium) or can forgo further sensory evidence due to a collapse into a single stable basin. The basins are illustrative data visualisations based on plots of the Dirichlet distribution **D**; they can be interpreted formally as the (free) energy landscape of a combined integration and decision-making network.[38]



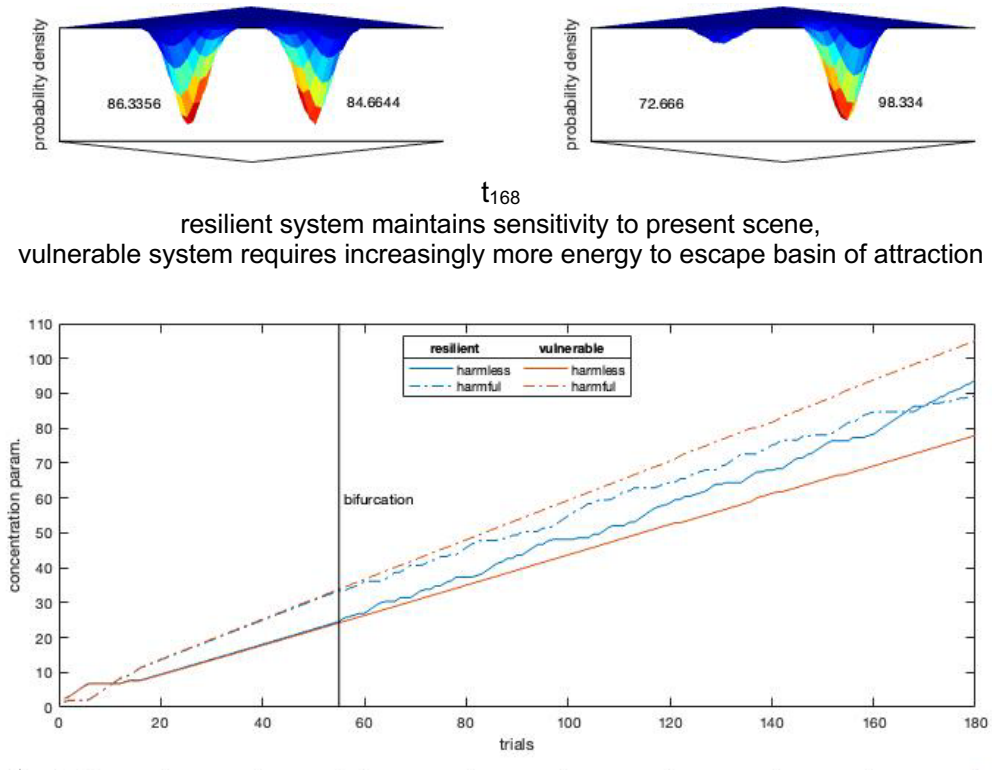


FIGURE 6. See main text for description.

In Fig. 7, we show that a successful intervention is one that restores context-sensitivity to behaviour, which facilitates the rebalancing of prior probabilities. This success may be facilitated with or without psychopharmacological support. For example, SSRIs may be used therapeutically to improve the potential for new learning.

However, psychotropic drugs (including SSRIs, benzodiazepines, or excessive alcohol consumption) can also show illusory success when they blunt context-sensitivity in ways that reduce PTSS, without new learning. Interventions of this sort are modelled via changing the precision of the likelihood mappings (representing sensory precision) and transition priors (representing prior precision). Numerically, these interventions were simulated by lowering precision in the corresponding **A** and **B** tensors (under a generative model based upon a partially observed Markov decision process).

The disappearance of symptoms in the absence of context-sensitivity is insufficient to rebalance priors, leading symptomatic behaviour to return at equal or greater strength upon cessation of chemical intake. Conversely, cessation of (e.g.) a psychotherapeutic treatment that has restored context-sensitivity even before priors are fully rebalanced has the potential to lead eventually to rebalanced priors. This contrast between successful and unsuccessful PTSD treatment closely matches randomised controlled trial (RCT) results.[15]

Each diagram in **Figure 7a-e** can be interpreted as follows:

- The numbers along the bottom axis indicate the trial number in blocks of 60, comprising two alternative sequences of 180 trials (3 blocks of 60 each). The sequences are clarified in the captions beneath each block. (Summary: 7a depicts the initial 60 trials, with two alternative continuations of 2x 60-trial blocks, either to 7b-7c or 7d-7e. In other words, the 60 trials of 7a are in common to both sequences, and 7b and 7d begin with an alternative trial 61 of 180; 7c and 7e begin with an alternative trial 121 of 180.)
- The top row indicates whether it is a harmless (□) or harmful (+) trial.

- The four ‘internal’ rows (with greyscale boxes) indicate the policy selection for each trial. A trial is comprised of three time-steps: a start state, and two policy selections that advance to a middle and final state, respectively. (Fig. 4 depicts the pathways from the start state to the middle state, and the final time step allows the agent to remain in the middle state for the final state, or move to the alternative non-start state.)
  The four available policies are all the permutations and combinations of the two actions: 👀=exploratory sensing (of the periphery) and 😘=sensory attenuation (of the periphery). In the rows below, from top-to-bottom, the policy for two timesteps is:
  👀-👀(row 1), 😘-😘(row 2), 👀-😘(row 3), 😘-👀(row 4). The grayscale within each trial column indicates the progressive resolution of uncertainty by belief-updating (within the policy space) until policy selection.

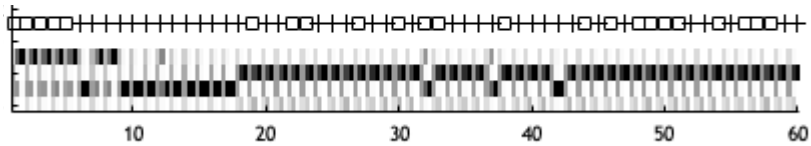


Figure 7a. Initial sequence of trials. Trial 1-5: harmless condition (□), 6-16: harmful condition (+). 17-60: random. Without intervention, the sequence continues in the manner of trials 50-60 (i.e., repeated selection of sensory attenuation policy under any condition). The evolution of contextual priors for this continuation without intervention (from trials 1-60 above and 61-180) is shown in the vulnerable system in Fig. 6.

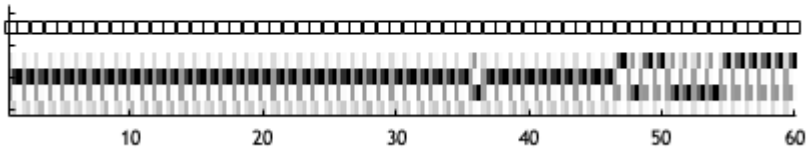


Figure 7b. Simulated treatment option to restore context-sensitivity. Controlled exposure to harmless condition for 60 trials. Towards the end of the sequence, the exploratory policy begins to emerge and the contextual priors begin rebalancing. This can be understood in terms of Fig. 6, as a shift from a vulnerable individual (that continues along the vulnerable trajectory in the simulation with no intervention) to what resembles the resilient individual (who never develops the imbalanced priors in the simulation).

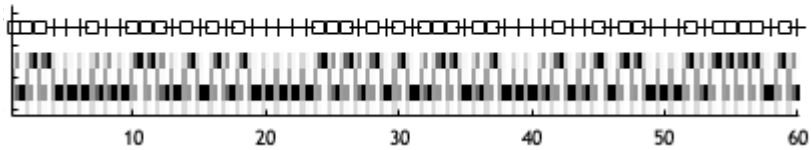


Figure 7c. Following the treatment intervention simulated in Fig. 7b, the remaining trials continue at random (harmless or harmful), maintaining balanced priors and appropriate context-sensitive policy selection.

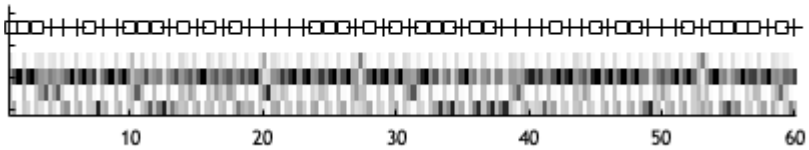


Figure 7d. This sequence is an alternative progression following the 60 trials of Fig. 7a. It approximates the effects of (e.g.) SSRIs (selective serotonin reuptake inhibitors) or alcohol in lowering the precision of sensory information and action. At the same time, the trials are random to illustrate the absence of a controlled environment.[39] The kind of response vigour seen in Fig. 7a trials 50-60 is diminished here, indicated by more grayscale (low precision policy selection) and a wider range of selected policies, which can be interpreted as a reduction in PTSS.

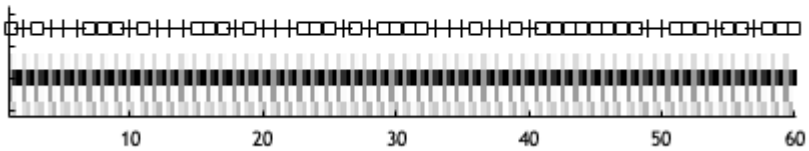


Figure 7e. This sequence continues from Fig. 7d with the same random alternations of conditions, but with a restoration of the initial sensory and action precision from Fig. 7a, representing the cessation of chemical intake (e.g., SSRIs or alcohol). In the Fig. 7d trials, the contextual priors were not rebalanced, so here, the context-insensitivity governs the policy selection. As described in the main paper, an exploratory policy is falsely ruled out as unnecessary, given what is falsely taken as sufficient information about context from the prior. This means that even in the harmless trials (□), the sensory attenuation policy is selected. In line with the full analysis in the main paper, this amounts to a self-reinforcing neuromodulatory 'vicious circle' of defensive policy selection and counterfactual blocking.

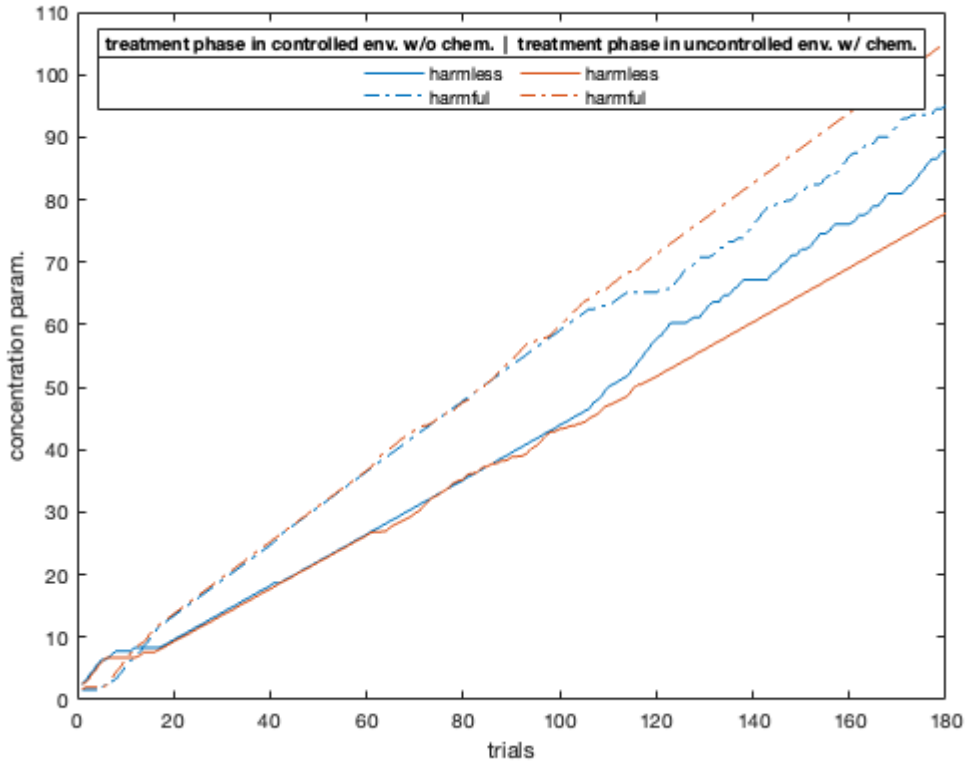


Figure 7f. Evolution of contextual priors (vulnerable), with alternative interventions in trials 61-120. Gradually converging plot lines (blue) depict the trial sequence of Fig. 7a-**b-c**, with a simulated intermediary phase of no chemical intervention under a controlled environment (e.g., psychotherapeutic treatment). Diverging plot lines (orange) depict the trial sequence of Fig. 7a-**d-e**, with a simulated intermediary phase of chemical intervention (e.g., SSRIs or alcohol) under an uncontrolled environment (simulated by lowering precision in the **A** and **B** tensors).

Discussion

The analysis and modelling presented here serves to formalise concepts that often remain underspecified in the narrative literature on PTSD and related psychopathology. In this work, context refers to elements of the sensory or cognitive periphery that provide evidence for the states of affairs prior to belief-updating. Belief-updating here refers to inferring – and learning about – the causes of sensations through neuronal message passing and experience-dependent plasticity, respectively. Our simulations demonstrate the interdependence between

inference and learning, and how biased inference can lead to aberrant learning and a failure to recognise changes in context. It is this failure we associate with the psychopathology of PTSD. Biased inference in this instance manifests as a failure of responding to epistemic affordances that leads to a failure to recognise the context has changed.

This kind of failure sits on a critical tipping point, leading to a phenomenal description of the ensuing inferences in terms of bistability, that can be related to the notion of canalisation in psychopathology.[40] While attractor basins are used in different ways to indicate neuronal dynamics, here, we have shown that plasticity can create unstable basins that bias the system toward sensory evidence accumulation to resolve ambiguity about the causes of sensations. In contrast, a collapse or canalisation into a single basin can cause a trajectory initiated by a sensory cue to become inescapably bound by the basin; resolving uncertainty through precise – but false – beliefs about the causes of sensations. Crucially, this resolution is self-maintaining because it precludes epistemic foraging or information seeking; thereby instantiating a vicious circle of confirmation bias. Restoring plasticity can make the system amenable to new learning, thus rebalancing priors and restoring the collapsed basin. The saddle point between basins corresponds to a reinstatement of context-sensitivity.

In a context where harm is highly likely, such as a combat zone, it is adaptive to have elevated ACh and NA, to remain vigilant against possible harm, and rapidly mobilize a defensive response for mitigating harm. Essentially, it is context-sensitive to be in a heightened alert state, 'better safe than sorry'. After a transition into a context where harm is unlikely, it does not mean harm is impossible. A home neighbourhood may be largely harmless, but occasionally harms can arise. Thus, context-sensitivity requires sensory disambiguation on a case-by-case basis, as a tap on the shoulder could be from a friend or foe. It would be maladaptive to immediately respond to a shoulder tap with a vigorous fight/flight response, as this would be unwarranted if it were a friend. Yet, it would also be maladaptive to be 'fearless' and disregard learning whether the tap is indeed from a foe, which warrants self-protection.

We have provided modelling-informed analysis of PTSD pathogenesis and related sequelae of traumatic stress and their relationship to diagnostic criteria in terms of vulnerability and resilience. Our key hypothesis is that the balance of plasticity-enhancement and plasticity-decay for aversive events is greater in the vulnerable. We have also provided several novel auxiliary hypotheses that can be tested in controlled experiments (see above and Online Methods).

Our analysis also provides a way to understand learning-based treatment mechanisms and their relation to (e.g.) monoaminergic interventions and substance abuse. Pharmacological interventions could enhance psychotherapies which support learning contextual disambiguation, through allocating attention to evidence for benign causes of current, including aversive events. This is a putative mechanism of action of entactogen-assisted psychotherapy. The combination of psychotherapy and psychopharmacology may be indispensable in some cases (e.g., due to TBI-related structural damage). At the same time, our account is consistent with psychotherapy providing sufficient treatment for trauma in many cases. ELEs fed back that understanding the interconnected aspects of our analysis can help mitigate self-stigma and shame (see Online Methods for ELE feedback), though establishing detailed mechanisms whereby social and therapy factors can help requires further work.

For onset prevention, the model suggests that arbitrary aspects of the environment during the post-traumatic period could hinder or induce impairment, depending on how strongly and/or frequently sensory cues mobilise a defensive response. For example, immediate debriefing following a traumatic experience could unduly remobilise a defensive response and strengthen memory of the adverse context. Social factors (e.g., support groups, transition homes) would have the potential to counterbalance adverse cued responses by strengthening the inference

of a present harmless context. Overall, our analysis lends support to biopsychosocial approaches to diagnosing and treating PTSD, or preventing its onset.

Online Methods

The foregoing account is interlinked with cognitive psychology concepts and supporting literature here. We use example scenarios for thematic organisation. Subsections progress through contextual mechanisms and processes, visual working memory (VWM), long-term memory (LTM) and generalisation, traumatic memory, and sleep.

Recognition dynamics involving foreground and background elements

Consider the example of being in a room with red wallpaper and a table with a deck of playing cards on it. This example can be linked to disparate literature using several terminological variations; namely, the room and wallpaper properties can be understood variously as context, background, scene, sensory periphery and/or global visual field. When interacting with the deck of cards, the deck can be considered a foreground or salient element, an object of attentional or sensory focus and/or part of the local visual field. In what follows, we treat the terms within each set as synonymous, acknowledging that some research operationalises distinctions internal to each set.

In an ordinary setting, relative to an agent's attentional focus and interaction, there are often focal elements that are salient (i.e., foreground) and surrounding scene components that can be categorized as 'context' (i.e., background). In any such event, there also may be experiential and affective relationships between the agent and salient elements (e.g., the cards were a gift from a friend). How agent-environment interactions during an event unfold in perception will impinge on how they are encoded in memory, including salient foreground, contextual background, affective, and goal-direct aspects (e.g., implicit or explicit reward). In turn, how they are recalled from memory will also differ according to these encoded aspects and the recall circumstances.

Experimental studies often manipulate foreground and background elements as stimulus and context, to investigate neuropsychological processes – for example, by displaying sequences of images to isolate a phenomenon of interest. One finding is that recognition of a paired object and scene mutually reinforce each other.[41] A pairing can be learned by repeated exposure to examples ('training'), such as a red room with a deck of cards and a blue room with a large matchbox. After training, seeing a deck of cards will lead to fast recognition of the red room (inference from object to context), and vice versa, from room colour to table contents (inference from context to object).

It is also the case that scenes can be recognised based on global texture. For example, the red room may be full of round tables and chairs, while the blue room has rows of theatre seats. These scenes not only have dimensional features (red vs. blue) with recognisable objects (table seating vs. rows), but also a global textural stimulus array that combines into a gestalt and facilitates rapid recognition.[42] Global context also can be multimodal, such that different ambient sounds in each room (auditory background), once learned, facilitate faster visual recognition.[43,44]

All of the above forms of recognition (stimulus-context associations) can be recast as disambiguation in probabilistic inference. On-going predictions govern visual scanning (sensory sampling) to accumulate evidence that lowers the uncertainty about a given hypothesis. This describes the unimpaired 'machinery' that gives rise to PTSD when deployed suboptimally.

Continuing with the unimpaired scenario, when there is an element in common across multiple contexts, such as a curtain hanging in both the red and blue rooms, visually focussing on the curtain provides an ambiguous indication of what room I am in. If I hypothesise that I am in the blue room, I predict I will see rows of seats when I scan the room. Seeing round tables violates my prediction and resolves my uncertainty. I now infer I am in the red room. (The caption to Fig. 8 illustrates this example.)

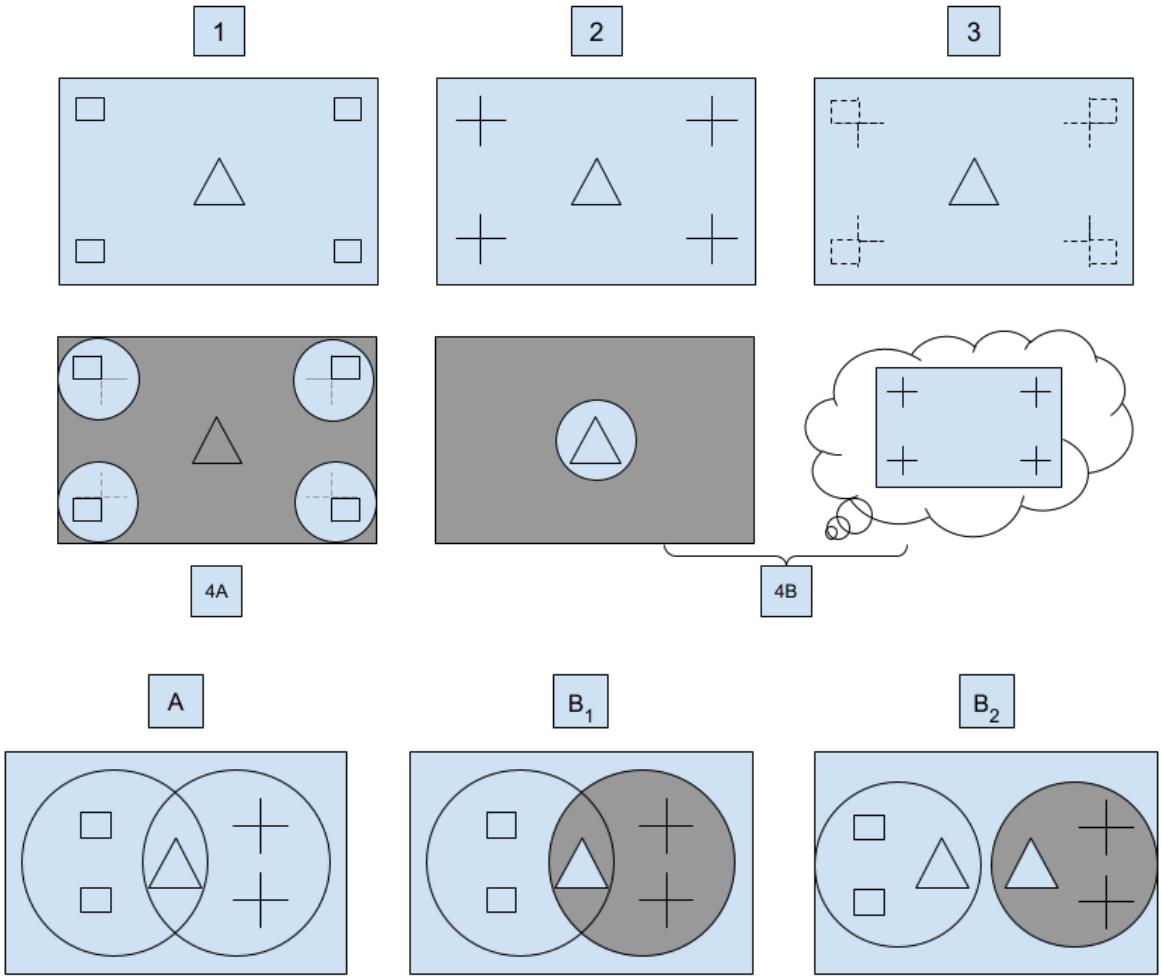


FIGURE 8. Two panels with a common element (△) and different backgrounds (□ vs. +) are labelled 1 and 2, respectively. After sufficient training to gain familiarity with panels 1 and 2, seeing the triangle in panel 3 suggests two alternative possibilities: it ambiguously matches panel 1 or 2. For this figure, it is stipulated that the veridical match is between panels 1 and 3. Even if panel 3 is predicted to match panel 2, attention can be directed to the periphery for sensory sampling (illustrated in panel 4A), to accumulate evidence about the veridical context. Once sensory evidence at the periphery is accumulated, it resolves the ambiguity of the background scene, leading to the inference that panel 3 context matches panel 1. A combination of sensory attenuation at the periphery with sensory focus on foreground, and high precision on recall of panel 2 – illustrated in panel 4B – can inform context disambiguation in favour of a false inference that panel 3 matches panel 2. See further below for a description of boxes A and $B_{1\text{-}2}$.

Multimodal (polysensory) behavioural interaction with elements of the environment is also relevant to stimulus-context association. Seeing an approaching object that looks as if it will soon hit me may coactivate the tactile sensation of being hit.[3,45] Thus, visual recognition is not limited to a purely abstract mental activity, but can be continuous with overt behaviours embedded in a context. For example, in a park, a defensive posture may be apt to protect oneself from being harmed by an apparently approaching ball initially seen at a distance.[4,46] Yet an equivalent visual sensation may cause an inapt defensive 'flinch' response during an immersive sports videogame when there will be no tactile impact.

This scenario also describes the unimpaired 'machinery' that gives rise to PTSD when it is impaired. A defensive reaction to current conditions can be understood as a consequence of prediction under a probabilistic (generative) model, as opposed to a stimulus-response account.[47] Under the active inference formulation, one is effectively working backwards from a future time point of feeling tactile pain to the current time point of an approaching projectile. Inferences are made about the underlying (hidden) states that generate the sequence of sensations (e.g., the projectile approach will appear as a growing ball and lead to damaging impact causing pain). On this basis, an action is selected that has the best chance of preventing a surprising outcome, i.e., being hit by a ball.

The context of the park vs. the videogame is relevant to the learned state sequence and resulting sensory outcome. When playing the videogame, what appears to be an incoming ball is a foreground cue that can lead to a false inference about the background context. If I (falsely) infer I am in the park, I expect the ball will hit me and cause a painful outcome. Thus, from a first-person perspective, a defensive flinch is apt (Bayes optimal), even though it is inapt to an observer who (correctly) infers I am playing a videogame.

Technically, this can be understood in terms of the complete class theorem;[48,49] namely, for every pair of behaviours and loss functions there are some prior beliefs that render my behaviour Bayes optimal. Our focus in this work is on prior beliefs about the context in which I am behaving. Next, we connect false inference about context to post-traumatic sensory attenuation.

Between 'bottom-up' and 'top-down': Visual working memory (VWM)

In a cluttered scene – akin to a naturalistic environment – individual object recognition may be diminished for most objects, especially when scanning a scene with unfocused attention. However, preferred (i.e., familiar or unsurprising) objects may remain undiminished despite clutter,[50] which can explain a 'pop-out' effect when scanning a crowd and spotting a familiar face. This preferential recognition occurs in the ventral stream, which has been found to be impaired in PTSD, resulting in a form of sensory attenuation.[51] A puzzle about sensory attenuation in PTSD is how it might relate to hypervigilance, a form of sensory hyperactivity. This puzzle can be solved from the account we elucidate here.

Psychology and neuroimaging studies operationalise various load concepts (attentional, perceptual, cognitive, working memory) to focus on specific differences (e.g., visual pathways vs. attention networks). Yet the common ground among them is a resource model that can be mapped to neuronal populations with biophysical limits. For example, all neurons within a population can be tuned to a single stimulus, or a subset is tuned to (e.g.) another stimulus.

From a technical perspective, this resource allocation is an emergent property of active inference in the following sense: the objective of active inference is to maximise the evidence for generative models of the lived world. Evidence (a.k.a., marginal likelihood) can be decomposed into accuracy and complexity. Therefore, model selection necessarily entails a minimisation of complexity, which can be read as minimising the degrees of freedom used to provide an accurate account of the sensorium.[52–54] This minimisation can be read as complying with Occam's principle and manifests as an apparent constraint on the allocation of representational resources (e.g., the quantisation of representations, implicit in receptive fields). Our current modelling does not aim to disentangle different aspects of resource allocation within active inference. For our purposes, it suffices to say that mitigating harm in a traumatic encounter commonly requires visuospatial attention and sensorimotor coordination; namely, efficient active vision.[55–58]

Visual working memory (VWM) is a useful construct for centring on load-limited interactions

related to visuospatial sensorimotor task demands. We treat VWM as a limited neuronal population susceptible to 'bottom-up' (sensory driven) and 'top-down' (recalled imagery driven) excitation.[59–61] Given a sensory contribution, some findings on visual pathways unconcerned with VWM indirectly bear on our model. For instance, there is empirical support for reduced processing of peripheral visual stimuli under higher task demands, i.e., higher cognitive load; this includes the direct interaction between optical information and visual cortices.[62–64] In our account, the reduced influence of sensory evidence on VWM in some PTSS phenomena corresponds to an increased contribution of recalled contextual imagery. (We model this as an explore/exploit trade-off mediated by the precision afforded new observations and prior expectations.)

In PTSD, increased bottom-up gain has been found to co-exist with increased top-down gain.[65] Two connectivity pathways, from vision (bottom-up) and mental imagery (top-down)[59,60] can be understood as converging in VWM.[61] Even if a visual element in VWM is absent from a 'bottom-up' sampling of the scene, VWM can robustly maintain visual information 'top-down', as if the item were present.[66] Moreover, the top-down influence of mental imagery on VWM can affect and distort concurrent bottom-up visual perception.[67]

In line with the scenario of a looming projectile above, VWM is not necessarily detached from motor behaviour. Mental imagery of a scene can even influence oculomotor movement, inducing saccades to fixation points that match saccades in perception of the same scene.[68,69] One can also pursue similar arguments to explain rapid eye movements during dreaming in REM sleep.[70] This coupling could explain why altered eye moments (as in EMDR therapy) would disrupt the vividness of mental imagery, and in turn reduce the magnitude of affective response to mental imagery of a traumatic experience.[71]

A synthesis of the above evidence points to a possible explanation for 'flashback' phenomenology: high sensory precision or gain (modulated by ACh[22], discussed in the main text) can facilitate visual attention to scene elements with enhanced salience (trauma reminders or 'cues') that are rapidly identified and rapidly elicit a trauma-context association. The traumatic context is characterised as one with a high probability of causes of harm ('harmful context').[16] The on-going imperatives of harm-mitigation requires flexibly attending to 'task-relevant' stimuli; i.e., sampling the visual scene to maximise information gain (a.k.a., Bayesian surprise).[72–75] This aspect of active vision follows the metaphor of an attentional 'spotlight' that must rapidly shift among momentary foreground elements. With each shift, the remaining background scene is relegated to a diminished 'task-irrelevant' periphery.[64] This restlessly roving spotlight would be indicative of hypervigilance, with increased sensory gain on a focal element and corresponding attenuation of the sensory periphery.

As we illustrate with our model simulations (see Results), under active inference, the sensory periphery is uninformative about context (from a traumatised first-person standpoint) because of high precision on the prior probability of the harmful context. In terms of belief-updating, given the (false) belief one is in a harmful context that matches traumatic recall, a strong top-down influence of trauma scene imagery on VWM – concurrent with foreground perception – can cause a distortive blend of the present scene and a past traumatic event.

These imagery-based elements would continuously support misperception and misrecognition of the current scene, allowing attention to be maintained on a significant foreground element while hindering attention to – and subsequent belief updating about – context (Fig. 8, box 4B). In turn, a behavioural response is elicited that would be appropriate to a repeat traumatic encounter, but is inappropriate to the veridical conditions. This formulation of sensory attenuation theory thus provides a testable hypothesis for the neurocognitive machinery of 'flashback' phenomenology, in that cued traumatic recall is predicted to attenuate peripheral sensing more than focal sensing (e.g., in a target/distractor paradigm). The theory also provides a testable hypothesis related to another common PTSD symptom, the exaggerated

startle response (ESR); namely, that alternatively valenced contexts for the same probe stimulus should differentially affect startle parameters for PTSD-impaired individuals to a lesser degree than those without.[76]

Between past and present: From memory to generalisation

An encounter with a hostile person or adverse place would be important to remember for harm mitigation in a subsequent encounter. On one hand, the widely accepted psychological account of this experiential process moves in a bottom-up fashion from perception to VWM to short-term and/or long-term memory (STM and LTM), and in top-down fashion from memory to recall and/or recognition. On the other hand, widely accepted hierarchical predictive processing accounts cast this overall process as involving a vast number of feedback and feedforward re-entrant loops throughout the brain. This speaks to a strong correspondence between the psychological constructs and interacting yet discrete neuronal populations involved in active sensing and learning.[29,61,77] Thus, our neurocomputational (active inference) model is compatible with both psychological and neurophysiological modes of explanation.[47,78] From this perspective, we draw on typically disconnected findings on PTSD including cognitive psychology and biological psychiatry, and related work on unimpaired mechanisms and processes.

In light of the above literature, it is clear that we can learn to associate foreground elements with background scenes, and foreground and background can independently activate one another. Such co-activation also applies to elements of episodic memories. Events involving multiple observed elements (e.g., specific visual landmarks, faces, ambient sounds, etc.) are experienced and encoded together. If an element overlaps between events (a sunset viewed on two separate holiday destinations), a memory of both events may be partly activated when seeing another sunset. That is, similar neuronal representations may be partially coactivated (e.g., 50-50).

An additional sensory stimulus associated with one memory but not the other (a specific scent) can in turn facilitate a (non-pathological) 'flashback' of the distinctive event memory.[79,80] This could describe the famous Proustian phenomenological effect of having a madeleine with tea and being mentally transported to a vivid childhood scene as if it were present again. The 'winner takes all' convergence on a decision amounts to an inference: inhibiting ruled out representations and disinhibiting a selected one.[81]

Cognitive psychological and neurobehavioural experiments are often designed to mimic a naturalistic progression that involves event-based experience, memory, generalisation, and recall. These can amount to learning that a task (e.g., pulling a lever) has a desirable (i.e., familiar or preferred) result in one context and an undesirable (ambiguous or surprising) result in another. Learning about the task and the context can allow a contextual cue to elicit the learned context-relevant interactions (e.g., approach vs. avoid). Human and nonhuman primate studies and computational simulations indicate that such task and context information is differentially encoded in memory.[82–84]

Hippocampal dynamics are implicated in both generalised task guidance and episodic recall.[29,80,85] This interaction between generalised task-oriented behaviour and context-specific episodic memories offers another example of unimpaired 'machinery' involved in PTSD impairment. For example, if you move heavy boxes, you may learn a generalisable task-relevant strategy for how to lift, hold, and carry. Months later, when moving furniture, an association between the two contexts can help guide your behaviour. At the same time, the association may also lead to recalling specific episodic moments ('flashes') from moving boxes.[86] The contextual similarity between both events allows for generalising the task behaviour, while the contextual break between them facilitates their ability to co-exist in distinct event memories.[85,87,88] Our modelling (see main text) shows how a harmful context can elicit

defensive behaviour and simultaneously promote introspective attention to event memory. (In the computational neuroanatomy of active inference, context is encoded in a **D** vector and associated with neuronal representations in the hippocampus.[89,90])

Impaired context disambiguation in post-traumatic pathology

Non-pathological 'flashes' can have a similar genesis as pathological 'flashbacks' (a common PTSD symptom). Consider what may occur if a quotidian but non-ubiquitous object is part of a traumatic event, say, a readily available jacket worn by an aggressor in a hostile encounter. There is a potential for a later encounter with someone else wearing the same type of jacket to elicit a traumatic memory.

Pursuing the above example, a *resilient* individual – who experienced the traumatic event – may form an initial trauma association with the jacket, which co-exists with other memories of seeing the same type of jacket on others. The association may be weakened over time, as the jacket cue is neuronally linked across distinct event boundaries. A repeat encounter with the cue is thus insufficient to disambiguate co-activated memories, so they do not become vivid or 'capture' attention. Attention is easily deployed to the sensory periphery to accumulate evidence about the present external surroundings. It is unlikely the present situation will be confused with the past event.

An illustration is given in Fig. 8, with a triangle standing in for the jacket. In Fig. 8 box A, the triangle is in common to both circles, depicting event memories as guiding contexts. Seeing the triangle evokes both events, with a square or cross context, respectively. Deploying attention to the sensory periphery (Fig. 8, box 4A) resolves the uncertainty about the present context.

Using the same example, a *vulnerable* individual may find the jacket cue capable of fully activating the trauma recollection, at the expense of any competing associations (Fig. 8, filled-in circle in box B_1; the contrast to B_2 is explained below). This full 'winner takes all' activation can predominate in VWM, increasing recall vividness and associated affective and motor responses, intermingled with present perception (i.e., flashback; Fig. 8, box 4B). The present salient object (triangle/jacket) interacts with the recalled traumatic context, mutually reinforcing the belief that the past event is present, in other words, a re-experiencing. This suggests a further testable hypothesis about post-traumatic impairment that differentiates between an active inference account, which predicts increased weight on trauma-encoded context, and dual representation theory, which predicts decreased weight.[91]

Under a sensory (periphery) attenuation policy, the vulnerable person does not learn about the veridical context. This reinforces a pathological persistent belief of being in the harmful context, given the cue (as illustrated in our model simulation in the main text). If re-experiencing is itself self-categorized as injurious, deliberate experiential avoidance will further impair learning. But even without deliberate experiential avoidance, the re-experiencing dynamics that limit peripheral awareness ultimately block inference and learning about benign causes of sensations (counterfactual blocking).[11] In contrast, if counterfactual evaluation of context is not blocked, it becomes possible to recognise the veridical context. Even minor indications that I am not in harmful context, if sustained (e.g., with adequate social support[92]), can facilitate a gradual updating and rebalancing of contextual priors that restores appropriate context-sensitivity. This rebalancing can be understood in terms of experience-dependent neuroplasticity.

In naturalistic conditions, the restored context-sensitivity would be adaptive as opposed to an 'extinguished fear' that could be maladaptive. For example, one may be momentarily susceptible to harm in an alleyway near home, even though the home context has a low likelihood of harm overall. A cue (jacket) that indicates a possible harmful encounter can be

confirmed by rapid exploratory sensing (friend or foe) leading to an equally rapid decision selecting or deselecting defensive behaviour.

Recent findings show that stressful initial recall (as could be expected in a post-traumatic period) can prevent the formation of new engrams of the event memory.[93] Without typical post-consolidation processes, subsequent trauma recall can reactivate the original experience-encoded memory trace, consistent with traumatic re-experiencing, in contrast to non-traumatic unpleasant memory recall.[94] Fig. 8, boxes B_1 and B_2, illustrate two ways in which the strength of an adverse event memory (dark grey circles) can furnish non-veridical context in a pathological flashback. Both can result in what is depicted in Fig. 8 box 4B at the expense of Fig. 8 box 4A, namely, an overly precise top-down influence on VWM concurrent with bottom-up peripheral sensory attenuation (of contextual information). At the same time, high sensory gain on the foreground (triangle/jacket) both cues the traumatic recall and maintains the task-guiding past-as-present dynamic. (Our computational model does not distinguish between Fig. 8 box B_1 and B_2, but future work could illustrate the relevant hippocampal differences.)

Traumatic memory and sleep

Our analysis and modelling provides fresh insight into the finding that dysregulated sleep following traumatic experience predicts the development of PTSD.[95] As described above, stressful recall results in the maintenance of a wakeful trauma experience memory trace that is reactivated on recall but not re-encoded.[93] There is another opportunity for re-encoding during sleep. Successful re-encoding facilitates a closed event boundary with selected associations neuronally linked across events (for instance, the jacket from the above example; see Fig. 8 box A and B_1).[88] However, if following a traumatic experience, subsequent sleep is disrupted, then the original traumatic memory trace may continue to be maintained (Fig. 8 box B_2).[24]

In a healthy scenario, memories are repeatedly re-activated ('retrieved') and re-encoded in a new set of synaptic connections that are activated when recalling the original event.[96] The healthy progression depends upon unstressed retrieval and sufficient NREM sleep, both of which may face interference following acute or extended traumatic experience.[93,95,97] For example, early life stress can lead to a memory trace (neuronal ensemble) being preferentially re-activated during related adult stress,[98] suggesting a contributing factor to CPTSD.

In the initial days following an event, some elements of the event-correlated neuronal ensemble overlap between its initial encoding and subsequent re-encoding. In a healthy progression, the re-encoded ensemble eliminates overlap with the originally encoded ensemble. This process can be understood formally as information compression or model reduction to minimise complexity (i.e., maximise model evidence in the absence of sensory input),[70] differentially driven by time and experiential learning factors.[99] These factors are captured indirectly (abstractly) by our modelling: the time factor can be understood as a decay rate which is potentially too low in a person vulnerable to post-traumatic syndromes. If there is insufficient decay over time, alternative synaptic weights can be strengthened by learning (e.g., in treatment intervention). In principle, this learning could reduce excessive re-activation of the original ensemble, which may subsequently support its re-encoding.

A single nightmare or even a period of nightmares (e.g., following a bereavement) is not necessarily pathological, but may contribute to a period of intrusive cued re-activations in a manner suggested by our model. While our framing and modelling deliberately excludes fear and extinction, related work analyses nightmares in terms of fear extinction linked to 'decontextualization' and 'recontextualization''.[100] These terms can be re-interpreted using our model as follows. If there is a cue in common to multiple memories, an overly precise contextual prior for 'harmful' may preferentially activate a traumatic memory that gives rise to a nightmare. Weakening the prior (synaptic strength) may reduce cued re-activation of that

memory, akin to decontextualization of the cue. Strengthening a competing contextual prior for 'harmless' may then allow preferential activation of non-traumatic memories, akin to recontextualization of the cue.

Expert by Lived Experience (ELE) feedback

Seven ELEs individually reviewed this work and provided feedback, summarized here. Each point contains quotations from a single ELE (distinct from order listed in acknowledgments), reproduced verbatim apart from minor grammatical adjustments. Quotations have been re-ordered and grouped under thematic subheadings. Emphasis added in bold to highlight key takeaways.

Ability of the analysis to capture lived experience phenomena

- Actually **made me feel understood** in a strange way. It's like—my brain is doing what it thinks it should, assigning too much weight to the wrong signals. That finally explains why I feel like I'm always overreacting, even when I rationally know I'm safe.
- **The science gave me language for things I've sensed for a long time but didn't know how to explain.** As someone with a background in military service and an interest in how the brain works, I really appreciated that this model goes beyond the old fear-conditioning stuff.
- This is **the first time I felt truly seen by a model**. When you described the brain clinging to old threat beliefs, **I felt like you were describing my daily life**. PTSD has always felt like being stuck in a story my body won't stop telling—even when everything around me says I'm safe. This is the first time a model actually made sense of why I'm still reacting the way I do.
- **For me,** PTSD has never been about fear so much as **being on guard all the time**. It's not just a reaction, it's a posture. I'm always scanning, always planning escape routes, trying to control everything. That made total sense when I was in a dangerous environment—but now, it's exhausting and doesn't fit my life anymore. **The way you framed this** as 'harm expectation' and defensive mobilization **really captured that**.
- [The] focus on persistent hyper-alertness **struck a chord with me**, [it is] **on the mark based on my personal experience**. [The] concept of impaired belief-updating makes sense to me and could explain why all the reassurance I've received hasn't helped. [How the analysis] describes flashbacks makes sense. [The] concept of self-reinforcing trauma cues makes a lot of sense, and **reflects my personal experience**.
- **This model really resonated** because it acknowledges that trauma doesn't happen in a vacuum. My harm expectations are shaped by racism, cultural identity, and systemic violence. The concept of moral injury is huge for me. It's not just about what happened, but how it conflicts with who I am, my values, and how I see the world. That's often left out of clinical models.
- [The analysis is] **enlightening and re-assuring from a human perspective**. [It is] grounding and empowering, in that it helps develop a better sense of situational awareness and sense of self [which] improves confidence in self-efficacy. [It is] presented in a highly cogent and intuitive manner with clarity and nuance. [It] enhances understanding, meaning and clinical value in its broadest sense.

ELE feedback *(continued).*

Potential of the analysis to help transform the self-perception of affected individuals

- The [analysis] **helps me trust that there's a logic behind my symptoms**, even if it's gone sideways. [The analysis] explains what my brain is doing, not just that it's doing something weird.
- **I've always been ashamed** of the triggers I have. I feel like [the analysis] kind of **sheds a new light on it**. Normally I just feel like the odd man out for how I feel.
- I appreciate that [the analysis] **doesn't blame the individual**.
- [The analysis] can be of tremendous help in **destigmatising illness, including mitigating self-stigma and shame**.

Suggestions to inform co-production of further research

- [Include] more voices—especially from survivors, veterans, and marginalized communities. Emphasize how relational, cultural, and societal factors shape belief systems and recovery.
- [Include] more veteran voices.
- I'm really curious how this could translate into therapy—especially in ways that are relational and culturally aware.
- I'd be really interested in treatments that help recalibrate this ['vicious circle'] process, rather than just manage the fallout.
- Keep making room for those of us whose trauma shows up more like hyper-preparation.
- Consider more directly how [institutions, cultures, and industries] can reinforce avoidance, over-medicalization, or mislabelling of trauma [which can] block or distort the healing process.

References

1. Linson, A., Parr, T. & Friston, K. J. Active inference, stressors, and psychological trauma: A neuroethological model of (mal) adaptive explore-exploit dynamics in ecological context. *Behav. Brain Res.* **380**, 112421 (2020).
2. Miles, S. R. *et al.* Hyperarousal symptoms linger after successful PTSD treatment in active duty military. *Psychol. Trauma Theory Res. Pract. Policy* **15**, 1398–1405 (2023).
3. Cooke, D. F. & Graziano, M. S. A. Sensorimotor Integration in the Precentral Gyrus: Polysensory Neurons and Defensive Movements. *J. Neurophysiol.* **91**, 1648–1660 (2004).
4. Cooke, D. F. & Graziano, M. S. A. Super-Flinchers and Nerves of Steel: Defensive Movements Altered by Chemical Manipulation of a Cortical Motor Area. *Neuron* **43**, 585–593 (2004).

5. Cisler, J. M., Dunsmoor, J. E., Fonzo, G. A. & Nemeroff, C. B. Latent-state and model-based learning in PTSD. *Trends Neurosci.* **47**, 150–162 (2024).

6. Tsuda, B., Pate, S. C., Tye, K. M., Siegelmann, H. T. & Sejnowski, T. J. Neuromodulators generate multiple context-relevant behaviors in a recurrent neural network by shifting activity flows in hyperchannels. *bioRxiv* 2021.05.31.446462 (2024) doi:10.1101/2021.05.31.446462.

7. Mobbs, D., Hagan, C. C., Dalgleish, T., Silston, B. & Prévost, C. The ecology of human fear: survival optimization and the nervous system. *Front. Neurosci.* **9**, (2015).

8. Maren, S., Phan, K. L. & Liberzon, I. The contextual brain: implications for fear conditioning, extinction and psychopathology. *Nat. Rev. Neurosci.* **14**, 417–428 (2013).

9. VanElzakker, M. B., Kathryn Dahlgren, M., Caroline Davis, F., Dubois, S. & Shin, L. M. From Pavlov to PTSD: The extinction of conditioned fear in rodents, humans, and anxiety disorders. *Extinction* **113**, 3–18 (2014).

10. Bouyeure, A. *et al.* Distinct representational properties of cues and contexts shape fear learning and extinction. (2025) doi:10.7554/elife.105126.1.

11. Moutoussis, M., Shahar, N., Hauser, T. U. & Dolan, R. J. Computation in Psychotherapy, or How Computational Psychiatry Can Aid Learning-Based Psychological Therapies. *Comput. Psychiatry* (2018) doi:10.1162/CPSY_a_00014.

12. Linson, A. & Friston, K. Reframing PTSD for computational psychiatry with the active inference framework. *Cognit. Neuropsychiatry* **24**, 347–368 (2019).

13. Seriès, P. Post-traumatic stress disorder as a disorder of prediction. *Nat. Neurosci.* **22**, 334–336 (2019).

14. Seriès, P., Veerapa, E. & Jardri, R. Can computational models help elucidate the link between complex trauma and hallucinations? *Hallucinations Neurobiol. Patient Exp.* **265**, 66–73 (2024).

15. Van der Kolk, B. *et al.* A randomized clinical trial of EMDR, fluoxetine and pill placebo in the treatment of PTSD: Treatment effects and long-term maintenance. *J. Clin. Psychiatry* **68**, 37–46 (2007).

16. Kaye, A. P., Rao, M. G., Kwan, A. C., Ressler, K. J. & Krystal, J. H. A computational model for learning from repeated traumatic experiences under uncertainty. *Cogn. Affect. Behav. Neurosci.* **23**, 894–904 (2023).

17. Harnett, N. G., Fleming, L. L., Clancy, K. J., Ressler, K. J. & Rosso, I. M. Affective Visual Circuit Dysfunction in Trauma and Stress-Related Disorders. *Biol. Psychiatry* **97**, 405–416 (2025).

18. Lange, R. D., Chattoraj, A., Beck, J. M., Yates, J. L. & Haefner, R. M. A confirmation bias in perceptual decision-making due to hierarchical approximate inference. *PLOS Comput. Biol.* **17**, e1009517 (2021).

19. Garcia-Rill, E. Disorders of the reticular activating system. *Med. Hypotheses* **49**, 379–387 (1997).

20. Hobson, J. A., Gott, J. A. & Friston, K. J. Minds and Brains, Sleep and Psychiatry. *Psychiatr. Res. Clin. Pract.* **3**, 12–28 (2021).

21. Mendoza Alvarez, M. *et al.* Systematic review: REM sleep, dysphoric dreams and nightmares as transdiagnostic features of psychiatric disorders with emotion dysregulation - Clinical implications. *Sleep Med.* **127**, 1–15 (2025).

22. Krueger, J. & Disney, A. A. Structure and function of dual-source cholinergic modulation in early vision. *J. Comp. Neurol.* **527**, 738–750 (2019).

23. Garcia-Rill, E. Chapter 5 - Posttraumatic stress and anxiety, the role of arousal. in *Arousal in Neurological and Psychiatric Diseases* (ed. Garcia-Rill, E.) 67–81 (Academic Press, 2019). doi:10.1016/B978-0-12-817992-5.00005-2.

24. Mellman, T. A., Kulick-Bell, R., Ashlock, L. E. & Nolan, B. Sleep events among veterans with combat-related posttraumatic stress disorder. *Am. J. Psychiatry* **152**, 110–115 (1995).

25. Gott, J. A., Stücker, S., Kanske, P., Haaker, J. & Dresler, M. Acetylcholine and metacognition during sleep. *Conscious. Cogn.* **117**, 103608 (2024).

26. Ben-Zion, Z. *et al.* Connectome-Based Predictive Modeling of PTSD Development Among Recent Trauma Survivors. *JAMA Netw. Open* **8**, e250331 (2025).

27. Zhou, L. *et al.* Organization of Functional Long-Range Circuits Controlling the Activity of Serotonergic Neurons in the Dorsal Raphe Nucleus. *Cell Rep.* **18**, 3018–3032 (2017).

28. Michels, L. *et al.* Prefrontal GABA and glutathione imbalance in posttraumatic stress disorder: Preliminary findings. *Psychiatry Res. Neuroimaging* **224**, 288–295 (2014).

29. Shohamy, D. & Adcock, R. A. Dopamine and adaptive memory. *Trends Cogn. Sci.* **14**, 464–472 (2010).

30. Nyberg, L. *et al.* Dopamine D2 receptor availability is linked to hippocampal–caudate functional connectivity and episodic memory. *Proc. Natl. Acad. Sci.* **113**, 7918–7923 (2016).

31. Schott, B. H. *et al.* The Dopaminergic Midbrain Participates in Human Episodic Memory Formation: Evidence from Genetic Imaging. *J. Neurosci.* **26**, 1407 (2006).

32. Bentley, P., Vuilleumier, P., Thiel, C. M., Driver, J. & Dolan, R. J. Cholinergic enhancement modulates neural correlates of selective attention and emotional processing. *NeuroImage* **20**, 58–70 (2003).

33. Picciotto, M. R., Higley, M. J. & Mineur, Y. S. Acetylcholine as a Neuromodulator: Cholinergic Signaling Shapes Nervous System Function and Behavior. *Neuron* **76**, 116–129 (2012).

34. Naegeli, C. *et al.* Locus Coeruleus Activity Mediates Hyperresponsiveness in Posttraumatic Stress Disorder. *Biol. Psychiatry* **83**, 254–262 (2018).

35. Dijkstra, N., von Rein, T., Kok, P. & Fleming, S. M. A neural basis for distinguishing imagination from reality. *Neuron* (2025) doi:10.1016/j.neuron.2025.05.015.

36. Monsalve-Mercado, M. M., Stine, G. M., Shadlen, M. N. & Miller, K. D. The geometry of the neural state space of decisions. *bioRxiv* 2025.01.24.634806 (2025) doi:10.1101/2025.01.24.634806.

37. Mirza, M. B., Adams, R. A., Friston, K. & Parr, T. Introducing a Bayesian model of selective attention based on active inference. *Sci. Rep.* **9**, 13915 (2019).

38. Khona, M. & Fiete, I. R. Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* **23**, 744–766 (2022).

39. Chiarotti, F., Viglione, A., Giuliani, A. & Branchi, I. Citalopram amplifies the influence of living conditions on mood in depressed patients enrolled in the STAR*D study. *Transl. Psychiatry* **7**, e1066–e1066 (2017).

40. Carhart-Harris, R. L. *et al.* Canalization and plasticity in psychopathology. *Neuropharmacology* **226**, 109398 (2023).

41. Brandman, T. & Peelen, M. V. Objects sharpen visual scene representations: evidence from MEG decoding. *Cereb. Cortex* **33**, 9524–9531 (2023).

42. Brady, T. F., Shafer-Skelton, A. & Alvarez, G. A. Global ensemble texture representations are critical to rapid scene perception. *J. Exp. Psychol. Hum. Percept. Perform.* **43**, 1160–

1176 (2017).

43. Williams, J. R., Markov, Y. A., Tiurina, N. A. & Störmer, V. S. What You See Is What You Hear: Sounds Alter the Contents of Visual Perception. *Psychol. Sci.* **33**, 2109–2122 (2022).

44. Wright, T. M., Pelphrey, K. A., Allison, T., McKeown, M. J. & McCarthy, G. Polysensory Interactions along Lateral Temporal Regions Evoked by Audiovisual Speech. *Cereb. Cortex* **13**, 1034–1043 (2003).

45. Beauchamp, M. S., Yasar, N. E., Frye, R. E. & Ro, T. Touch, sound and vision in human superior temporal sulcus. *NeuroImage* **41**, 1011–1020 (2008).

46. Butter, C. M., Buchtel, H. A. & Santucci, R. Spatial attentional shifts: Further evidence for the role of polysensory mechanisms using visual and tactile stimuli. *Neuropsychologia* **27**, 1231–1240 (1989).

47. Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P. & Pezzulo, G. Active inference: a process theory. *Neural Comput.* **29**, 1–49 (2017).

48. Wald, A. An Essentially Complete Class of Admissible Decision Functions. *Ann. Math. Stat.* **18**, 549–555 (1947).

49. Brown, L. D. A Complete Class Theorem for Statistical Problems with Finite Sample Spaces. *Ann. Stat.* **9**, 1289–1300 (1981).

50. Reddy, L. & Kanwisher, N. Category Selectivity in the Ventral Visual Pathway Confers Robustness to Clutter and Diverted Attention. *Curr. Biol.* **17**, 2067–2072 (2007).

51. Mueller-Pfeiffer, C. *et al.* Atypical visual processing in posttraumatic stress disorder. *NeuroImage Clin.* **3**, 531–538 (2013).

52. Da Costa, L. *et al.* Active inference on discrete state-spaces: A synthesis. *J. Math. Psychol.* **99**, 102447 (2020).

53. Parr, T., Da Costa, L. & Friston, K. Markov blankets, information geometry and stochastic thermodynamics. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **378**, 20190159 (2019).

54. Da Costa, L., Parr, T., Sengupta, B. & Friston, K. Neural Dynamics under Active Inference: Plausibility and Efficiency of Information Processing. *Entropy* **23**, (2021).

55. Wurtz, R. H., McAlonan, K., Cavanaugh, J. & Berman, R. A. Thalamic pathways for active vision. *Trends Cogn. Sci.* **15**, 177–184 (2011).

56. Parr, T. & Friston, K. J. The active construction of the visual world. *Neuropsychologia* **104**, 92–101 (2017).

57. Mirza, M. B., Adams, R. A., Mathys, C. & Friston, K. J. Human visual exploration reduces uncertainty about the sensed world. *PLOS ONE* **13**, e0190429 (2018).

58. Rao, R. P. N., Gklezakos, D. C. & Sathish, V. Active Predictive Coding: A Unifying Neural Model for Active Perception, Compositional Learning, and Hierarchical Planning. *Neural Comput.* **36**, 1–32 (2023).

59. Dijkstra, N., Zeidman, P., Ondobaka, S., van Gerven, M. A. J. & Friston, K. Distinct Top-down and Bottom-up Brain Connectivity During Visual Perception and Imagery. *Sci. Rep.* **7**, 5677 (2017).

60. Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C. & de Lange, F. P. Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Curr. Biol.* **23**, 1427–1431 (2013).

61. Tomić, I. & Bays, P. M. A dynamic neural resource model bridges sensory and working memory. *eLife* **12**, RP91034 (2024).

62. Schwartz, S. *et al.* Attentional Load and Sensory Competition in Human Vision: Modulation of fMRI Responses by Load at Fixation during Task-irrelevant Stimulation in the Peripheral Visual Field. *Cereb. Cortex* **15**, 770–786 (2005).

63. Stolte, M., Bahrami, B. & Lavie, N. High perceptual load leads to both reduced gain and broader orientation tuning. *J. Vis.* **14**, 9–9 (2014).

64. Torralbo, A., Kelley, T. A., Rees, G. & Lavie, N. Attention induced neural response trade-off in retinotopic cortex under load. *Sci. Rep.* **6**, 33041 (2016).

65. Clancy, K., Ding, M., Bernat, E., Schmidt, N. B. & Li, W. Restless 'rest': intrinsic sensory hyperactivity and disinhibition in post-traumatic stress disorder. *Brain* **140**, 2041–2050 (2017).

66. Harrison, S. A. & Tong, F. Decoding reveals the contents of visual working memory in early visual areas. *Nature* **458**, 632–635 (2009).

67. Teng, C. & Kravitz, D. J. Visual working memory directly alters perception. *Nat. Hum. Behav.* **3**, 827–836 (2019).

68. S. A. Brandt & L. W. Stark. Spontaneous Eye Movements During Visual Imagery Reflect the Content of the Visual Scene. *J. Cogn. Neurosci.* **9**, 27–38 (1997).

69. Spivey, M. J. & Geng, J. J. Oculomotor mechanisms activated by imagery and memory: eye movements to absent objects. *Psychol. Res.* **65**, 235–241 (2001).

70. Hobson, J. A. & Friston, K. J. Waking and dreaming consciousness: Neurobiological and functional considerations. *Prog. Neurobiol.* **98**, 82–98 (2012).

71. Andrade, J., Kavanagh, D. & Baddeley, A. Eye-movements and visual imagery: A working memory approach to the treatment of post-traumatic stress disorder. *Br. J. Clin. Psychol.* **36**, 209–223 (1997).

72. Itti, L. & Baldi, P. Bayesian surprise attracts human attention. *Vis. Atten. Psychophys. Electrophysiol. Neuroimaging* **49**, 1295–1306 (2009).

73. Sun, Y., Gomez, F. & Schmidhuber, J. Planning to Be Surprised: Optimal Bayesian Exploration in Dynamic Environments. in *Artificial General Intelligence* (eds. Schmidhuber, J., Thórisson, K. R. & Looks, M.) 41–51 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2011).

74. Parr, T. & Friston, K. J. Working memory, attention, and salience in active inference. *Sci. Rep.* **7**, 14678 (2017).

75. Parr, T. & Friston, K. J. Attention or salience? *Atten. Percept.* **29**, 1–5 (2019).

76. Bradley, M. M., Cuthbert, B. N. & Lang, P. J. Pictures as prepulse: Attention and emotion in startle modification. *Psychophysiology* **30**, 541–545 (1993).

77. Cotton, K. & Ricker, T. J. Examining the relationship between working memory consolidation and long-term consolidation. *Psychon. Bull. Rev.* **29**, 1625–1648 (2022).

78. Smith, R., Friston, K. J. & Whyte, C. J. A step-by-step tutorial on active inference and its application to empirical data. *J. Math. Psychol.* **107**, 102632 (2022).

79. Horner, A. J. & Burgess, N. The associative structure of memory for multi-element events. *J. Exp. Psychol. Gen.* **142**, 1370–1383 (2013).

80. Horner, A. J., Bisby, J. A., Bush, D., Lin, W.-J. & Burgess, N. Evidence for holistic episodic recollection via hippocampal pattern completion. *Nat. Commun.* **6**, 7462 (2015).

81. Koolschijn, R. S. *et al.* Memory recall involves a transient break in excitatory-inhibitory balance. *eLife* **10**, e70071 (2021).

82. Holton, E., Braun, L., Thompson, J. A. F., Grohn, J. & Summerfield, C. Humans and neural

networks show similar patterns of transfer and interference during continual learning. Preprint at https://doi.org/10.31234/osf.io/98ksw_v1 (2025).

83. Park, J., Holmes, C. D. & Snyder, L. H. Compositional architecture: Orthogonal neural codes for task context and spatial memory in prefrontal cortex. *bioRxiv* 2025.02.25.640211 (2025) doi:10.1101/2025.02.25.640211.

84. Wills, T. J., Lever, C., Cacucci, F., Burgess, N. & O'Keefe, J. Attractor Dynamics in the Hippocampal Representation of the Local Environment. *Science* **308**, 873–876 (2005).

85. McKenzie, S. *et al.* Hippocampal Representation of Related and Opposing Memories Develop within Distinct, Hierarchically Organized Neural Schemas. *Neuron* **83**, 202–215 (2014).

86. Ólafsdóttir, H. F., Carpenter, F. & Barry, C. Task Demands Predict a Dynamic Switch in the Content of Awake Hippocampal Replay. *Neuron* **96**, 925-935.e6 (2017).

87. Favila, S. E., Chanales, A. J. H. & Kuhl, B. A. Experience-dependent hippocampal pattern differentiation prevents interference during subsequent learning. *Nat. Commun.* **7**, 11066 (2016).

88. Zotow, E., Bisby, J. A. & Burgess, N. Behavioral evidence for pattern separation in human episodic memory. *Learn. Mem.* **27**, 301–309 (2020).

89. Friston, K. *et al.* Active inference and learning. *Neurosci. Biobehav. Rev.* **68**, 862–879 (2016).

90. Friston, K. & Buzsáki, G. The Functional Anatomy of Time: What and When in the Brain. *Trends Cogn. Sci.* **20**, 500–511 (2016).

91. Brewin, C. R. & Burgess, N. Contextualisation in the revised dual representation theory of PTSD: A response to Pearson and colleagues. *J. Behav. Ther. Exp. Psychiatry* **45**, 217–219 (2014).

92. Sarto-Jackson, I. *The Making and Breaking of Minds: How Social Interactions Shape the Human Mind*. (Vernon Press, 2022).

93. Heinbockel, H., Wagner, A. D. & Schwabe, L. Post-retrieval stress impairs subsequent memory depending on hippocampal memory trace reinstatement during reactivation. *Sci. Adv.* **10**, eadm7504.

94. Perl, O. *et al.* Neural patterns differentiate traumatic from sad autobiographical memories in PTSD. *Nat. Neurosci.* **26**, 2226–2236 (2023).

95. Mellman, T. A., Bustamante, V., Fins, A. I., Pigeon, W. R. & Nolan, B. REM Sleep and the Early Development of Posttraumatic Stress Disorder. *Am. J. Psychiatry* **159**, 1696–1701 (2002).

96. Chang, H. *et al.* Sleep microstructure organizes memory replay. *Nature* **637**, 1161–1169 (2025).

97. Bollmann, L., Baracskay, P., Stella, F. & Csicsvari, J. Sleep stages antagonistically modulate reactivation drift. *Neuron* (2025) doi:10.1016/j.neuron.2025.02.025.

98. Balouek, J.-A. *et al.* Reactivation of Early-Life Stress-Sensitive Neuronal Ensembles Contributes to Lifelong Stress Hypersensitivity. *J. Neurosci.* **43**, 5996 (2023).

99. Geva, N., Deitch, D., Rubin, A. & Ziv, Y. Time and experience differentially affect distinct aspects of hippocampal representational drift. *Neuron* **111**, 2357-2366.e5 (2023).

100. McNamara, P., Wildman, W. J., Hodulik, G. & Rohr, D. A neurocomputational theory of nightmares: the role of formal properties of nightmare images. *SLEEP Adv.* **2**, zpab009 (2021).

Acknowledgements

The authors thank the Experts by Lived Experience (ELEs) who individually reviewed this work and provided feedback, including Rory Burke, Cody Cerne, and those who requested anonymity. AL received research funding support from the Open University School of Computing & Communications, and was hosted during manuscript editing by the Konrad Lorenz Institute for Evolution & Cognition Research. KF is supported by funding from the Wellcome Trust (Ref: 226793/Z/22/Z). The Functional Imaging Laboratory is funded by a Wellcome Discovery Grant.

Supplementary information

Simulations were performed in Matlab using a standard routine, spm_MDP_VB_X.m, included with the SPM software (http://www.fil.ion.ucl.ac.uk/spm/). Code available upon reasonable request (will be included in version of record). No Generative AI was used in the creation of this manuscript.