

NirWire — Upstream Neuroplasticity in Conditioned Reactive Patterns Through Conversational AI

Eranga Bandara^a, Ross Gore^a, Asanga Gunaratna^b, Ravi Mukkamala^a, Sachini Rajapakse^c, Isurunima Kularathna^c, Pramoda Karunarathna^c, Chalani Rajapakse^c, Sachin Shetty^a, Christopher K. Rhea^a, Ng Wee Keong^c, Kasun De Zoysa^d, Amin Hass^f, Shaifali Kaushik^l, Wathsala Herath^g, Preston Samuel^k, Anita H. Clayton^j, Atmaram Yarlagadda^l

^a*Old Dominion University, Norfolk, VA, USA*

^b*AI Motion Labs, Melbourne, Australia*

^c*Nanyang Technological University, Singapore*

^d*University of Colombo, Sri Lanka*

^e*IcicleLabs.AI*

^f*Accenture Technology Labs, Arlington, VA, USA*

^g*Agentsway.AI*

^h*GSI Scandinavia AB*

ⁱ*Lithuanian University of Health Sciences*

^j*Department of Psychiatry and Neurobehavioral Sciences,*

University of Virginia School of Medicine, Charlottesville, VA, USA

^k*Blanchfield Army Community Hospital, Fort Campbell, KY, USA*

^l*McDonald Army Health Center, Newport News, VA, USA*

Abstract

Conditioned reactive patterns – Hebbian-encoded neurone paths built

Email addresses: cmedawer@odu.edu (Eranga Bandara), rgore@odu.edu (Ross Gore), asanga.gunaratna@complianceoslab.app (Asanga Gunaratna), rmukkama@odu.edu (Ravi Mukkamala), sachini.rajapakse@iciclelabs.ai (Sachini Rajapakse), isurunima.kularathna@iciclelabs.ai (Isurunima Kularathna), pramoda.karunarathna@iciclelabs.ai (Pramoda Karunarathna), chalani.rajapakse@iciclelabs.ai (Chalani Rajapakse), sshetty@odu.edu (Sachin Shetty), crhea@odu.edu (Christopher K. Rhea), awkng@ntu.edu.sg (Ng Wee Keong), kasun@ucsc.cmb.ac.lk (Kasun De Zoysa), amin.hassanzadeh@accenture.com (Amin Hass), Skaus2000@gmail.com (Shaifali Kaushik), wathsala.herath@agentsway.ai (Wathsala Herath), preston.1.samuel.mil@health.mil (Preston Samuel), AHC8V@uvahealth.org (Anita H. Clayton), atmaram.yarlagadda.civ@health.mil (Atmaram Yarlagadda)

through repetition, single-event intensity, or epigenetic inheritance firing automatically below awareness – underlie a broad spectrum of suffering, from PTSD and combat trauma to everyday anger, relationship avoidance, and performance anxiety. Existing interventions (CBT, prolonged exposure, EMDR) are validated but act downstream of the reactive cascade, after the path has fired; long-term potentiation (LTP) continues to operate on the path with every episode, and the conditioned path itself is not structurally targeted. We propose NirWire, a theoretical framework and conversational AI system for upstream neuroplasticity in conditioned reactive patterns. We first develop an end-to-end account of how these patterns arise as physical neurone paths – from synaptic encoding, to compression into implicit-belief routing, to millisecond-scale affective routing, to the opposing mechanisms of LTP and long-term depression (LTD). NirWire then targets the feeling tone gap: the 100-200ms prefrontal-amygdala window at which top-down modulation can intercept the cascade before LTP operates. Guided observation at this window produces LTD of the underlying connections, accumulating into structural reorganisation – upstream neuroplasticity, not downstream management. We make three contributions: (1) a situation-keyed compressed routing model of implicit belief encoding, analogous to Huffman coding; (2) encoding-depth-dependent stimulus generalisation, whereby heavily encoded paths fire from a single attribute; and (3) a five-function on-device conversational AI architecture operationalising upstream intervention. The framework is illustrated through a worked post-traumatic stress case; a functional prototype has been built and technically validated, and we set out the studies required to validate its clinical efficacy.

Keywords: Agentic AI, False Memory, Forensic Psychology, Responsible AI, Explainable AI, LLM, Model Context Protocol

1. Introduction

Every day, millions of people experience psychological reactions they did not choose and cannot easily explain. A nurse eight months after a mass casualty incident finds herself unable to return to the emergency room — not because she lacks courage or professional commitment, but because the sound of an ambulance siren activates a cascade of fear and panic before any conscious thought has formed [1]. A partner reacts with sudden hurt and withdrawal to a neutral comment — not because the comment was unkind,

but because repeated experiences in which a particular tone of voice preceded conflict have encoded an automatic routing from “that tone” to “I am being rejected” that fires in under 200 milliseconds [2, 3]. A soldier returned from deployment brakes hard at the sight of any object on a roadside — not because he cannot distinguish debris from an improvised explosive device, but because one overwhelming combat event has encoded a vehicle-threat path that overrides deliberate tactical assessment before it can engage [4]. These are not character flaws, failures of willpower, or signs of psychological weakness. They are *conditioned reactive patterns* — Hebbian-encoded neurone paths built into the brain’s physical architecture through repetition, single-event intensity, or epigenetic inheritance [5, 6]. They share a defining property: they fire automatically, below the threshold of conscious awareness, producing outcomes the person did not choose and often cannot predict. The encoding is physical, the activation is automatic, and the cascade completes before deliberate prefrontal processing can engage.

The neurobiological basis of such patterns is well established. Hebbian long-term potentiation (LTP) provides the physical substrate for experience-dependent encoding: co-activated neurons strengthen their connections, and repeated or intense activation widens the resulting path, lowers its activation threshold, and increases its firing speed [5, 7, 3]. The amygdala performs rapid affective appraisal of incoming signals within roughly 100–200ms, generating a pre-cognitive affective signal that initiates the reactive cascade before conscious awareness establishes itself [8, 9, 10]. The implicit beliefs encoded in these paths are not consciously held propositions but compressed automatic routing codes: the most frequently or intensely activated paths become the fastest, most automatic routes in the brain’s affective architecture [11, 12]. We develop this end-to-end mechanism in full in Section 3; here we note only its central consequence — that these paths deepen with every triggered episode, which is why untreated post-traumatic stress disorder (PTSD) worsens rather than resolves over time, and why avoidance reliably intensifies the underlying pattern rather than relieving it [4, 13, 14].

Existing pharmacological and psychological interventions are clinically validated and effective. Cognitive behavioural therapy (CBT) produces reliable symptom reductions across anxiety, depression, and PTSD [15, 16]; prolonged exposure and Eye Movement Desensitisation and Reprocessing (EMDR) demonstrate significant efficacy in PTSD [13, 17]; and selective serotonin reuptake inhibitors reduce amygdala reactivity and improve affect regulation [18, 19]. Yet all of these interventions share a structural character-

istic: they intervene *downstream* of the reactive cascade, after the encoding path has fired to completion. CBT restructures the narrative constructed around the feeling after the cascade has mobilised; prolonged exposure extinguishes the conditioned association, but activation still occurs and LTP continues to operate on the path during every session; pharmacological agents modulate neurotransmission without altering the specific synaptic architecture of the conditioned path. The consequence is that the conditioned path itself — the physical synaptic structure that generates the cascade — is not structurally targeted by any existing intervention. This is not a criticism of existing treatments, which represent significant clinical achievements; it is the identification of a specific mechanistic gap. No existing intervention targets the upstream moment where structural dissolution of the conditioned path becomes possible.

That upstream moment has a precise neurobiological identity. Affective neuroscience identifies a brief prefrontal-amygdala regulatory window — approximately 100–200ms after stimulus registration — during which top-down modulation can intercept the downstream cascade before it completes its routing to the reactive destination [20, 21, 22, 23]. We refer to this window as the *feeling tone gap*: the interval between the arising of the pre-cognitive affective signal and the completion of the reactive cascade. In untrained awareness this gap is effectively invisible, and conscious awareness registers only the downstream product — the anger, the fear, the anxiety — rather than the process that generated it; with systematic metacognitive training, the capacity to engage this window increases substantially [12, 24, 25]. Crucially, engaging the gap does not merely modulate the cascade output. When the cascade is intercepted before completion, the reactive path fails to undergo LTP-mediated strengthening, and the opposing synaptic mechanism — long-term depression (LTD) — operates instead, physically weakening the connection [3, 26, 27]. Accumulated across many sessions, these interceptions produce structural neural reorganisation. This is the distinction between *suppression* — managing the output of a completed cascade, leaving the path and its LTP intact — and *dissolution* — intercepting the cascade upstream, producing LTD-mediated weakening of the path itself.

We propose NirWire — a theoretical framework and conversational AI system for upstream neuroplasticity in conditioned reactive patterns. This paper first develops a complete end-to-end account of how psychological reactive patterns arise as physical neurone paths — from the single neurone and synapse, to the neurone path as a chain of connected neurons, to the deeply

encoded paths that constitute implicit beliefs, to the millisecond-scale affective routing that activates them below awareness, and to the opposing mechanisms of LTP and LTD that respectively deepen and dissolve them [19, 28]. Building on this model, NirWire operationalises upstream intervention at the feeling tone gap through guided metacognitive observation: rather than managing the cascade output, the NirWire agent guides the user to observe the neurological machinery generating the cascade in real time, at the moment of activation, before the routing completes [29]. This observing position — seeing the path rather than being run by it — constitutes the upstream interception that initiates LTD dissolution. Conversational AI is uniquely suited to this task: the feeling tone gap occurs in the moment of activation, and a system on the user’s own device provides proximity to the intervention site that no scheduled clinical session can match [30, 31]; the specificity of conditioned patterns requires calibrated, personalised contact events rather than generic stimuli [14]; and an on-device, privacy-preserving deployment with zero data egress suits contexts where data sensitivity is paramount, including military and clinical settings [32, 33].

This paper makes three theoretical contributions. First, we describe implicit belief encoding as a situation-keyed compressed routing system — analogous to Huffman encoding in computer science [34] — in which the incoming situational context (tone, facial expression, body language, words, location, sound, smell, and associated memory) is matched against encoded path signatures through approximate pattern matching to generate a pre-cognitive affective routing decision, extending existing accounts of rapid affective appraisal [8, 9] by specifying the structure of the situational key. Second, we propose *encoding-depth-dependent stimulus generalisation*: the situational specificity required to activate a conditioned path is inversely proportional to its encoding depth, so that thinly encoded paths require the full complement of original attributes while heavily encoded paths activate from a single matching attribute — accounting for the clinical observation that trauma responses trigger from apparently unrelated stimuli [4, 35]. Third, we describe a five-function conversational AI architecture — intake mapping, personalised psychoeducation, stimulus ladder construction, guided observation sessions, and real-time neurological narration — that operationalises upstream neuroplasticity at the feeling tone gap in a privacy-preserving on-device format.

The remainder of this paper is organised as follows. Section 2 reviews related work and positions NirWire against existing trauma therapies, contemplative and constructionist approaches, and digital and on-device systems.

Section 3 develops the theoretical foundations — the complete end-to-end account from the single neurone through path formation, situations and the senses, deeply encoded implicit beliefs, situation-keyed affective routing, and the opposing mechanisms of LTP and LTD — and introduces the four moments of observation. Section 4 describes the NirWire system — its five functions and its on-device technical architecture — and illustrates them through a worked case. Section 5 discusses positioning relative to existing treatments, limitations, and future directions, and Section 6 concludes.

2. Related Work

NirWire draws on and departs from five strands of prior work: established trauma-focused psychotherapies, metacognitive and contemplative approaches to emotion regulation, constructionist accounts of affect, digital and conversational mental-health systems, and on-device language models. We review each in turn and then position NirWire against representative systems in Table 1.

Trauma-focused psychotherapies. The dominant evidence-based treatments for post-traumatic stress disorder operate through extinction learning and cognitive restructuring. Prolonged exposure therapy repeatedly activates the fear memory under safe conditions to promote inhibitory learning [36], and Eye Movement Desensitisation and Reprocessing (EMDR) targets memory reconsolidation through dual-attention stimulation [37]. The Ehlers and Clark cognitive model attributes persistent PTSD to a disturbance in the way the trauma memory is represented and appraised [38]. These approaches are clinically effective and remain the standard of care. However, each intervenes after the reactive cascade has been activated — the fear response is deliberately evoked and then processed — rather than intercepting it upstream, at the pre-cognitive window before the cascade completes. In the synaptic terms we develop in Section 3, the path is repeatedly fired and then re-appraised; the structural dissolution of the path through interception prior to firing is not the operative mechanism.

Metacognitive and contemplative approaches. A second strand targets the relationship between awareness and reactivity rather than the content of the reaction. Mindfulness-based cognitive therapy links relapse prevention to metacognitive awareness — the capacity to relate to thoughts and feelings as transient mental events rather than facts [39]. Neuroimaging of experienced meditators shows altered default-mode network activity

consistent with reduced self-referential elaboration [40], and the enactive tradition frames cognition as embodied and observation as constitutive of experience [41]. This literature is the closest conceptual antecedent to NirWire’s observation-based mechanism: it identifies the observing stance as regulatory and gestures at an early interceptive window. What it does not provide is a specific synaptic account tying that observing stance to long-term depression (LTD) of an identified reactive path, a personalised and calibrated delivery method, or a scalable technical implementation. NirWire operationalises the contemplative insight as a mechanistic, individually calibrated intervention.

Constructionist accounts of affect. The theory of constructed emotion holds that affective experience is assembled from core affect and prior conceptual knowledge rather than triggered as a fixed program [42]. This account is compatible with our situation-keyed encoding model and with the existence of a pre-conceptual affective signal preceding the constructed emotional experience. It is, however, a descriptive theory of how affect arises; it does not propose an intervention that acts on the encoded routing itself, nor a mechanism for its structural weakening.

Digital and conversational mental-health systems. Fully automated conversational agents have demonstrated feasibility and symptom benefit at scale. Woebot delivered cognitive-behavioural content and reduced depressive symptoms in a randomised trial [43], and Wysa showed engagement and mood benefits through an empathy-oriented conversational agent [44]. Virtual-reality exposure systems such as BRAVEMIND deliver graded trauma cues for combat-related PTSD [45]. These systems establish that scalable, engaging digital delivery is achievable, but they port existing downstream protocols — CBT content, exposure cues — into new delivery channels; none targets the feeling tone gap or an LTD-based dissolution mechanism, and most are cloud-dependent rather than privacy-preserving on-device.

On-device and mental-health language models. Recent work explores language models specialised for mental-health text analysis, such as MentaLLaMA for interpretable analysis of social-media language [46], and compact models such as Phi-3 Mini that make capable on-device inference feasible on consumer hardware [47]. This work supplies part of NirWire’s technical substrate — private, on-device conversational capability — but is not itself an intervention framework: MentaLLaMA analyses rather than intervenes, and Phi-3 Mini is a general-purpose model without any associated therapeutic mechanism.

Positioning. Table 1 compares representative systems across the dimensions that define NirWire’s contribution. The pattern is consistent: trauma therapies act downstream; contemplative and constructionist approaches identify the observing stance and the constructed nature of affect but supply no synaptic dissolution mechanism or delivery method; and digital systems provide scalable delivery but transport downstream protocols. NirWire is distinguished not by any single column but by their conjunction — an upstream, feeling-tone-gap intervention grounded in an explicit LTP/LTD account, structured by a three-layer observation scaffold and a situation-keyed encoding model, and delivered through a privacy-preserving on-device conversational agent.

Table 1: Comparison of NirWire with representative related work. ✓ = supported; ✗ = not supported; ~ = partially supported.

Work	Trauma / PTSD focus	Upstream intervention	Feeling tone gap	Three-layer observation	LTD dissolution	Situation-keyed encoding	On-device privacy	Mobile delivery
Foa et al. [36]	✓	✗	✗	✗	~	✗	✗	✗
Shapiro [37]	✓	✗	✗	✗	~	✗	✗	✗
Ehlers & Clark [38]	✓	✗	✗	✗	✗	~	✗	✗
Teasdale et al. [39]	✗	~	~	~	✗	✗	✗	✗
Brewer et al. [40]	✗	~	~	✗	~	✗	✗	✗
Varela et al. [41]	✗	~	✗	✗	✗	✗	✗	✗
Barrett [42]	✗	~	~	~	~	~	✗	✗
Fitzpatrick et al. [43]	✗	✗	✗	✗	✗	✗	✗	✓
Inkster et al. [44]	✗	✗	✗	✗	✗	✗	✗	✓
Rizzo et al. [45]	✓	✗	✗	✗	~	✗	✗	✗
Yang et al. [46]	~	✗	✗	✗	✗	✗	✗	✗
Abdin et al. [47]	✗	✗	✗	✗	✗	✗	✓	✓
NirWire (proposed)	✓	✓	✓	✓	✓	✓	✓	✓

3. Theoretical Foundations

This section develops the end-to-end account on which NirWire rests. We proceed from the smallest unit — the single neurone and its synapse — to the neurone path, to the deeply encoded paths that constitute implicit beliefs, to the millisecond-scale affective routing that activates them below awareness, and finally to the opposing synaptic mechanisms of long-term potentiation and long-term depression that respectively deepen and dissolve them. The argument is cumulative: each subsection supplies a component that the NirWire framework (Section 4) later operationalises.

3.1. The Neurone Path: The Physical Substrate of Psychological Experience

The foundational claim of the framework is that psychological experience is physical. A thought, a habit, a fear, a belief, or a skill is not an abstract entity residing apart from the brain; it is the activity of a specific configuration of connected neurons [5]. Two neurons that fire in close temporal succession strengthen the synaptic connection between them, so that subsequent activation of the first more readily activates the second [7, 3]. Repeated across chains of neurons, this process produces a *neurone path*: a physical sequence of connected neurons that, once activated at its origin, propagates activation along its length to a characteristic outcome (Figure 1).

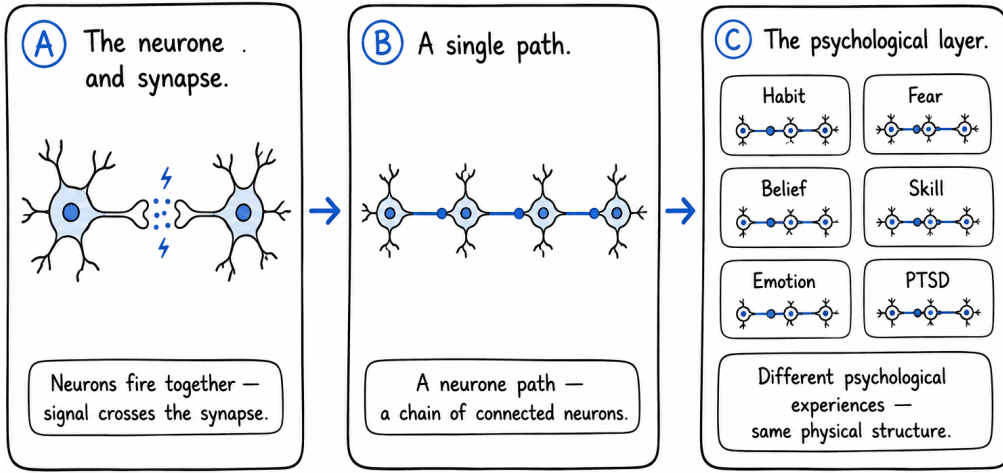


Figure 1: What neurone paths are. A single neurone communicates with the next by releasing neurotransmitters across the synapse (left). A chain of such connected neurons forms a neurone path (centre). Psychological phenomena as diverse as habits, fears, beliefs, skills, emotions, and post-traumatic stress share the same underlying physical structure — they differ in content and encoding depth, not in kind (right).

The psychological significance of a neurone path lies in what it does when it fires. A path activated by a particular situation routes toward a characteristic experience and behaviour. Critically, this routing occurs whether or not the person chooses it: once the path is activated at its origin, propagation is automatic. The subjective experience — the thought, the feeling, the urge — is the activation of the path, not a separate event that follows it (Figure 2).

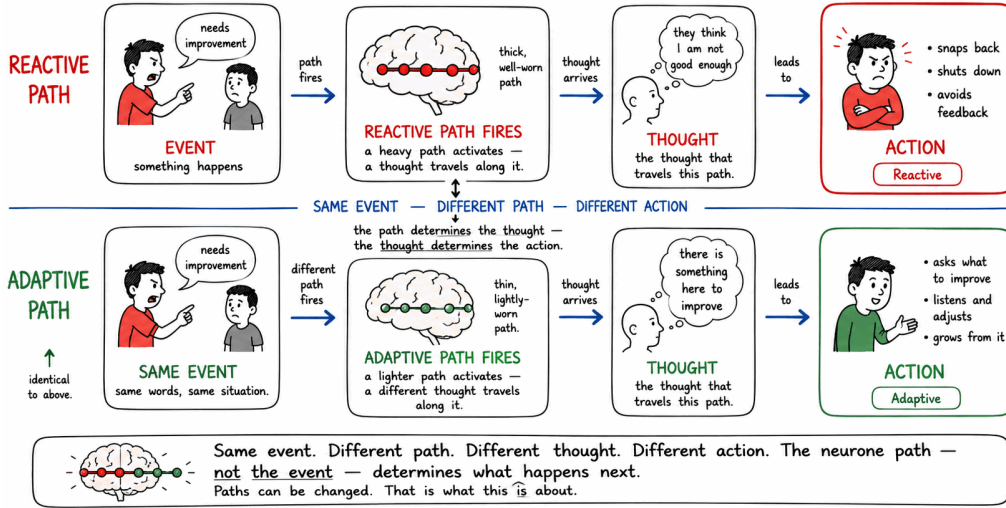


Figure 2: How a neurone path works, from event to action. The same event can activate different paths in different individuals, and the activated path — not the event itself — determines the resulting thought and action. A heavily encoded path (top) routes rapidly to a reactive outcome; a lighter path (bottom) routes to a considered response. The path determines the thought; the thought determines the action.

This reframing has a direct clinical consequence. If a reactive pattern is a physical path, then changing the pattern is a matter of changing the path — not of willpower, insight alone, or the suppression of an unwanted output. The remainder of this section establishes how paths form, how they deepen, and how they can be dissolved.

3.2. How Neurone Paths Form

Neurone paths are encoded through three mechanisms. The first is *repetition*: a situation and a response that co-occur repeatedly strengthen their connection incrementally, each repetition adding a small increment of synaptic efficacy [5, 7]. A child told repeatedly that their efforts are not good enough encodes an inadequacy path gradually, over years, until it activates automatically in adulthood whenever their work is judged — no single episode is decisive, but each instance adds a small increment until the path fires on its own. The second is *single-event intensity*: under conditions of extreme arousal, stress hormones accelerate synaptic consolidation such that a single event can produce the encoding that would otherwise require thousands of repetitions [4, 35]. The nurse who experiences one overwhelming mass-casualty night encodes a danger path in a single exposure. The third

is *genetic and epigenetic inheritance*: predispositions and the consequences of ancestral experience can be transmitted, biasing the ease with which particular paths form [6] (Figure 3).

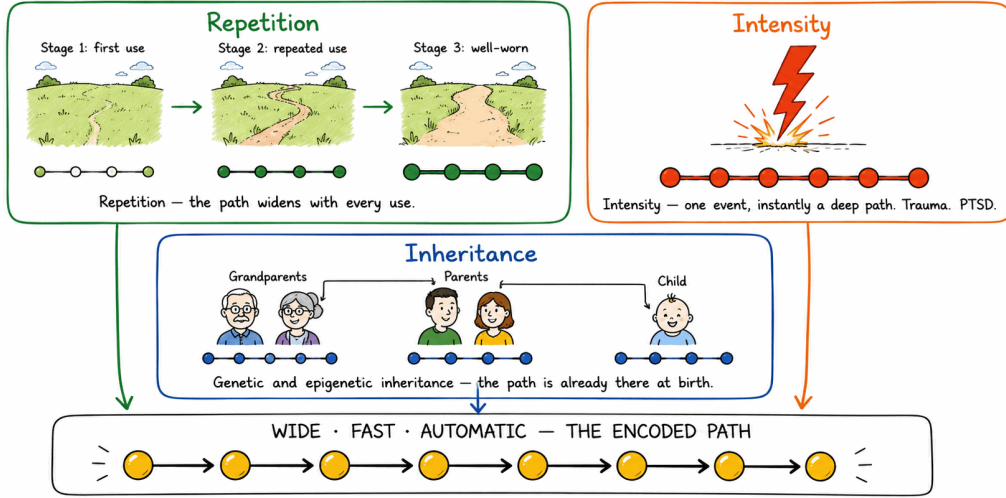


Figure 3: How neurone paths form. Three mechanisms encode paths into the brain's physical architecture: repetition (many small increments of synaptic strengthening over time), single-event intensity (one extreme event producing maximal encoding through stress-hormone-accelerated consolidation), and genetic or epigenetic inheritance (transmitted predispositions that bias path formation).

These mechanisms are not mutually exclusive; a given path may be inherited as a predisposition, established through repetition, and then intensified by a single event. What matters for the framework is that all three produce the same kind of physical structure, and all three are subject to the same mechanisms of deepening and dissolution described below.

3.3. Not All Paths Are Problems

The framework does not treat neurone paths as pathological in general. The overwhelming majority of encoded paths are adaptive or neutral: the skill of reading, the ability to ride a bicycle, the recognition of a loved one's face, and the capacity for language are all deeply encoded paths, and their automaticity is precisely what makes them useful. Deep encoding is not the problem; it is the basis of all learning and competence (Figure 4).

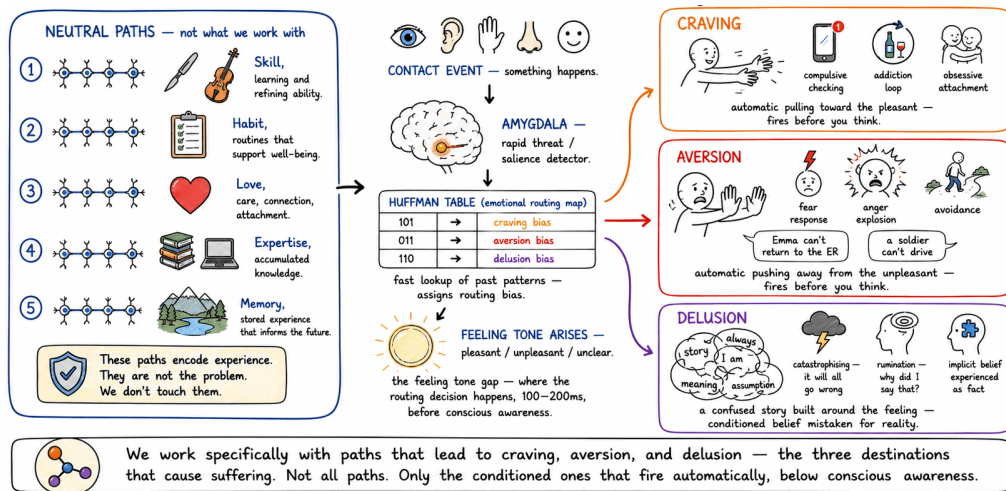


Figure 4: Not all paths are problems. The same physical mechanism of encoding underlies adaptive paths (skills, knowledge, recognition) and maladaptive reactive paths. Only paths that route automatically to reactive suffering — disproportionate, involuntary, and unwanted responses — are targets for intervention. The framework preserves adaptive encoding and targets only the specific reactive paths that generate suffering.

The target of intervention is a specific subset: paths that route automatically to reactive suffering — responses that are disproportionate to the situation, involuntary, and unwanted by the person who experiences them. This scoping is important both clinically and ethically: the aim is not to flatten affect or erase memory, but to dissolve the specific paths that produce reactions the person themselves wishes to be free of.

3.4. What Is a Situation? Situations Reach Us Through the Senses

Before describing how situations are encoded and matched, we must be precise about what a situation is. We propose that a situation is not an abstract whole but the specific set of sensory inputs through which an event reaches the brain. These inputs arrive through a closed set of channels: five *exteroceptive* channels that register the external world — sight, hearing, smell, taste, and touch — and at least one *internal* channel, memory, which registers self-generated content such as a surfacing recollection or a recalled voice. A felt inner state, such as the sense of being alone or unsupported, functions as a further internal channel. This closed-set formulation is deliberate: it converts an otherwise arbitrary list of situational features into a principled enumeration of the channels through which any situation can be

registered at all (Figure 5).

A clarification on terminology is warranted here. The brain does not index or retrieve whole “situations” as discrete objects; mechanistically, what propagates and is matched is a *pattern of sensory signals*, appraised rapidly and approximately by the amygdala and associated circuits against encoded patterns [8, 9]. We use “situation” throughout as a deliberate simplifying abstraction — a convenient name for that pattern of co-occurring sensory signals — because it makes the framework legible without altering its mechanistic content. When we speak of a situation being matched to a path, we mean that the pattern of sensory signals characterising that situation is what the matching operates on.

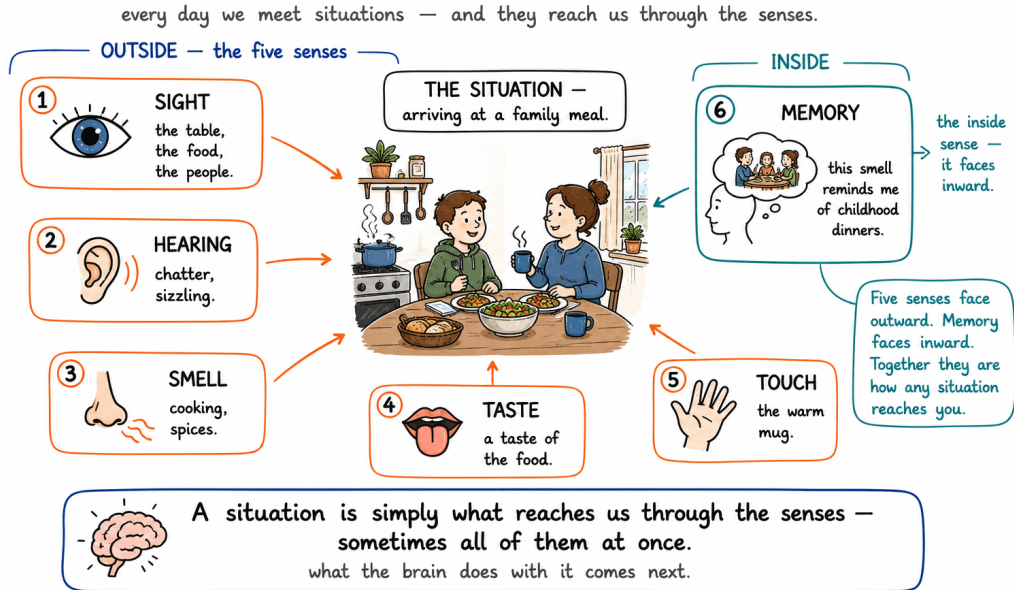


Figure 5: What a situation is. An everyday situation — here, arriving at a family meal — reaches the brain through a closed set of sensory channels: the five exteroceptive senses (sight, hearing, smell, taste, touch) and the internal channel of memory, which faces inward and can supply content with no external input at all. A situation is constituted by whichever channels are active, integrated together into a single recognised event.

Two consequences follow. First, a situation is composed only of the channels that are actually active; different situations recruit different subsets, and no situation requires all channels at once. Second, because memory is a genuine internal channel, a situation can be assembled partly or wholly from internally generated content — which is why a recollection alone, with the

external world entirely quiet, can constitute a situation capable of activating a path. This internal channel is central to the account of generalisation and trauma developed below. Our formulation here is a sensory-channel model grounded in multimodal integration; it is worth noting, as a historical antecedent rather than a foundation, that classical Buddhist analysis likewise treated the mind as a sixth sense base — alongside the five physical senses — whose objects are mental phenomena including memories [48, 49]. We do not develop this convergence here; it is treated separately in future work.

3.5. A Situation Links to a Neurone Path

A situation, once registered through its active channels, is matched against the brain’s store of encoded paths. When the pattern of sensory inputs — appraised rapidly by the amygdala — matches the signature of an encoded path, that path is activated. The situation is thus the *key*, and the neurone path is what the key unlocks. Crucially, it is the pattern of sensory channels — not the event in the abstract — that performs the matching, and it is the activated path, not the event, that determines the resulting experience and behaviour (Figure 6).

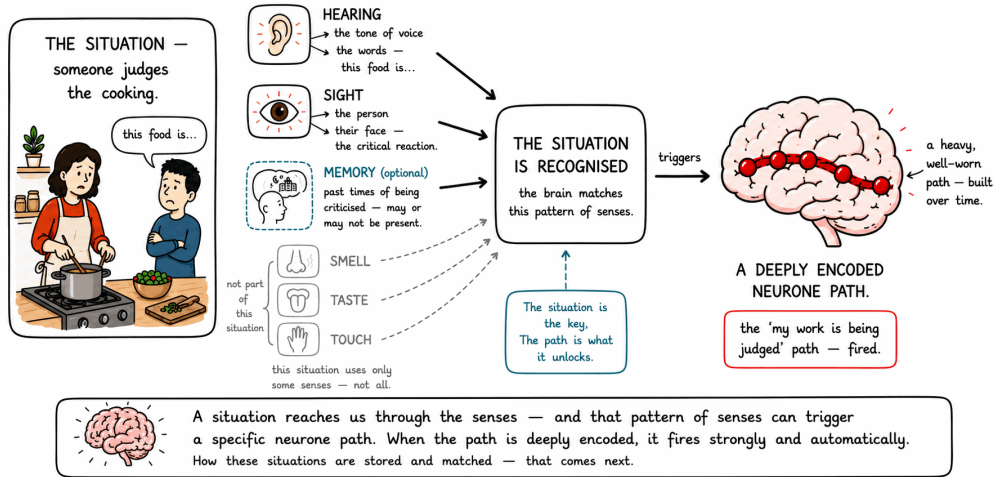


Figure 6: A situation links to a neurone path. A specific situation — someone commenting on a cook’s food — reaches the brain through only the channels that carry it: hearing (the tone and the words) and sight (the person and their facial reaction), with memory as an optional internal contributor. The other senses are not part of this situation. This pattern of active channels matches, and thereby fires, a deeply encoded neurone path — here, the implicit-belief path “my work is being judged.”

This figure also illustrates the point made above: only some channels carry a given situation. The cook’s situation is composed of hearing and sight; smell, taste, and touch play no part in it, even though food is present. The situational key is therefore composed of only the channels that carried the path’s formative attributes — a subset of the closed channel set, not the whole set. This subset structure is what makes the encoding model that follows both precise and tractable.

3.6. *Deeply Encoded Paths as Implicit Beliefs: A Situation-Keyed Encoding Model*

The most deeply encoded reactive paths function as *implicit beliefs*: automatic routing assumptions that determine how a situation is interpreted before any deliberate reasoning occurs. A person does not consciously decide that criticism means they are inadequate; the felt sense of inadequacy arrives first, in milliseconds, and conscious reasoning proceeds from it [10, 11]. The implicit belief is not a proposition the person holds; it is the path that fires.

We propose that this routing operates as a situation-keyed compressed encoding system, analogous to Huffman encoding in information theory [34]. In Huffman encoding, the most frequently occurring symbols are assigned the shortest codes, minimising the average length of the encoded message. In the brain’s affective architecture, the most frequently or intensely encoded paths become the shortest, fastest, most automatic routes: they fire first, at the lowest stimulus intensity, and below conscious awareness. The *key* to this lookup is the situation, decomposed — as established in Section 3.4 — into sense-tagged attributes across the closed channel set: what is heard (tone, words), seen (face, person, place), smelled, and internally registered (memory, inner state). The amygdala performs approximate pattern matching of these incoming attributes against encoded path signatures, and returns a pre-cognitive affective routing decision — a feeling tone — in under 200 milliseconds [8, 9] (Figure 7).

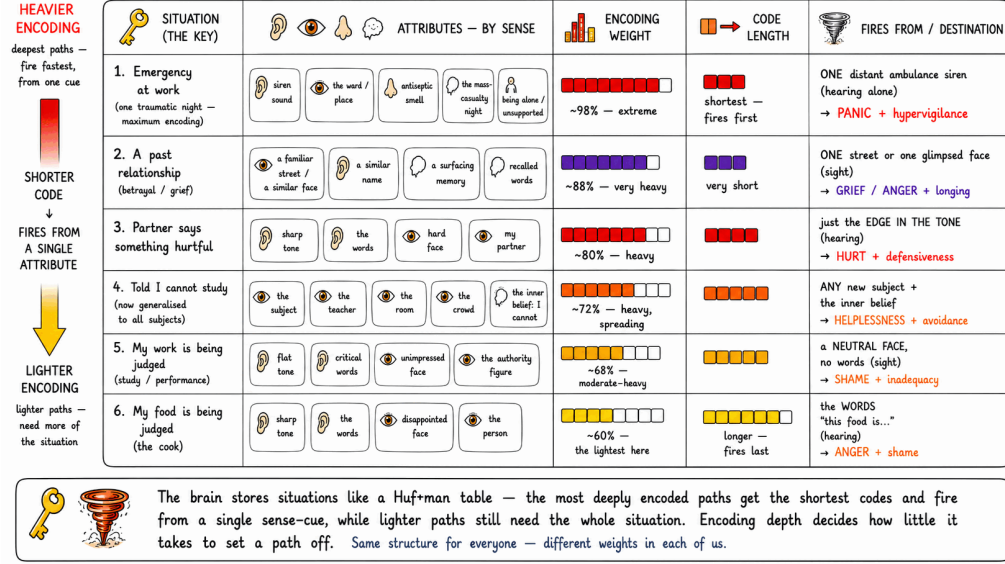


Figure 7: The situation-keyed encoding model. Each row represents a conditioned reactive path indexed by a situational key comprising attributes such as tone, facial expression, words, location, sound, and memory. Deeply encoded paths (top rows) carry the greatest weight and fire fastest and most automatically, analogous to the shortest codes in Huffman encoding. Incoming situations are matched approximately against these keys to generate a pre-cognitive affective routing decision in under 200 milliseconds.

This model extends existing accounts of rapid affective appraisal [8, 9, 11] by specifying the structure of the situational key and the compression principle that determines which paths dominate. It also sets up the central clinical phenomenon addressed next: the behaviour of these keys as encoding depth increases.

3.7. Encoding-Depth-Dependent Stimulus Generalisation: The Tornado Effect

We propose that the situational specificity required to activate a conditioned path is inversely proportional to its encoding depth. A thinly encoded path requires the full complement of original situational attributes to fire: the tone, the face, the words, and the location must all be present. As encoding depth increases, progressively fewer attributes are required, until a heavily encoded path fires from a single matching attribute. We term this *encoding-depth-dependent stimulus generalisation*, and offer the metaphor of a tornado to make it intuitive: a light breeze is felt only on direct contact, whereas a

tornado exerts its pull far beyond its visible boundary [50]. The deeper the encoding, the wider the path’s pull radius (Figure 8).

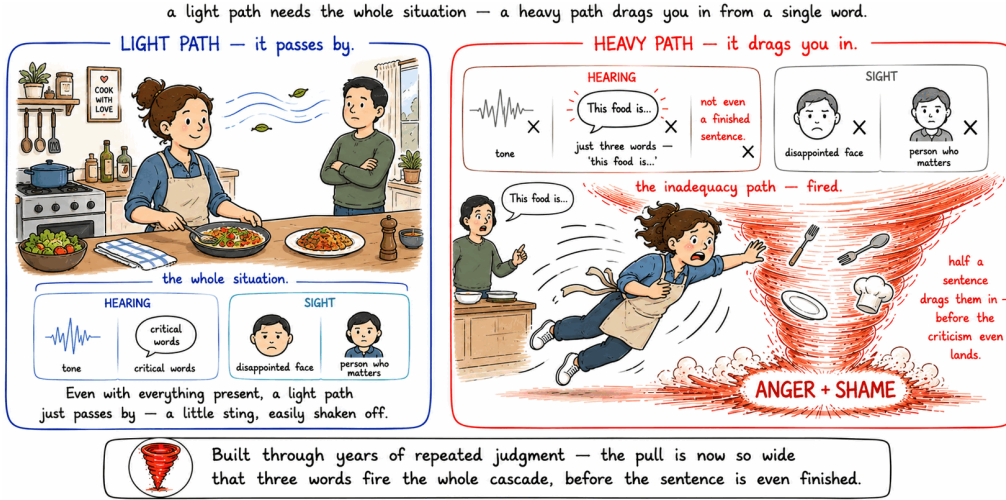


Figure 8: The tornado effect in a repetition-encoded path (cook scenario). A normally encoded path requires all situational attributes — tone, facial expression, the specific person, and critical words — to be present to fire (left). After deep encoding through years of repeated criticism, the path fires from a single attribute: the words “this food is. . .” alone activate the full reactive cascade, regardless of tone, speaker, or setting (right).

The same principle explains the hallmark clinical feature of trauma: activation from stimuli that share almost nothing with the original event. When a path is encoded at maximal intensity in a single event, its pull radius becomes extremely wide, and a single peripheral attribute — a sound, a smell — suffices to fire the entire cascade [4, 35] (Figure 9).

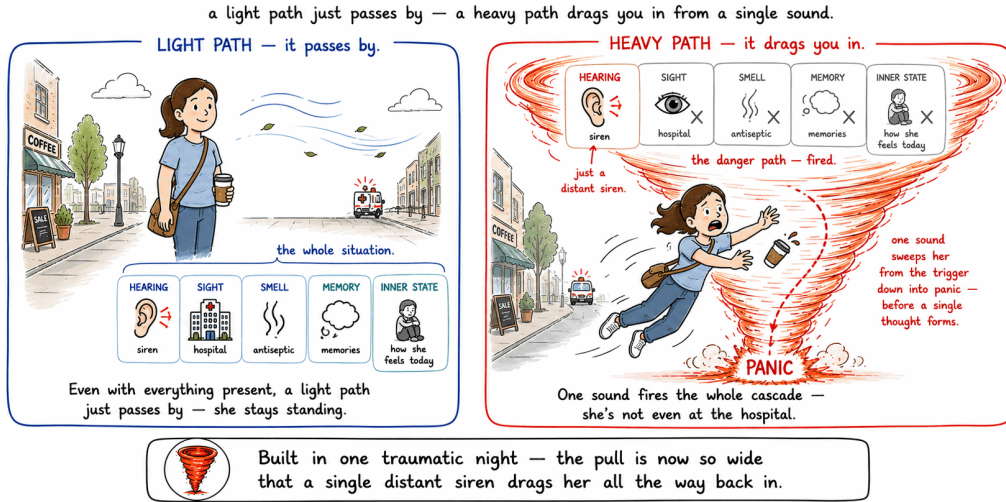


Figure 9: The tornado effect in an intensity-encoded path (hospital scenario). Following a single mass-casualty event encoded at maximal intensity, the danger path acquires an extremely wide pull radius. A distant ambulance siren — a single attribute, absent the original people, place, and circumstances — is sufficient to activate the full reactive cascade. This is the mechanistic basis of trauma responses triggering from apparently unrelated stimuli.

Encoding-depth-dependent stimulus generalisation reframes a puzzling clinical observation as a direct consequence of encoding depth, and predicts that reducing encoding depth — narrowing the path — should correspondingly narrow the pull radius, restoring the requirement for more complete situational matching before activation.

3.8. Affective Routing and the Feeling Tone Gap

We now specify the temporal structure of activation, because it is here that intervention becomes possible. Following situational registration, the amygdala generates a pre-cognitive affective signal — a bare feeling tone of pleasantness, unpleasantness, or neutrality — within approximately 100–200ms, prior to full prefrontal engagement [8, 9, 10]. This feeling tone is not yet an emotion, a thought, or a reaction; it is the bare affective valence that precedes and shapes them. The constructed emotional experience — the narrative of what is happening and what it means — is assembled afterward [11] (Figure 10).

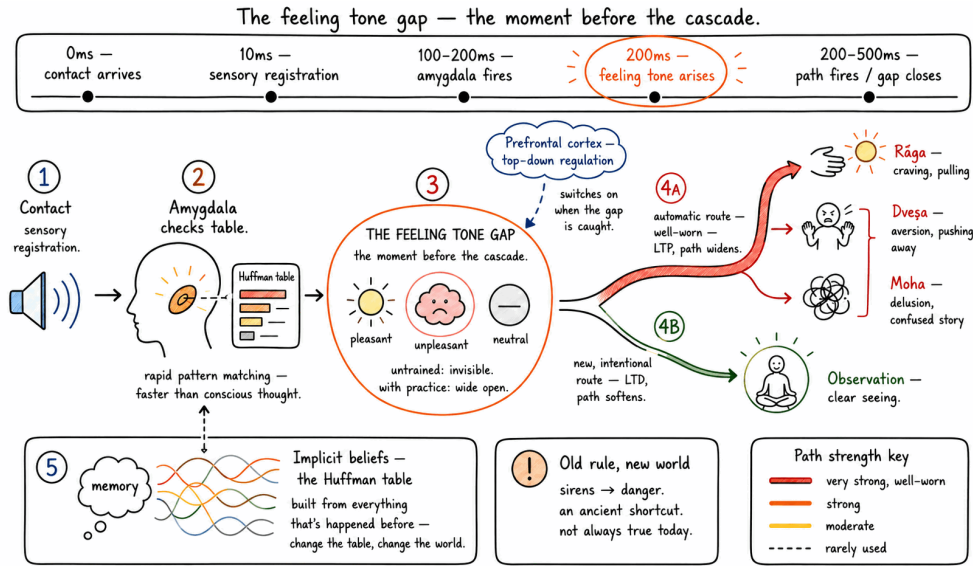


Figure 10: How the brain routes thoughts, and the feeling tone gap. A situation is registered and matched against encoded paths; the amygdala returns a pre-cognitive feeling tone within 100–200ms; and only afterward is the conscious thought and reactive behaviour constructed. The brief prefrontal-amygdala regulatory window between the feeling tone and the completion of the reactive cascade — the feeling tone gap — is the sole point at which the cascade can be intercepted upstream.

Between the arising of the feeling tone and the completion of the reactive cascade lies a brief prefrontal-amygdala regulatory window during which top-down modulation can intercept the cascade before it completes [20, 21, 22, 23]. We call this window the *feeling tone gap*. In untrained awareness it is effectively invisible: the feeling tone and the reaction appear as a single seamless event. The central proposal of the framework is that this gap is the only point at which upstream intervention — and therefore structural dissolution — is possible.

3.9. How Paths Deepen: Long-Term Potentiation

Each time a reactive path fires to completion, long-term potentiation operates on the synaptic connections along its length: AMPA receptors are inserted into the post-synaptic membrane, and the path is strengthened [7, 3, 26]. The path becomes wider, its activation threshold lower, its firing faster and more automatic, and its pull radius — following the tornado principle — wider. This is why reactive patterns worsen with repetition rather than

exhausting themselves, and why avoidance, far from relieving a fear, reliably intensifies it: each avoidance episode fires and therefore strengthens the very path it seeks to escape [13, 14] (Figure 11).

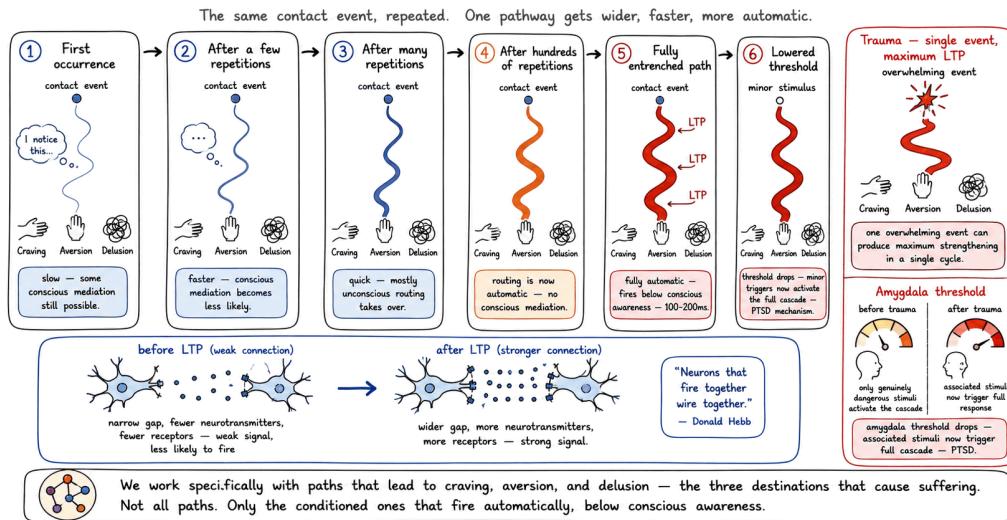


Figure 11: How paths deepen through long-term potentiation. Each completed firing of a reactive path strengthens its synaptic connections, widening the path, lowering its activation threshold, and increasing its firing speed and automaticity. Untreated reactive patterns deepen over time because every triggered episode functions as a strengthening event.

We illustrate path deepening in two contrasting cases, presented separately because they represent the two principal formation mechanisms. The first is the cook, whose inadequacy path was built by *repetition*. No single episode of criticism was decisive; instead, each instance added a small increment of long-term potentiation, and over years these increments accumulated into a deeply encoded path. Every subsequent completed firing — each adult episode in which criticism of the cook’s food fires the path to a full reaction — continues this deepening, widening the path and lowering its threshold further (Figure 12).

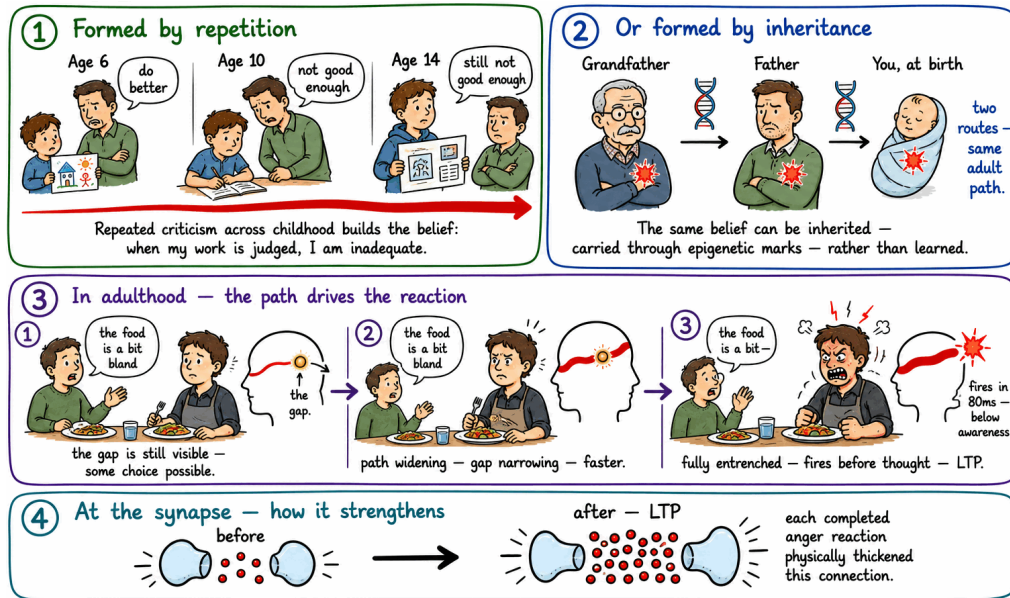


Figure 12: Path deepening through repetition (the cook). The inadequacy path is built incrementally: each episode of criticism adds a small increment of long-term potentiation, and over years these accumulate into a deeply encoded path. Every completed firing thereafter deepens it further, widening the path and lowering its activation threshold.

The second is the hospital case (Emma), whose danger path was built by *single-event intensity*. Here one overwhelming night produced, in a single exposure, the depth of encoding that repetition would take years to achieve — because extreme physiological arousal accelerates synaptic consolidation [4, 35]. Thereafter every triggered episode of the panic cascade operates as a further potentiation event, so that untreated the path deepens over time rather than fading (Figure 13).

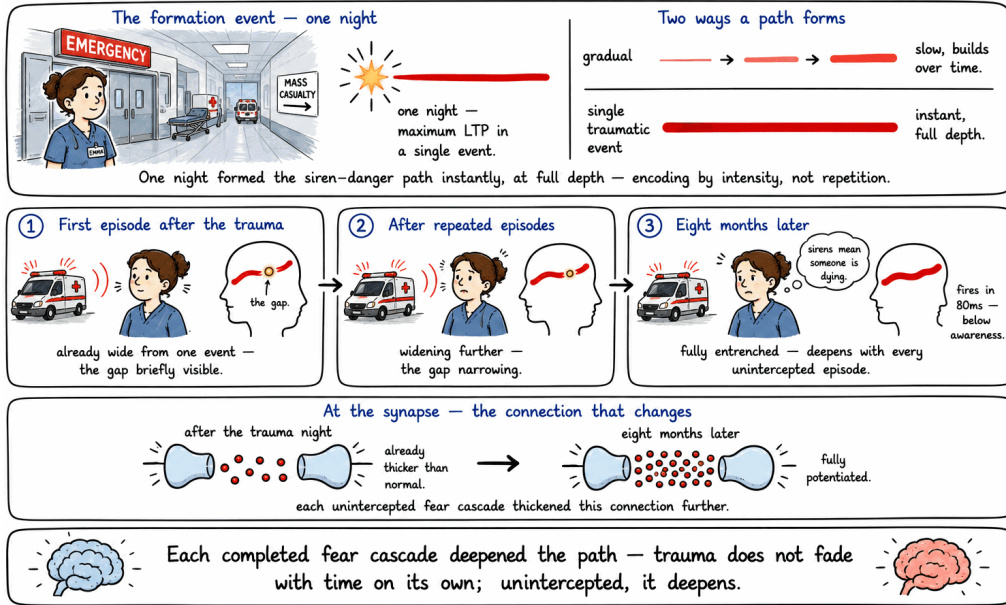


Figure 13: Path deepening through single-event intensity (the hospital case, Emma). A single night of extreme arousal produces, in one exposure, the encoding depth that repetition would require years to build. Each subsequent triggered episode of the panic cascade acts as a further potentiation event, deepening the path rather than exhausting it.

The two cases arrive at the same endpoint — a deeply encoded reactive path — by different temporal routes: many small increments in the cook, one large increment in Emma. The clinical implication is identical for both: any intervention which allows the path to continue firing to completion — however it manages the aftermath — leaves the deepening mechanism intact. Structural change requires interrupting the firing itself.

3.10. Which Paths Require Rewiring

Because most encoded paths are adaptive, the framework requires a principled criterion for identifying intervention targets. A path is a candidate for rewiring when its activation is disproportionate to the situation, involuntary, and productive of suffering the person wishes to be free of. Adaptive paths — skills, knowledge, meaningful emotional responses — are explicitly preserved (Figure 14).

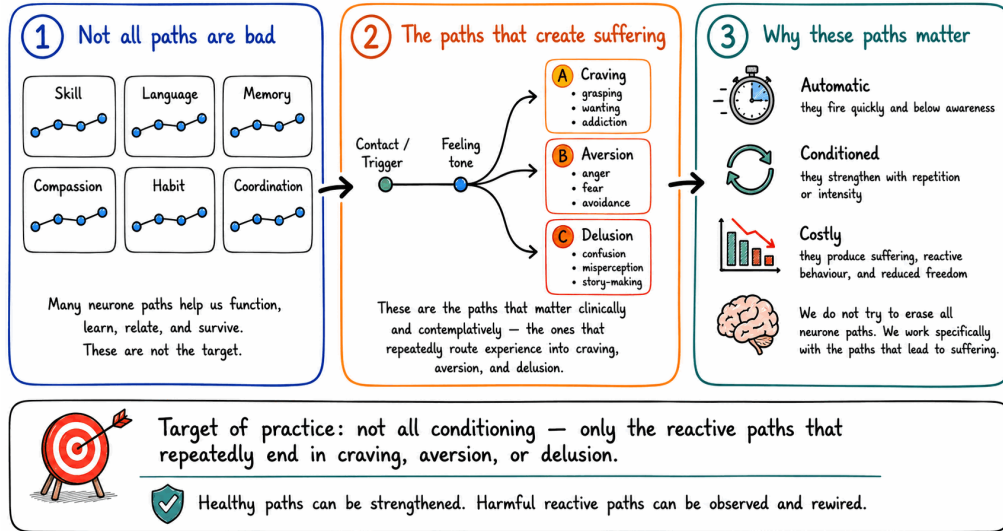


Figure 14: Which neurone paths require rewiring. Intervention targets are restricted to paths that route automatically to reactive suffering: disproportionate, involuntary, and unwanted responses. Adaptive and neutral paths are preserved. Identifying the specific target path — its situational key and reactive destination — is the first step of the NirWire intake process.

Identifying the specific target path, including its situational key and its characteristic reactive destination, is the first operational task of the NirWire system, described in Section 4.

3.11. How Paths Dissolve: Long-Term Depression and the Four Moments of Observation

Long-term depression is the synaptic mechanism opposite to potentiation: under conditions of activation without completion, AMPA receptors are internalised from the post-synaptic membrane, phosphatases activate, and the synaptic connection physically weakens [3, 26, 27, 51]. The framework’s central mechanistic claim is that engaging the feeling tone gap — observing the path as it begins to fire, rather than allowing it to complete — creates precisely these conditions. Each interception produces a small increment of LTD. Individually these increments are negligible; accumulated across hundreds of sessions, they produce measurable structural reorganisation, narrowing the path, raising its threshold, and reducing its pull radius.

The interception is achieved through structured observation, which we decompose into four moments: noticing the bare feeling tone as it arises;

recognising it as the activation of a path rather than a fact about the world; sustaining observation without allowing the cascade to complete; and witnessing the path subside un-acted-upon [12, 24, 25, 52]. This is the mechanism by which observation — so often described in general terms — is here given a specific synaptic interpretation (Figure 15).

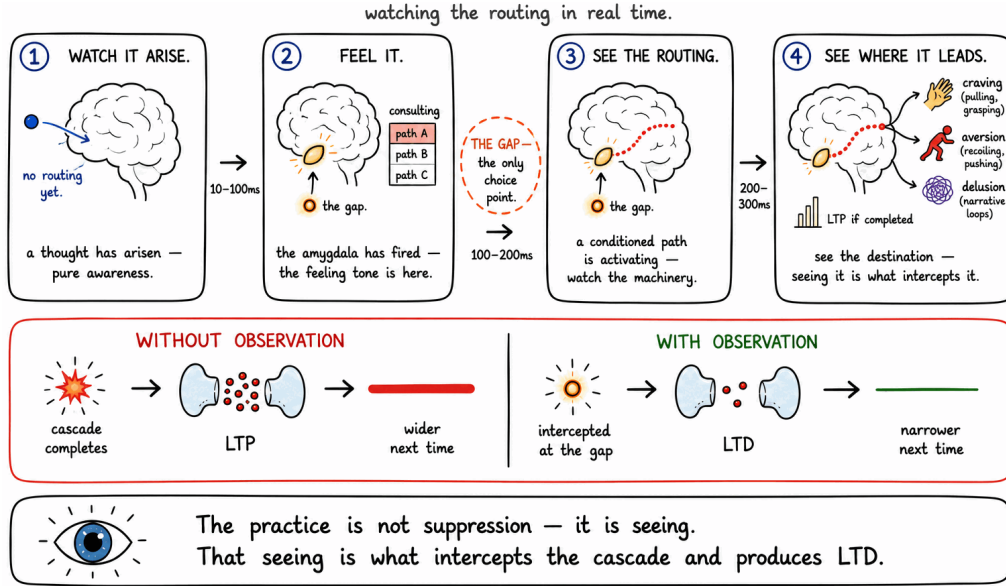


Figure 15: How paths dissolve through long-term depression, via the four moments of observation. Observation at the feeling tone gap intercepts the reactive cascade before completion: noticing the bare feeling tone, recognising it as path activation rather than fact, sustaining observation without completing the cascade, and witnessing the path subside. Each interception produces a small increment of long-term depression, weakening the path.

This mechanism defines the distinction between suppression and dissolution. Suppression manages the behavioural output of a cascade that has already fired to completion, leaving the path — and its potentiation — intact. Dissolution intercepts the cascade upstream, at the feeling tone gap, so that the path does not fire to completion and long-term depression operates instead. Only the latter produces structural change in the path itself.

A single interception changes little; the significance of the mechanism is cumulative. Across repeated sessions, two processes run in parallel. The reactive path, no longer permitted to fire to completion, undergoes incremental long-term depression and progressively weakens. At the same time, the act

of observation is itself a neural activity, and each time it is performed it is strengthened through the same Hebbian potentiation that builds any path — so the capacity to observe grows stronger with practice. Over many sessions these opposing trajectories cross: the observing capacity, initially weak, becomes stronger than the reactive path, which has narrowed. That this kind of repeated mental training produces measurable structural change in the brain is well established — experience-dependent grey-matter change follows sustained practice in domains from spatial navigation to juggling [53, 54], and contemplative training in particular is associated with structural and functional reorganisation in regions supporting attention and emotion regulation [55, 56]. The framework proposes that guided observation at the feeling tone gap engages precisely this capacity for structural reorganisation, directed specifically at the reactive path (Figure 16).

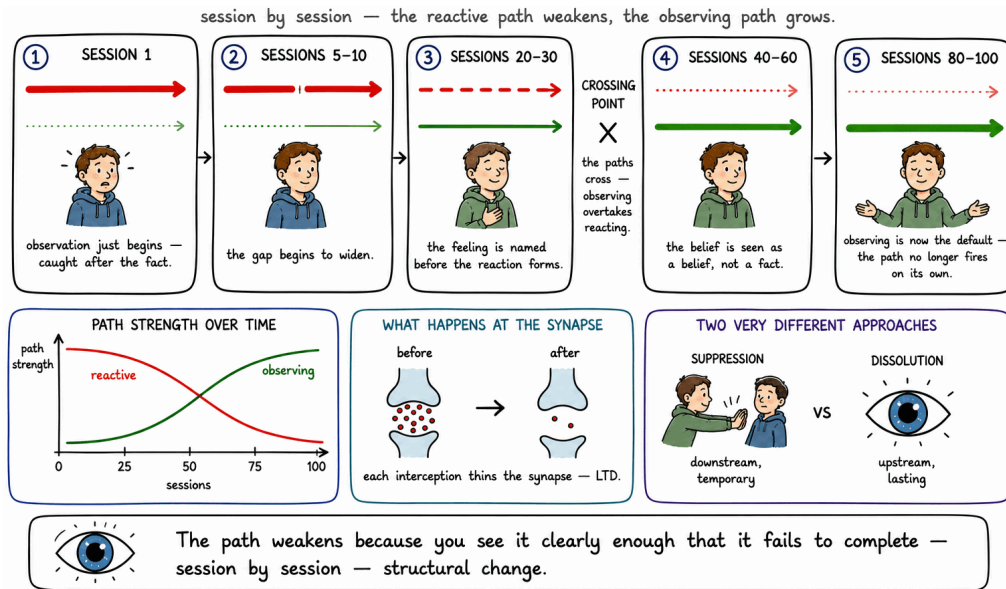


Figure 16: How observation rewires the brain across sessions. Repeated interception at the feeling tone gap drives two parallel trajectories: the reactive path weakens through accumulated long-term depression (its width and pull radius contracting session by session), while the observing capacity strengthens through potentiation each time it is exercised. Over many sessions the trajectories cross — the observing capacity overtakes the reactive path — after which the situation that once triggered the automatic cascade no longer does so. The structural change is cumulative and gradual, not a single event.

We illustrate this dissolution process in the same two cases used above.

For the cook, repeated observation of the “my work is being judged” path at the feeling tone gap — catching it as it begins to fire rather than completing the reaction — accumulates long-term depression across sessions, until food criticism no longer fires the full cascade automatically (Figure 17).

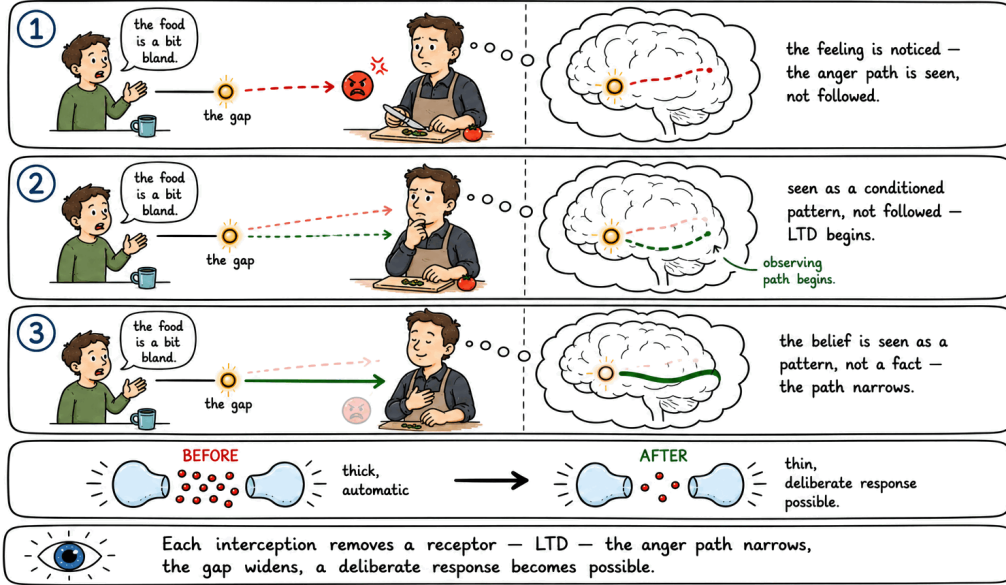


Figure 17: Path dissolution through observation (the cook). Repeated observation at the feeling tone gap — catching the “my work is being judged” path as it begins to fire, rather than completing the reaction — accumulates long-term depression across sessions. The path narrows, its threshold rises, and its pull radius contracts, until food criticism no longer fires the full reactive cascade automatically.

Because the mechanism is defined by interception at the gap rather than by how the path was originally encoded, it applies identically to the intensity-encoded case. For Emma, repeated observation of the danger path at the feeling tone gap — noticing the panic as it begins to arise, before the cascade completes — accumulates the same long-term depression, session by session, narrowing a path that was built in a single night (Figure 18).

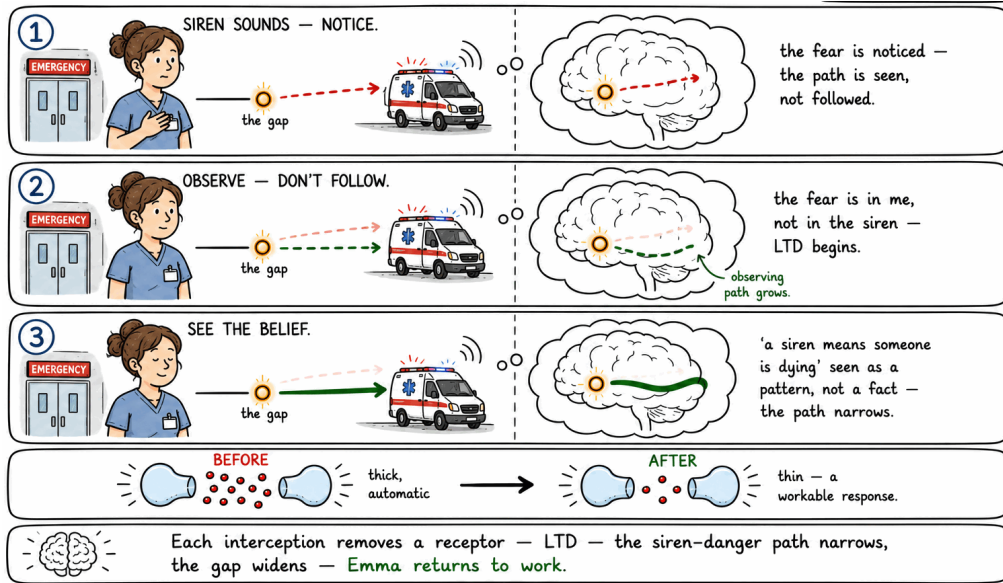


Figure 18: Path dissolution through observation (the hospital case, Emma). Though the danger path was encoded in a single intense event, it dissolves through the same mechanism as the cook’s: repeated interception at the feeling tone gap accumulates long-term depression, narrowing the path and contracting its pull radius, until a siren no longer fires the full panic cascade. The encoding history differs; the dissolution mechanism is identical.

Taken together, the components of this section constitute the end-to-end model that NirWire operationalises: reactive patterns are physical paths; paths are encoded by repetition, intensity, and inheritance; a situation reaches the brain through a closed set of sensory channels and, as a situation-keyed implicit belief, activates a path whose pull radius widens with encoding depth; the path fires through a pre-cognitive feeling tone that opens a brief regulatory gap; potentiation deepens it when it fires to completion; and long-term depression dissolves it when it is intercepted at the gap through structured observation. Section 4 describes the system that puts this model into practice.

4. The NirWire Framework

The theoretical account developed in Section 3 identifies a specific and previously unexploited intervention point: the feeling tone gap, at which a reactive path can be intercepted before it fires to completion, so that long-

term depression rather than long-term potentiation operates on it. NirWire is the system we propose to operationalise intervention at this point. This section describes its core reframing of the therapeutic relationship, its five functions, and its technical architecture.

4.1. Core Reframe: The User as Scientist of Their Own Brain

NirWire does not position the user as a patient receiving treatment. It positions them as an observer of their own neurological machinery — a scientist studying the reactive paths that fire within them. This reframing is not merely motivational; it follows directly from the mechanism. Dissolution requires the user to occupy the observing position at the feeling tone gap: to watch the path begin to fire rather than to be carried by it. A framework that positioned the user as a passive recipient of an intervention would be working against the very stance the mechanism depends upon. The role of the NirWire agent is therefore not to treat the user but to teach them to see clearly enough, and at the right moment, that the rewiring happens as a natural consequence of sustained observation (Figure 19).

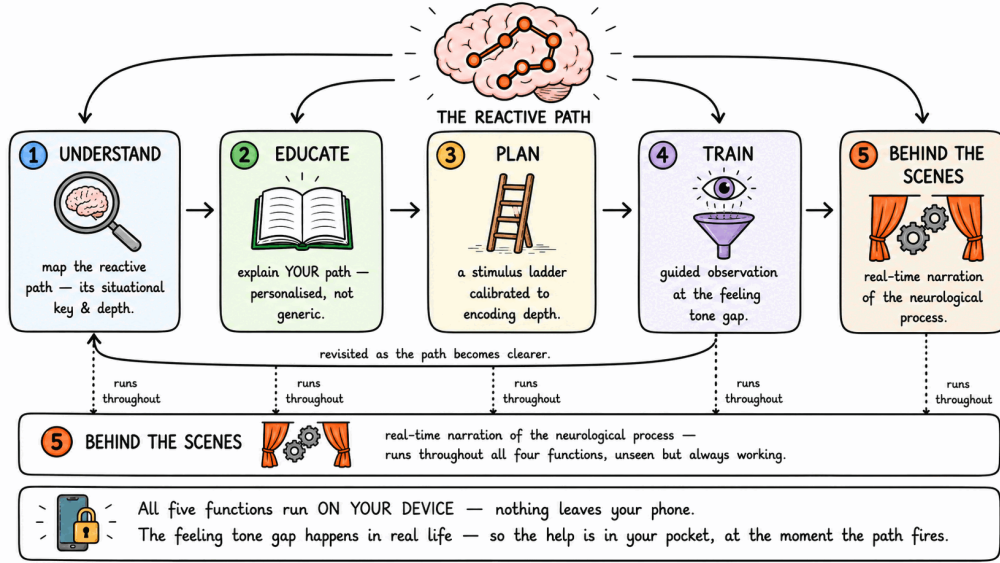


Figure 19: The NirWire framework at a glance. Five functions operate on the reactive path identified in the theoretical model: Understand maps the person’s reactive architecture; Educate delivers personalised psychoeducation; Plan constructs a calibrated training programme; Train guides observation at the feeling tone gap; and Behind the Scenes narrates the underlying neurological process in real time. All functions run on-device, with no data leaving the user’s hardware.

The five functions described below are not a rigid sequence but a set of capabilities that the agent draws on as the user progresses. In practice a user moves from understanding to education to planning to training, but the Behind the Scenes narration operates throughout, and the agent revisits earlier functions as the reactive architecture becomes clearer.

4.2. Function 1 — Understand: Mapping the Reactive Architecture

The first function is an intake conversation that maps the specific reactive path to be addressed. Drawing on the situation-keyed model of Section 3.6, the agent works with the user to identify the situational key of their reactive pattern: which sensory attributes are present when the reaction fires — what is heard, seen, smelled, or internally recalled — and which single attributes are sufficient to trigger it, revealing the path’s encoding depth. It establishes the reactive destination the path routes to, the path’s likely formation history (repetition, single-event intensity, or inherited predisposition), and the current activation intensity. The output is a personalised map of the reactive

architecture: the target path, its situational key, its encoding depth, and its starting level (Figure 20).

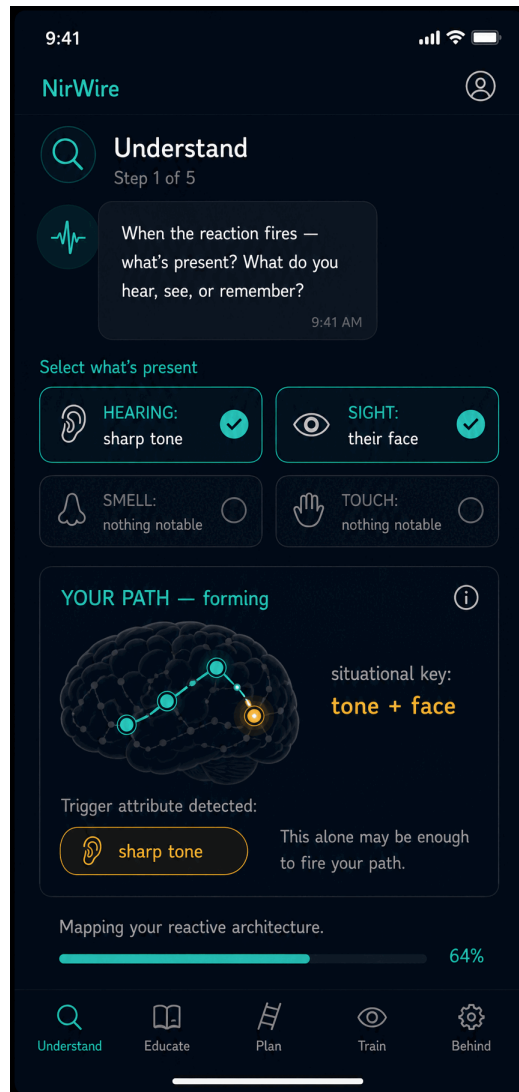


Figure 20: Function 1 — Understand. The intake conversation elicits the situational key of the reactive pattern, identifying which sensory attributes are present and which single attribute is sufficient to trigger the path.

Critically, this mapping precedes education. NirWire does not begin with generic psychoeducation; it first learns the specific architecture of the indi-

vidual’s path, so that everything subsequent is tailored to that person’s brain rather than delivered as general information.

4.3. Function 2 — Educate: Personalised Psychoeducation

Once the reactive architecture is mapped, the agent delivers psychoeducation built around the user’s own path. Rather than explaining reactive patterns in the abstract, it explains *their* pattern: what their specific path is, how it was likely encoded, why the reaction feels involuntary (because it completes below the threshold of conscious awareness), and what dissolution would consist of [57, 58]. The theoretical model of Section 3 is delivered here not as a lecture but as an account of the user’s own experience, which both motivates the training and equips the user with the conceptual vocabulary — feeling tone, the gap, the path, dissolution — that the training will require (Figure 21).



Figure 21: Function 2 — Educate. Psychoeducation is personalised to the user’s own mapped path — what it is, how it was encoded, and why it fires automatically — rather than delivered as generic information.

4.4. Function 3 — Plan: The Calibrated Stimulus Ladder

The Plan function constructs a personalised training programme centred on a graded stimulus ladder. Because the encoding-depth model (Section 3.7) predicts that a heavily encoded path fires from a single attribute

while a lighter one requires the full situation, the ladder is calibrated to the path's encoding depth: it begins with contact events mild enough to open the feeling tone gap without overwhelming the user's capacity to observe, and progresses toward fuller situational matches as observation stabilises. The programme specifies short daily sessions and longer weekly sessions, with a clear advancement criterion — stable observation across several consecutive sessions before the stimulus intensity is increased. This graded structure is consistent with established principles of inhibitory learning during exposure [14], while differing in its target: the ladder here serves observation at the gap, not extinction through repeated completion (Figure 22).

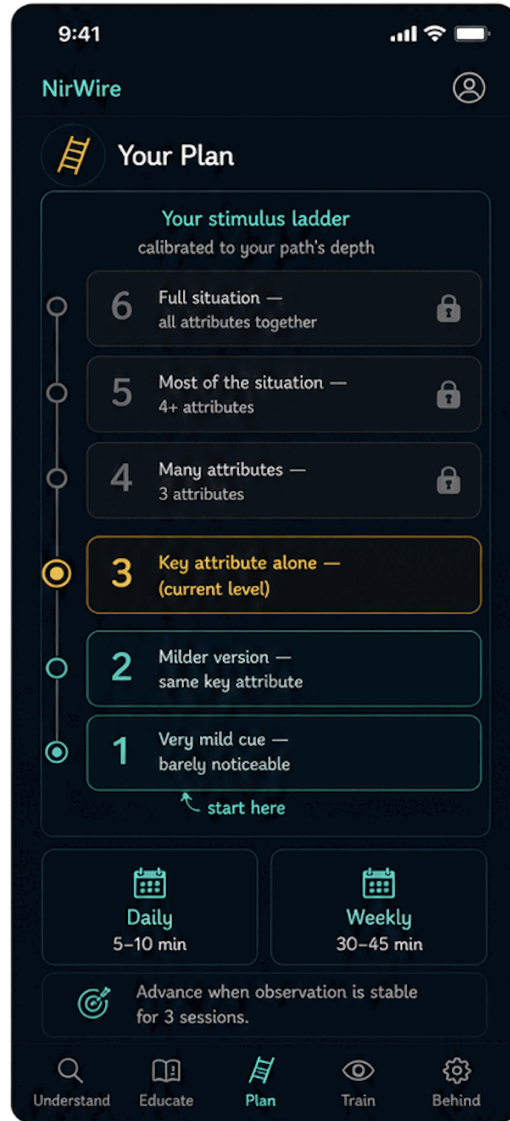


Figure 22: Function 3 — Plan. A stimulus ladder calibrated to the path’s encoding depth, with daily and weekly sessions and an advancement criterion based on stable observation rather than elapsed time.

4.5. Function 4 — Train: Guided Observation at the Feeling Tone Gap

The Train function is where dissolution occurs. In a training session, the agent generates a calibrated contact event at the user’s current ladder level and guides them through the four moments of observation established

in Section 3.11: noticing the bare feeling tone as it arises, recognising it as the activation of a path rather than a fact about the world, sustaining observation without allowing the cascade to complete, and witnessing the path subside un-acted-upon. Each such interception is a small increment of long-term depression. The guided structure sustains the observing position that untrained awareness cannot easily hold, and the calibration ensures the contact is strong enough to open the gap but not so strong that the cascade completes before observation can engage (Figure 23).

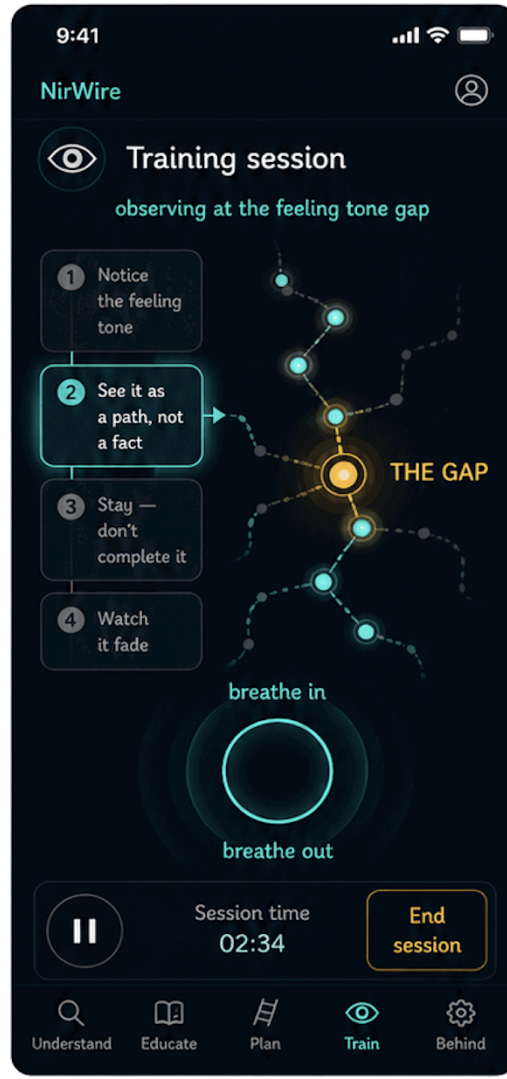


Figure 23: Function 4 — Train. A calibrated contact event is generated and the agent guides the user through the four moments of observation at the feeling tone gap, sustaining the observing position through which long-term depression accumulates.

Over repeated sessions, the cumulative process described in Section 3.11 unfolds: the reactive path weakens through accumulated depression while the observing capacity strengthens through practice, until the two trajectories cross and the path no longer fires automatically to the situations that once triggered it.

4.6. Function 5 — Behind the Scenes: Real-Time Neurological Narration

Running alongside the training is a narration function that names the neurological process as it occurs: identifying the moment the situational key is matched, the feeling tone arising, the path beginning to activate, and the gap opening. This narration serves two purposes. It sustains the observing position by giving the user a concrete, mechanistic frame for what they are experiencing — reinforcing that they are watching a path fire rather than confronting a fact — and it deepens metacognitive engagement, which the literature associates with the decentring capacity that the training seeks to build [12, 24, 25]. The narration is what makes the framework’s name literal: the user is shown the wiring behind their own reactions (Figure 24).

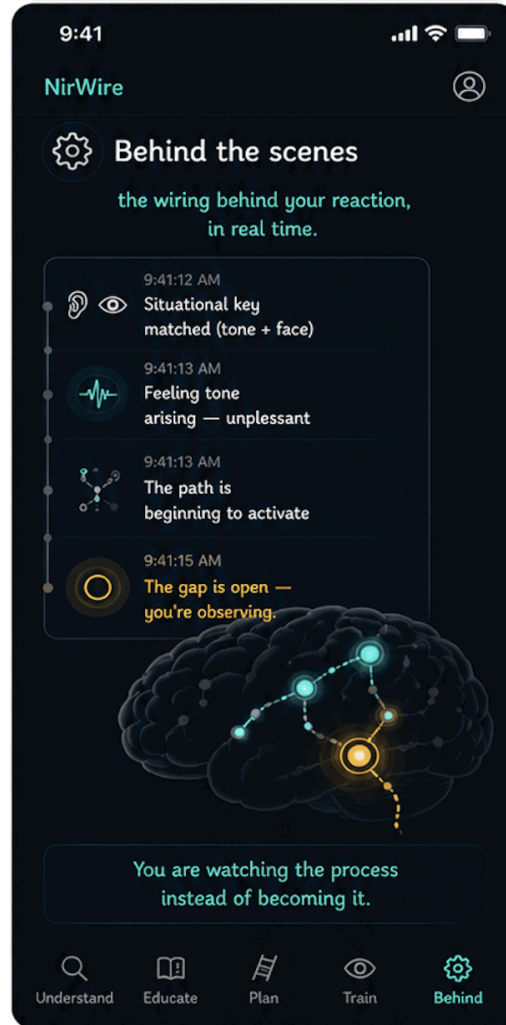


Figure 24: Function 5 — Behind the Scenes. Real-time narration names the neurological process as it unfolds — the situational match, the feeling tone, the path activating, the gap opening — sustaining the observing position and deepening metacognitive engagement.

4.7. Technical Architecture

NirWire is designed for strictly on-device deployment. All inference runs on the user’s own hardware, and session content — which is necessarily intimate, comprising the user’s most sensitive reactive patterns — never leaves the device. There is no server tier and no routine data egress [58]. This privacy-preserving design responds both to the sensitivity of the material

and to the documented importance of trust and data security in the adoption of digital mental-health tools [30, 31, 32], and it makes the framework deployable in settings where external transmission is prohibited or inadvisable. Where clinical hand-off is desired, it is achieved through encrypted, user-controlled export rather than through any live connection.

The proximity that on-device deployment affords is not merely a privacy consideration but a mechanistic one. Because the feeling tone gap occurs in the moment of activation — in the situations of ordinary life rather than in a scheduled clinical hour — a system carried on the user’s own device can support observation at or near the moment a path actually fires, providing a proximity to the intervention site that periodic clinical contact cannot match [30, 31, 59].

A two-tier inference design

Running large language models on a phone is constrained by memory, thermal, and battery budgets, and running several concurrently is infeasible for sustained use. NirWire therefore adopts a two-tier design that separates an always-on lightweight default from an optional high-accuracy mode that the user explicitly enables (Figure 25).

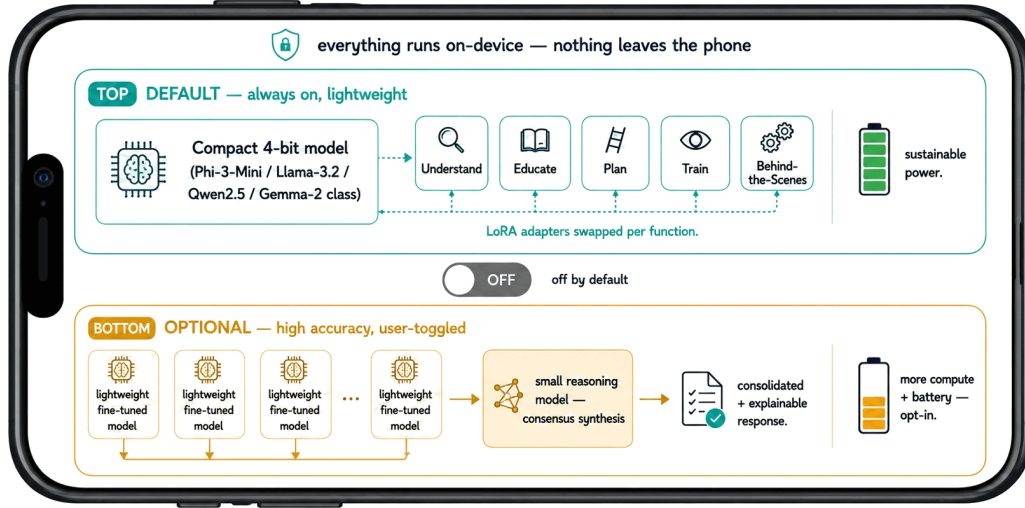


Figure 25: The two-tier on-device architecture. The default tier runs a single compact quantised model that handles all five functions within a phone’s battery and thermal budget. An optional, user-toggled high-accuracy tier runs a consortium of lightweight fine-tuned models whose outputs are reconciled by a small reasoning model through consensus-driven synthesis, yielding more robust and explainable responses at a higher compute and battery cost. Both tiers run entirely on-device; nothing leaves the phone.

The **default tier** runs a single compact, instruction-tuned model quantised to 4-bit precision — a model in the small on-device class such as Phi-3-Mini [47], Llama-3.2-1B/3B [60], Qwen2.5-1.5B/3B, or Gemma-2-2B. A single such model, with function-specific behaviour supplied by lightweight swappable adapters (below), delivers the full five functions within a sustainable power envelope and is the always-on experience for the ordinary user.

The **optional high-accuracy tier** is a user-toggled mode intended for moments where robustness matters most — for example, the initial Understand mapping, or a difficult training session. When enabled, it runs a *consortium* of several lightweight fine-tuned models whose independent outputs are reconciled by a small reasoning-oriented model through consensus-driven synthesis, an approach we have developed and validated technically in prior work on high-stakes decision support [61, 62, 63]. Because this mode holds multiple models in memory and performs additional reasoning passes, it is materially more compute- and battery-intensive than the default; it is therefore off by default, clearly surfaced as a higher-cost option, and left entirely under the user’s control. The heavier consortia used in server-class deployments of this method are not used here; NirWire’s consortium is composed

of lightweight models chosen to remain on-device [64].

Consensus-driven reasoning for responsible, explainable output

The purpose of the consortium is not merely accuracy but responsibility and explainability, which are essential in a sensitive mental-health context. Rather than relying on a single model’s response, the high-accuracy tier gathers independent outputs from several specialised models and passes them to a reasoning model that identifies agreement and disagreement, reconciles them, and produces a consolidated response accompanied by an interpretable rationale [62, 65]. This consensus-driven process reduces single-model bias, mitigates overconfident or spurious outputs, and exposes the reasoning behind a response rather than presenting it as an opaque verdict — properties that matter especially when the system is narrating a user’s own neurological process back to them (Figure 26).

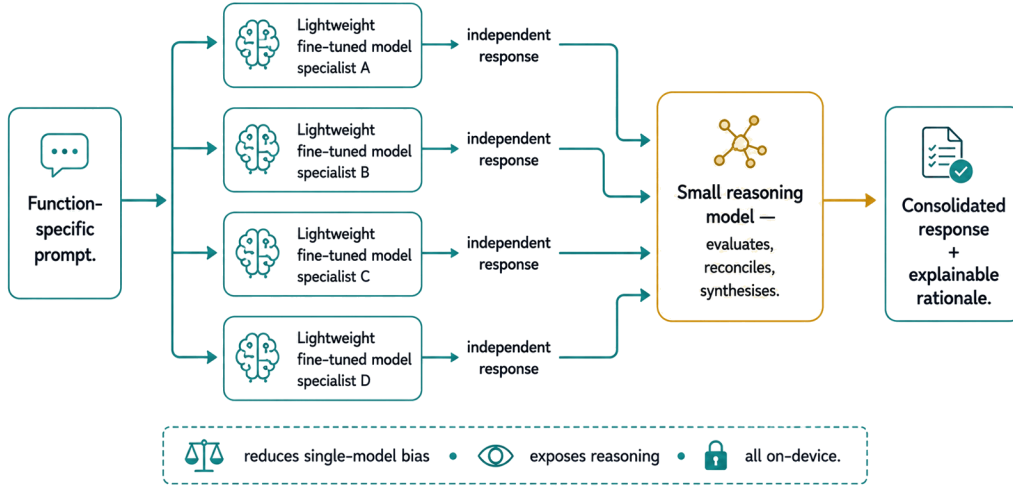


Figure 26: Consensus-driven reasoning in the optional high-accuracy tier. A function-specific prompt is dispatched to several lightweight fine-tuned models, each producing an independent response. A small reasoning model then evaluates, reconciles, and synthesises these into a single consolidated output with an explicit rationale, reducing single-model bias and supporting explainable, responsible decision support. All components are lightweight and run on-device.

Function specialisation through parameter-efficient fine-tuning

The five functions are specialised through parameter-efficient fine-tuning rather than through separate full models. Using Low-Rank Adaptation

(LoRA) [66] with 4-bit quantisation (QLoRA) [67], a small adapter is trained for each function — intake mapping, personalised psychoeducation, ladder construction, observation guidance, and neurological narration — on curated function-specific data. Because each adapter is a small fraction of the base model’s parameters, adapters can be stored compactly on-device and swapped as the conversation moves between functions, so the default tier can present function-specialised behaviour while keeping a single base model resident in memory [68]. Fine-tuning is performed offline; the deployed system’s behaviour is fixed during use, consistent with on-device and air-gapped operation (Figure 27).

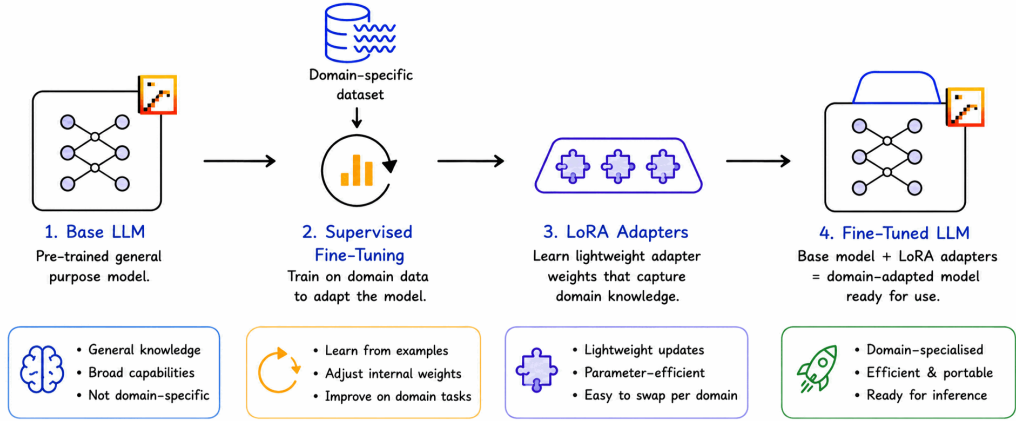


Figure 27: Overview of the supervised fine-tuning pipeline used to adapt a base LLM to domain-specific tasks. A curated, domain-specific dataset is used to perform supervised fine-tuning of a base LLM, with Low-Rank Adapter (LoRA) modules applied to enable parameter-efficient training. The resulting fine-tuned model retains the general linguistic capabilities of the base model while incorporating domain-specific knowledge, supporting efficient deployment and updates in resource-constrained and air-gapped environments.

Implementation status

A functional prototype of the framework has been built, and its technical components have been implemented and tested: the fine-tuning pipeline has been exercised, function-specific adapters produce appropriately structured outputs, the consensus-driven reasoning tier reconciles multiple model outputs into consolidated responses with rationales, and on-device inference has been verified as feasible within mobile resource constraints. We emphasise that this is technical validation of the system, not clinical validation of its

effect: the framework’s therapeutic efficacy has not been evaluated in a clinical study, and such evaluation is reserved for the future directions described in Section 5.

5. Discussion

We have presented NirWire as a theoretical framework and an implemented system: an end-to-end account of how conditioned reactive patterns arise as physical neurone paths, and a conversational AI architecture that targets those paths at the feeling tone gap. In this section we position the framework relative to existing treatments, state its limitations plainly, and outline the directions in which it should be tested and extended.

5.1. Positioning Relative to Existing Treatments

NirWire is not proposed as a replacement for established treatments but as a complement to them. Cognitive behavioural therapy, prolonged exposure, EMDR, and pharmacotherapy are clinically validated and effective, and nothing in the framework disputes their value. The claim is narrower and specific: these interventions operate downstream of the reactive cascade — restructuring the narrative, extinguishing the conditioned association, or modulating neurotransmission after the path has fired — whereas NirWire targets the upstream moment, before the cascade completes, at which long-term depression rather than long-term potentiation operates on the path. These are not competing mechanisms but complementary ones. A person may restructure the appraisal that follows a reaction (downstream) while also learning to intercept the path that generates it (upstream), and there is no reason the two cannot proceed together. Indeed, the framework suggests that downstream interventions may become more effective when the path generating the cascade is simultaneously being weakened at its source, since a narrower path produces a less forceful cascade to manage.

This complementarity also clarifies what NirWire is not. It is not a crisis service, and it does not perform diagnosis or treatment in the clinical sense. It is a training system for a specific skill — observation at the feeling tone gap — whose repeated exercise the framework proposes as a route to structural change. Where clinical care is needed, NirWire is intended to sit alongside it, with hand-off supported through user-controlled export.

5.2. *Relationship to Contemplative Practice*

The observational mechanism at the centre of the framework has a long history outside neuroscience. Contemplative traditions have described sustained, non-reactive observation of arising experience for millennia, and modern mindfulness-based clinical approaches have adapted these practices and demonstrated their effects on relapse and emotion regulation [52, 12, 55, 56]. NirWire’s contribution is not to introduce observation as such but to give it a specific mechanistic target and a calibrated, personalised delivery: rather than general open awareness, the system directs observation to a particular identified path, at the particular moment it fires, calibrated to that path’s encoding depth, and narrated in terms of the underlying process. In this sense NirWire can be read as an attempt to operationalise the mechanism that contemplative practice discovered, and to make it accessible without requiring the extended training that traditional practice demands. The deeper correspondence between the framework’s account and contemplative theories of mind is substantial enough to warrant separate treatment and is not developed here.

5.3. *Limitations*

The most important limitation is that this is a theoretical framework paper accompanied by a technically implemented system, not a clinical outcomes study. The framework’s central therapeutic claim — that guided observation at the feeling tone gap produces cumulative long-term depression of reactive paths and thereby lasting change — is grounded in established synaptic and affective neuroscience, but it has not been validated in a clinical trial. We have verified that the system can be built and that its components function as intended; we have not shown that using it improves clinical outcomes. That question is open, and the framework should be read as a set of testable proposals rather than as demonstrated results.

Several further limitations follow. First, the two central constructs we introduce — the situation-keyed Huffman encoding model and encoding-depth-dependent stimulus generalisation (the tornado effect) — are conceptual models, offered as analogies that organise and predict phenomena, not as literal descriptions of neural data structures. Their value lies in the predictions they generate and the interventions they motivate, and those predictions require independent empirical test. Second, the mechanistic account links observation at the feeling tone gap to long-term depression by drawing on established properties of these processes, but the specific claim that

metacognitive interception produces measurable LTD of an identified reactive path has not been directly demonstrated and would require targeted neuroscientific study. Third, the framework depends on the user’s capacity to occupy and sustain the observing position; individuals differ in this capacity, and it is not yet known for whom, and under what conditions, the training is effective or appropriate. Fourth, the practical demands of on-device operation impose real constraints on model capability, and the optional high-accuracy mode carries a compute and battery cost that limits how often it can be used.

5.4. Future Directions

The clearest next step is clinical evaluation. The framework generates specific, testable predictions — that session-by-session activation intensity should decline as training proceeds, that this decline should be more durable than the symptom management produced by suppression-based strategies, and that reductions should generalise in the manner the encoding-depth account predicts. These can be examined in controlled studies across the domains the framework addresses, from post-traumatic stress to everyday reactive patterns, with standardised outcome measures and appropriate comparison conditions. Because the framework positions itself as complementary to existing care, the natural design compares standard care with standard care augmented by NirWire, rather than setting the two in opposition.

Beyond outcome trials, three lines of work would strengthen the framework. Neuroimaging and electrophysiological studies could test the core mechanistic claim directly, seeking measurable correlates of path weakening and of the strengthening observing capacity across training. Measurement work could establish session-level behavioural markers — activation intensity, latency, and generalisation breadth — as tractable proxies for the underlying structural change, since these are observable without imaging and could support both research and adaptive personalisation within the system. And design work could refine the calibration of the stimulus ladder and the observation scaffold to individual differences in observing capacity, which the limitations above identify as a key unknown. Together these directions would move the framework from a grounded proposal toward an empirically validated intervention.

6. Conclusion and Future Work

This paper set out to do two things: to develop a complete, end-to-end account of how conditioned reactive patterns arise as physical structures in the brain, and to propose a system that acts on those structures at the one point where their dissolution becomes possible. The account runs from the single neurone and synapse, through the neurone path, to the deeply encoded paths that function as implicit beliefs; through the sensory channels by which a situation reaches the brain and the situation-keyed, compressed matching that fires a path in under 200 milliseconds; to the opposing synaptic mechanisms of long-term potentiation, which deepens a path each time it fires to completion, and long-term depression, which weakens it when the cascade is intercepted upstream. The intervention point that this account identifies — the feeling tone gap, the brief window between the arising of a pre-cognitive feeling tone and the completion of the reactive cascade — is the site at which NirWire operates.

Around this account we proposed three contributions: a situation-keyed compressed encoding model of implicit belief, analogous to Huffman coding; encoding-depth-dependent stimulus generalisation, the tornado effect, which explains why deeply encoded paths fire from a single fragment of a situation and generalise far beyond their origin; and a five-function conversational AI architecture — Understand, Educate, Plan, Train, and Behind the Scenes — that operationalises guided observation at the feeling tone gap. We described how this system is realised on-device, with a lightweight always-on model and an optional, user-controlled consensus-driven consortium for higher-accuracy moments, so that a person’s most sensitive material never leaves their own hardware. The distinction that runs through the whole framework is between *suppression*, which manages the output of a reaction that has already fired, and *dissolution*, which weakens the path that produces it: existing interventions largely do the former, and NirWire is designed to add the latter.

We have been careful to state what this work is and is not. It is a theoretical framework, accompanied by a system that has been built and technically tested; it is not a demonstration of clinical efficacy. The central therapeutic claim — that repeated observation at the feeling tone gap produces cumulative long-term depression of reactive paths, and thereby lasting change — is grounded in established neuroscience but remains to be validated in people. The value of the framework, for now, lies in the clarity of the mechanism it proposes and the specific, testable predictions that follow from it.

Those predictions define the future work. The most immediate need is clinical evaluation: controlled studies, across domains from post-traumatic stress to everyday reactive patterns, comparing standard care with standard care augmented by NirWire, and tracking whether session-by-session activation intensity declines, whether that decline is more durable than suppression-based management, and whether improvements generalise as the encoding-depth account predicts. Beyond outcomes, three lines of work would strengthen the framework: neuroimaging and electrophysiological studies to test the mechanistic claim directly, by seeking measurable correlates of a weakening reactive path and a strengthening observing capacity; measurement work to establish session-level behavioural markers — activation intensity, latency, and generalisation breadth — as practical proxies for structural change and as signals for adaptive personalisation; and design work to calibrate the stimulus ladder and observation scaffold to individual differences in the capacity to observe, which remains a key unknown. A separate strand, noted only briefly here, concerns the correspondence between this framework and long-standing contemplative accounts of mind, which merits dedicated treatment in its own right.

The brain becomes what it repeats. Every reactive pattern is a path that was built, and what was built can be weakened. NirWire is a proposal for how to do that deliberately — by learning to see a path clearly, at the moment it begins to fire, often enough that it no longer runs on its own.

References

- [1] J. E. LeDoux, *Anxious: Using the Brain to Understand and Treat Fear and Anxiety*, Viking, New York, 2015.
- [2] P. Fries, A mechanism for cognitive dynamics: neuronal communication through neuronal coherence, *Trends in Cognitive Sciences* 9 (10) (2005) 474–480. doi:10.1016/j.tics.2005.08.011.
- [3] R. C. Malenka, M. F. Bear, LTP and LTD: An embarrassment of riches, *Neuron* 44 (1) (2004) 5–21. doi:10.1016/j.neuron.2004.09.012.
- [4] R. K. Pitman, A. M. Rasmusson, K. C. Koenen, L. M. Shin, S. P. Orr, M. W. Gilbertson, M. R. Milad, I. Liberzon, Biological studies of post-traumatic stress disorder, *Nature Reviews Neuroscience* 13 (11) (2012) 769–787. doi:10.1038/nrn3339.

- [5] D. O. Hebb, *The Organisation of Behavior: A Neuropsychological Theory*, Wiley, New York, 1949.
- [6] T. L. Roth, F. D. Lubin, A. J. Funk, J. D. Sweatt, Lasting epigenetic influence of early-life adversity on the BDNF gene, *Biological Psychiatry* 65 (9) (2009) 760–769. doi:10.1016/j.biopsych.2008.11.028.
- [7] T. V. P. Bliss, G. L. Collingridge, A synaptic model of memory: long-term potentiation in the hippocampus, *Nature* 361 (1993) 31–39. doi:10.1038/361031a0.
- [8] J. E. LeDoux, *The Emotional Brain: The Mysterious Underpinnings of Emotional Life*, Simon and Schuster, New York, 1996.
- [9] A. Öhman, The role of the amygdala in human fear: Automatic detection of threat, *Psychoneuroendocrinology* 30 (10) (2005) 953–958. doi:10.1016/j.psyneuen.2005.03.019.
- [10] R. B. Zajonc, Feeling and thinking: Preferences need no inferences, *American Psychologist* 35 (2) (1980) 151–175. doi:10.1037/0003-066X.35.2.151.
- [11] L. F. Barrett, *How Emotions Are Made: The Secret Life of the Brain*, Houghton Mifflin Harcourt, New York, 2017.
- [12] J. D. Teasdale, R. J. Howard, S. G. Cox, Y. Ha, M. J. Brammer, S. C. R. Williams, S. A. Checkley, Functional MRI study of the cognitive generation of affect, *American Journal of Psychiatry* 156 (2) (1999) 209–215. doi:10.1176/ajp.156.2.209.
- [13] E. B. Foa, J. D. Huppert, S. P. Cahill, Emotional processing theory: An update, *Pathological Anxiety: Emotional Processing in Etiology and Treatment* (2007) 3–24.
- [14] M. G. Craske, K. Kircanski, M. Zelikowsky, J. Mystkowski, N. Chowdhury, A. Baker, Optimizing inhibitory learning during exposure therapy, *Behaviour Research and Therapy* 46 (1) (2008) 5–27. doi:10.1016/j.brat.2007.10.003.

- [15] S. G. Hofmann, A. Asnaani, I. J. J. Vonk, A. T. Sawyer, A. Fang, The efficacy of cognitive behavioral therapy: A review of meta-analyses, *Cognitive Therapy and Research* 36 (5) (2012) 427–440. doi:10.1007/s10608-012-9476-1.
- [16] A. C. Butler, J. E. Chapman, E. M. Forman, A. T. Beck, The empirical status of cognitive-behavioral therapy: A review of meta-analyses, *Clinical Psychology Review* 26 (1) (2006) 17–31. doi:10.1016/j.cpr.2005.07.003.
- [17] F. Shapiro, EMDR therapy humanitarian assistance programs: Treating the psychological, physical, and societal effects of adverse experiences worldwide, *Journal of EMDR Practice and Research* 8 (4) (2014) 181–186. doi:10.1891/1933-3196.8.4.181.
- [18] C. J. Harmer, G. M. Goodwin, P. J. Cowen, Why do antidepressants take so long to work? a cognitive neuropsychological model of antidepressant drug action, *British Journal of Psychiatry* 195 (2) (2009) 102–108. doi:10.1192/bjp.bp.108.051193.
- [19] E. Bandara, R. Gore, S. Shetty, R. Mukkamala, C. Rhea, A. Yarlagadda, S. Kaushik, L. De Silva, A. Maznychenko, I. Sokolowska, et al., Standardization of neuromuscular reflex analysis—role of fine-tuned vision-language model consortium and openai gpt-oss reasoning llm enabled decision support system, *arXiv preprint arXiv:2508.12473* (2025).
- [20] K. N. Ochsner, J. J. Gross, The cognitive control of emotion, *Trends in Cognitive Sciences* 9 (5) (2005) 242–249. doi:10.1016/j.tics.2005.03.010.
- [21] J. J. Gross, Antecedent- and response-focused emotion regulation: Divergent consequences for experience, expression, and physiology, *Journal of Personality and Social Psychology* 74 (1) (1998) 224–237. doi:10.1037/0022-3514.74.1.224.
- [22] J. J. Gross, Emotion regulation: Current status and future prospects, *Psychological Inquiry* 26 (1) (2015) 1–26. doi:10.1080/1047840X.2014.940781.
- [23] A. Aldao, S. Nolen-Hoeksema, S. Schweizer, Emotion-regulation strategies across psychopathology: A meta-analytic review, *Clinical Psychology Review* 30 (2) (2010) 217–237. doi:10.1016/j.cpr.2009.11.004.

- [24] D. M. Fresco, M. T. Moore, M. H. M. van Dulmen, Z. V. Segal, S. H. Ma, J. D. Teasdale, J. M. G. Williams, Initial psychometric properties of the experiences questionnaire: Validation of a self-report measure of decentering, *Behavior Therapy* 38 (3) (2007) 234–246. doi:10.1016/j.beth.2006.08.003.
- [25] A. Wells, *Metacognitive Therapy for Anxiety and Depression*, Guilford Press, New York, 2009.
- [26] A. Citri, R. C. Malenka, Synaptic plasticity: Multiple forms, functions, and mechanisms, *Neuropsychopharmacology* 33 (1) (2008) 18–41. doi:10.1038/sj.npp.1301559.
- [27] C. Lüscher, R. C. Malenka, NMDA receptor-dependent long-term potentiation and long-term depression (LTP/LTD), *Cold Spring Harbor Perspectives in Biology* 4 (6) (2012) a005710. doi:10.1101/cshperspect.a005710.
- [28] E. Bandara, R. Gore, A. Yarlagadda, A. H. Clayton, P. Samuel, C. K. Rhea, S. Shetty, Standardization of psychiatric diagnoses—role of fine-tuned llm consortium and openai-gpt-oss reasoning llm enabled decision support system, *arXiv preprint arXiv:2510.25588* (2025).
- [29] E. Bandara, R. Gore, X. Liang, S. Rajapakse, I. Kularathne, P. Karunarathna, P. Foytik, S. Shetty, R. Mukkamala, A. Rahman, et al., Agentsway—software development methodology for ai agents-based teams, *arXiv preprint arXiv:2510.23664* (2025).
- [30] J. Torous, J. Nicholas, M. E. Larsen, J. Firth, H. Christensen, Clinical review of user engagement with mental health smartphone apps: evidence, theory and improvements, *Evidence Based Mental Health* 21 (3) (2017) 116–119. doi:10.1136/eb-2018-102891.
- [31] D. C. Mohr, M. Zhang, S. M. Schueller, Personal sensing: Understanding mental health using ubiquitous sensors and machine learning, *Annual Review of Clinical Psychology* 13 (2017) 23–47. doi:10.1146/annurev-clinpsy-032816-044949.
- [32] A. S. Miner, A. Milstein, S. Schueller, R. Hegde, C. Mangurian, E. Linos, Smartphone-based conversational agents and responses to

- questions about mental health, interpersonal violence, and physical health, *JAMA Internal Medicine* 176 (5) (2016) 619–625. doi:10.1001/jamainternmed.2016.0400.
- [33] E. Bandara, A. Gunaratna, R. Gore, A. H. Clayton, C. K. Rhea, S. Rajapakse, I. Kularathna, S. Shetty, R. Mukkamala, X. Liang, et al., Toward zero-egress psychiatric ai: On-device llm deployment for privacy-preserving mental health decision support, *arXiv preprint arXiv:2604.18302* (2026).
 - [34] D. A. Huffman, A method for the construction of minimum-redundancy codes, *Proceedings of the IRE* 40 (9) (1952) 1098–1101. doi:10.1109/JRPROC.1952.273898.
 - [35] R. J. McNally, Progress and controversy in the study of posttraumatic stress disorder, *Annual Review of Psychology* 54 (2003) 229–252. doi:10.1146/annurev.psych.54.101601.145112.
 - [36] E. B. Foa, et al., *Prolonged exposure therapy for PTSD*, Oxford University Press (2007).
 - [37] F. Shapiro, *Eye movement desensitization and reprocessing*, Guilford Press (2001).
 - [38] A. Ehlers, D. M. Clark, A cognitive model of posttraumatic stress disorder, *Behaviour Research and Therapy* 38 (2000) 319–345.
 - [39] J. D. Teasdale, et al., Metacognitive awareness and prevention of relapse in depression, *Journal of Consulting and Clinical Psychology* 70 (2002) 275–287.
 - [40] J. A. Brewer, P. D. Worhunsky, J. R. Gray, Y.-Y. Tang, J. Weber, H. Kober, Meditation experience is associated with differences in default mode network activity and connectivity, *Proceedings of the National Academy of Sciences* 108 (50) (2011) 20254–20259.
 - [41] F. J. Varela, E. Thompson, E. Rosch, *The Embodied Mind: Cognitive Science and Human Experience*, MIT Press, Cambridge, MA, 1991.
 - [42] L. F. Barrett, *How emotions are made: The secret life of the brain* (2017).

- [43] K. K. Fitzpatrick, et al., Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent, *JMIR Mental Health* 4 (2017).
- [44] B. Inkster, S. Sarda, V. Subramanian, An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study, *JMIR mHealth and uHealth* 6 (11) (2018) e12106. doi:10.2196/12106.
- [45] A. S. Rizzo, et al., Ptsd and the virtual iraq/afghanistan system, *Journal of CyberTherapy and Rehabilitation* 3 (2010).
- [46] K. Yang, et al., Mentallama: interpretable mental health analysis, arXiv preprint arXiv:2309.13567 (2024).
- [47] M. Abdin, et al., Phi-3 technical report, arXiv preprint arXiv:2404.14219 (2024).
- [48] B. Bodhi (Ed.), *A Comprehensive Manual of Abhidhamma: The Abhidhammattha Sangaha of Ācariya Anuruddha*, Buddhist Publication Society, Kandy, Sri Lanka, 2000, see treatment of the six sense bases (saḷāyatana), including mano (mind) as the sixth sense whose objects are mental phenomena.
- [49] R. Gethin, *The Foundations of Buddhism*, Oxford University Press, Oxford, 1998.
- [50] R. Gore, E. Bandara, S. Shetty, A. E. Musto, P. Rana, A. Valencia-Romero, C. Rhea, L. Tayebi, H. Richter, A. Yarlagadda, et al., Proof-of-tbi-fine-tuned vision language model consortium and openai-o3 reasoning llm-based medical diagnosis support system for mild traumatic brain injury (tbi) prediction, arXiv preprint arXiv:2504.18671 (2025).
- [51] M. D. Bhaskaran, B. N. Smith, Effects of KCNQ channel modulation on synaptic transmission and plasticity in nucleus tractus solitarius, *European Journal of Neuroscience* 37 (12) (2013) 1285–1294. doi:10.1111/ejn.12140.
- [52] Z. V. Segal, J. M. G. Williams, J. D. Teasdale, *Mindfulness-Based Cognitive Therapy for Depression: A New Approach to Preventing Relapse*, Guilford Press, New York, 2002.

- [53] E. A. Maguire, D. G. Gadian, I. S. Johnsrude, C. D. Good, J. Ashburner, R. S. J. Frackowiak, C. D. Frith, Navigation-related structural change in the hippocampi of taxi drivers, *Proceedings of the National Academy of Sciences* 97 (8) (2000) 4398–4403. doi:10.1073/pnas.070039597.
- [54] B. Draganski, C. Gaser, V. Busch, G. Schuierer, U. Bogdahn, A. May, Changes in grey matter induced by training, *Nature* 427 (6972) (2004) 311–312. doi:10.1038/427311a.
- [55] B. K. Hölzel, J. Carmody, M. Vangel, C. Congleton, S. M. Yerramsetti, T. Gard, S. W. Lazar, Mindfulness practice leads to increases in regional brain grey matter density, *Psychiatry Research: Neuroimaging* 191 (1) (2011) 36–43. doi:10.1016/j.psychresns.2010.08.006.
- [56] R. J. Davidson, J. Kabat-Zinn, J. Schumacher, M. Rosenkranz, D. Muller, S. F. Santorelli, F. Urbanowski, A. Harrington, K. Bonus, J. F. Sheridan, Alterations in brain and immune function produced by mindfulness meditation, *Psychosomatic Medicine* 65 (4) (2003) 564–570.
- [57] E. Bandara, A. Gunaratna, R. Gore, A. Rahman, R. Mukkamala, S. Shetty, S. Rajapakse, I. Kularathna, P. Foytik, S. H. Bouk, et al., Ai trust os—a continuous governance framework for autonomous ai observability and zero-trust compliance in enterprise environments, *arXiv preprint arXiv:2604.04749* (2026).
- [58] E. Bandara, R. Gore, A. Gunaratna, R. Mukkamala, N. Siriwardanagea, S. Rajapakse, I. Kularathna, P. Karunarathna, W. Herath, C. Rajapakse, et al., Mindgap: A conversational ai framework for upstream neuroplastic intervention in post-traumatic stress disorder, *arXiv preprint arXiv:2605.14660* (2026).
- [59] E. Bandara, R. Gore, A. Gunaratna, S. Rajapakse, I. Kularathna, R. Mukkamala, S. Shetty, X. Liang, A. Hass, T. Hewa, et al., Think before you act—a neurocognitive governance model for autonomous ai agents, *arXiv preprint arXiv:2604.25684* (2026).
- [60] A. Grattafiori, A. Dubey, A. Jauhri, et al., The Llama 3 herd of models, *arXiv preprint arXiv:2407.21783* (2024).
- [61] A. Yarlagadda, E. Bandara, R. Gore, A. H. Clayton, P. Samuel, C. K. Rhea, S. Shetty, R. Mukkamala, X. Liang, A. Hass, A. Rahman, Train

the trainers: An agentic AI framework for peer-based mental health support in battlefield environments, arXiv preprint arXiv:2605.16269 (2026).

- [62] E. Bandara, T. Hewa, R. Gore, S. Shetty, R. Mukkamala, P. Foytik, A. Rahman, S. H. Bouk, X. Liang, A. Hass, Towards responsible and explainable AI agents with consensus-driven reasoning, arXiv preprint arXiv:2512.21699 (2025).
- [63] E. Bandara, R. Gore, A. Yarlagadda, A. H. Clayton, P. Samuel, C. K. Rhea, S. Shetty, Standardization of psychiatric diagnoses: Role of a fine-tuned LLM consortium and OpenAI-GPT-OSS reasoning LLM enabled decision support system, arXiv preprint arXiv:2510.25588 (2025).
- [64] E. Bandara, T. Hewa, R. Gore, S. Shetty, R. Mukkamala, P. Foytik, A. Rahman, S. H. Bouk, X. Liang, A. Hass, et al., Towards responsible and explainable ai agents with consensus-driven reasoning, arXiv preprint arXiv:2512.21699 (2025).
- [65] E. Bandara, R. Gore, P. Foytik, S. Shetty, R. Mukkamala, A. Rahman, X. Liang, S. H. Bouk, A. Hass, S. Rajapakse, et al., A practical guide for designing, developing, and deploying production-grade agentic ai workflows, arXiv preprint arXiv:2512.08769 (2025).
- [66] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, W. Chen, LoRA: Low-rank adaptation of large language models, International Conference on Learning Representations (ICLR) (2022).
- [67] T. Detrmers, A. Pagnoni, A. Holtzman, L. Zettlemoyer, QLoRA: Efficient finetuning of quantized LLMs, Advances in Neural Information Processing Systems (NeurIPS) 36 (2023).
- [68] E. Bandara, A. Hass, R. Gore, S. Shetty, R. Mukkamala, S. H. Bouk, X. Liang, N. W. Keong, K. De Zoysa, A. Withanage, et al., Astride: A security threat modeling platform for agentic-ai applications, arXiv preprint arXiv:2512.04785 (2025).