Trends in Cognitive Sciences

Towards an integrated science of autobiographical memory and mental health

--Manuscript Draft--

Manuscript Number:	
Article Type:	Opinion
Keywords:	autobiographical memory; treatment; predictive processing; cognitive behavior therapy; RDoC; transdiagnostic
Corresponding Author:	Caitlin Hitchcock UNITED KINGDOM
First Author:	Caitlin Hitchcock
Order of Authors:	Caitlin Hitchcock
	Tim Dalgleish
Abstract:	Autobiographical memory– the capacity to consciously recollect how things were in the past– plays a driving role in our perceptions of ourselves and the world. We propose that the core function of such remembering is to optimise our predictions of our unfolding future. As such, autobiographical memory underpins mental health and mental ill health, as a function of our experiences in the past. Here, we integrate established clinical therapeutic models of autobiographical memory into a broader predictive processing approach. This provides a transdiagnostic framework that accommodates diverse autobiographical memory phenomena in clinical psychology and generates proposals for therapeutic intervention. Finally, we submit that the case is now strong to consider autobiographical memory as a core research domain for mental health.

MRC Cognition and Brain Sciences Unit

UNIVERSITY OF CAMBRIDGE

18th September 2020

Dear Dr Drayton,

Please find enclosed our pre-approved submission for an opinion piece for *Trends in Cognitive Sciences* entitled 'Towards an integrated science of autobiographical memory and mental health'.

In the paper, we propose a transdiagnostic predictive framework to integrate diverse autobiographical memory phenomena from encoding to retrieval, and make a case for autobiographical memory as a core transdiagnostic research domain for mental health. In doing so, we situate autobiographical memory within both the predictive brain framework advocated in cognitive neuroscience and computational psychiatry, and Beckian cognitive theories which underlie psychotherapeutic techniques.

We believe this opinion article will stimulate interdisciplinary collaboration between clinical psychologists, computational psychiatrists, and cognitive neuroscientists, and contribute towards bridging the gap between clinical practice and cognitive science.

We look forward to your feedback.

Best wishes,

Dr Caitlin Hitchcock and Professor Tim Dalgleish
MRC Cognition and Brain Sciences Unit, University of Cambridge

Towards an integrated science of autobiographical memory and mental health

Caitlin Hitchcock

Tim Dalgleish

Medical Research Council Cognition and Brain Sciences Unit, University of Cambridge,

Cambridgeshire and Peterborough NHS Foundation Trust

Correspondence: Caitlin Hitchcock and Tim Dalgleish, MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge, CB2 7EF; Caitlin.hitchcock@mrc-cbu.cam.ac.uk; tim.dalgleish@mrc-cbu.cam.ac.uk; +44 1223 273 744

Key words: autobiographical memory, treatment, predictive processing, cognitive behavior therapy, RDoC, transdiagnostic

Abstract

Autobiographical memory – the capacity to consciously recollect how things were in the past – plays a driving role in our perceptions of ourselves and the world. We propose that the core function of such remembering is to optimise our predictions of our unfolding future. As such, autobiographical memory underpins mental health and mental ill health, as a function of our experiences in the past. Here, we integrate established clinical therapeutic models of autobiographical memory into a broader predictive processing approach. This provides a transdiagnostic framework that accommodates diverse autobiographical memory phenomena in clinical psychology and generates proposals for therapeutic intervention. Finally, we submit that the case is now strong to consider autobiographical memory as a core research domain for mental health.

The double-edged sword of autobiography

Human minds are unique in their ability to create, manipulate and reflect upon mental representations of themselves and the world that are decoupled from present reality[1]. For almost half our waking hours, our minds are filled with things other than the activity we are currently engaged in[2]. Much of this time is spent immersed in our autobiographical past, processing its relevance for who we are now, what we are doing, and for what we anticipate doing down the road. These personal recollections come in many guises, ranging from autobiographical themes grounded in generalised self-relevant knowledge to vivid autonoetic memories of specific episodes recreated in rich sensory detail[3]. Our autobiographical memory (**AM; see Glossary**) thereby colours all our thoughts, judgements, imaginations, narratives and decisions, from the trivial to profound. In its more abstruse forms, AM intertwines with our models of self to generate a complex narrative identity across time, richly populated with semantic self-knowledge[4]. Most critically, this ability to recollect and re-inhabit the past supports the mind's enormous computational power as a prediction machine, allowing us to anticipate and pre-empt our unfolding future[5].

Unsurprisingly, then, the nature and content of AM is intertwined with our mental health. If we recruit the past to predict dynamic interactions between ourselves and the world in the unfolding present and future, then how we feel about those interactions will be closely governed by how propitious or otherwise our autobiography tells us these predicted interactions will be. For many situations, this is clearly advantageous. However, when lessons from the past are a poor match for our current lives then these predictions will mislead us about the here and now, with detrimental effects on mental health[6] – AM is a doubled-edged sword.

Here we outline a **transdiagnostic** predictive framework to integrate diverse AM phenomena associated with mental health and its management. We discuss AM encoding and retrieval processes, before outlining a set of AM-informed therapeutic principles. We close by making the case for AM's inclusion in the National Institute of Mental Health's Research Domain Criteria matrix[7] as a core transdiagnostic research domain for mental health.

A generalised predictive framework for understanding AM

If we assume the core role of AM is to optimise predictions about our interactions with the world in the unfolding present and future, then it is pragmatic to situate AM processes within a general **predictive processing** framework of mind[5]. Hierarchical prediction models in cognitive science have a long history dating back to Helmholz's proposals of unconscious inference[8] and encompass cybernetics[9] and Powers' influential perceptual control theory[10]. Predictive processing approaches now aspire to a complete theory of mental life[11] including AM[5] and diverse aspects of affect[12, 13], mental health[14-16] and therapeutics[17] realised both mathematically and neurally. Our ambition here is to present an intentionally non-technical predictive processing account of the myriad interactions between AM and mental health.

At the heart of predictive processing approaches is a dynamic hierarchy of active prediction models about the interactions of the self and world, realised across multiple temporal and spatial scales, from a narrative self across the lifespan, through broad spatio-temporal contexts (the self in a particular relationship, role, location, or time period), down to immediate experiential encounters with the rapidly unfolding present[5, 18] (Figure 1; **Key Figure**). Each level of this hierarchy is dynamically populated from long-term stores (probability spaces) of self and world information that have different degrees of relevance for current circumstances. Thus, for example, 'the self-in-place' – the hierarchy of models that is active[19] – in a work context will likely be different to the self-in-place in an intimate relationship context. Similarly, the self-in-place during a depressive episode will differ markedly from the self-in-place when depression is in remission.

Within this hierarchical framework, nested models generate a stream of feedforward probabilistic predictions (Figure 1). Any divergences between these descending predictions and bottom-up input from the levels below are coded as '**prediction-errors'** that vary in how much they diverge from prior predictions. Ascending prediction-errors are used to iteratively refine the hierarchy of models to generate updated predictions in a continuous interactive dynamic. The system can therefore be conceptualised as a bidirectional hierarchical cascade where higher layers

predict what occurs at the layers below and lower layers return prediction-errors to the layers above[5].

Critically for our understanding of AM, feedforward predictions from higher hierarchical levels modulate how information is prioritised and processed at lower-levels right down to our direct interface with the external (or internal) world. First, stored information, including AM, that is consistent with the prediction models is selectively accessed. Second, we selectively process the unfolding present (e.g., selective attention, perception) in line with model predictions. Finally, we act or 'behave' within our external (actions in the world[5]) or internal (e.g., thought narratives, chains of AM recollection[7]) milieus so as to confirm these predictions; a process known as **active inference**[20].

An important feature of the predictive processing approach is a weighting of the degree of confidence –'**precision**' – applied to our predictions. The more precise the predictions, the lower the respective weighting or precision that is placed on any ascending prediction-errors signalling the extent to which the prediction models might be wrong. Models are more or less precise as a function of the extent to which they provide a good account of regularities across our entire previous experience in that context or situation. Highly precise prediction models are therefore resistant to modification in the face of prediction-errors as each error counts for little against the weight of past regularities. In contrast, models with lower precision are more open to modification by ascending prediction-errors or even replacement in the hierarchy with models that have greater precision in the unfolding context[5].

One advantage of such a prediction framework is that it locates understanding of AM within a broader neurocognitive architecture that seeks to account for the entirety of mental life[5, 21]. The framework can also be readily mapped onto existing conceptualisations of AM within both cognitive and clinical psychology; for example, the **self-memory system** – the leading theory of AM – proposes a 'working self' that is compiled from long-term stores of self-knowledge, which selectively recruits AMs that accord to active goal states[4].

Glossary

Active inference: The proposal that action in the world or in the mind (thinking, problem-solving etc.) fulfils predictions within self-organising or self-evidencing systems.

AM: Autobiographical memory for personal life experiences. Autobiographical information is stored in a fluid manner, such that autobiographical knowledge can be dynamically retrieved as single-episode event memories that are rich in specific detail, or as generalised representations of the past that summarise categories of events or extended time periods. Provides a basis for self-identity.

Beckian cognitive theory: An influential model of the architecture of mind underlying common mental health problems. Maladaptive schemas code dysfunctional core beliefs about the self, world and others that drive maladaptive perceptions, cognitions and behaviours.

Cognitive behaviour therapy (CBT): A family of psychological treatments based on Beckian cognitive theory of emotional disturbance. Treatment focuses on altering dysfunctional behaviours and updating cognitive representations that maintain symptoms.

Consolidation: The time-dependent process by which recently encoded experiences are transformed into long-term memory.

Kindling: Depressogenic responses become triggered more easily after repeated depressive episodes, increasing vulnerability to relapse.

Phenomenological centre: Self-related or world-related model that is currently dominating awareness. Typically focussed on current concerns.

Precision: A weighting of the confidence placed by the system in predictions or prediction errors. The inverse of variance in the world.

Prediction error: A difference index of the discrepancy between predictions and sensory or mental input.

Predictive processing: A framework for understanding neurocognitive functioning in which the mind and brain are constantly generating and updating a hierarchical mental model of the self in the world. The model generate predictions of sensory and mental input that are compared to

actual input. These comparisons result in prediction errors that are then used to update and revise mental models.

Schema: A cognitive framework used to understand the world, inter-relations between objects, and the place of the self in the world. In mental health disorders, schemas tend to be dysfunctional and are a key determinant of symptoms.

Self-evidencing: Information is selectively attended to, encoded, and accessed in a manner which supports active schemas or higher-order prediction models. A property of self-organising systems.

Self-memory system: An influential model of autobiographical memory proposed by Conway and Pleydell-Pearce (2000). The self-memory system is composed of an autobiographical knowledge base and a working self.

Sensitisation: Repeated exposure to a stimulus (in this case, depressive episodes) increases responsivity such that changes develop with faster onset, increased magnitude and longer duration.

Transdiagnostic: The proposal that a construct (a system, theory, process, symptom, behaviour) applies to more than one diagnostic grouping (e.g., depression, posttraumatic stress disorder, within psychiatry).

Similarly, within **Beckian cognitive theories** of mental health[22], higher-order models of self and world are characterised as '**schemas**' that drive affect and exert a top-down influence on lower-level processing in the form of cognitive biases, thinking distortions and behaviour[23]. Schemas are representational summaries of past experience, and mental health problems such as depression or anxiety are proposed to arise when 'maladaptive ' schemas become consolidated as a function of a history of negative experiences. Within a prediction framework, schemas equate to higher-order self-world prediction models, and the high degree of consolidation of dysfunctional schemas equates to increased precision. Schema-congruent biases in cognition and behaviour (including AM) and reduced precision of schema-incongruent information equate to the '**self-evidencing**' of these prediction models[16].

Beckian theory[22] emphasises the key role of active schemas in determining the degree of 'generality' of mental experience, including AMs. So, in addition to relevant specific episodic AMs (e.g. '*the time I failed an important test at school'*), when schemas dominate processing our minds are drawn to congruent general themes and regularities within AM – so-called 'categorical AMs'[24] (e.g., '*I was useless at school*') – that are coded at a comparable level of spatio-temporal granularity as the schemas themselves. Within the broader hierarchical prediction framework, we can therefore conceptualise a '**phenomenological centre**' within the hierarchy of models, the location of which determines the degree of generality of self- and world-related mental content that currently dominates awareness.

The phenomenological centre can shift across the hierarchy with concordant shifts in the granularity of mental content. AM content can therefore range from mental reflections on broad AM themes across the life course down to autonoetic reliving of specific past episodes. For those with current mental health difficulties (Figure 1) the phenomenological centre is typically populated from contextually-broader (schema-level) higher-order prediction models about the self and world that are central to ongoing concerns.

Beckian models[22] account for mental health risk with the proposal that, when mental ill health is prodromal or in remission, maladaptive schemas are latent but liable to become activated when the individual encounters significant schema-congruent information in the world. Within a prediction framework, schema-latency equates to an elevated degree of accessibility of competing self-world models within long-term (probabilistic) storage (Figure 1). As a consequence of this competition from stored models, the hierarchy of models in place during vulnerable states has reduced precision. As a result, when negative experiences are encountered the resultant prediction-errors are relatively more precise, increasing the probability that the prediction models in place require modification. When the latent 'dysfunctional' models offer a more precise account of current experience they will become activated, put 'in place', and the downward spiral into mental health difficulties will commence.

Encoding autobiographical memories

Newly encountered to-be-encoded events will vary in how well they align with active predictions, with the degree of divergence coded as prediction-error. The precision of prior prediction models – how much confidence is placed on them relative to competing alternatives – determines the precision or weighting placed on these prediction-errors. For highly-precise predictions, prediction-errors will have relatively low precision and their capacity to shift prior prediction models will be low (Figure 1).

We can apply these principles to a number of permutations relevant to mental health. During experiences of common mental health difficulties – for example, the depressive mindset in Figure 1 – the set of hierarchical maladaptive prediction models will have relatively high precision. This reflects the 'goodness-of-fit' of these maladaptive models in capturing regularities in adversity across past experience. Increased recurrence of depression will increase the precision of these dysfunctional models over time (**kindling**[25]), rendering them increasingly susceptible to activation when congruent events are encountered (**sensitisation**[26]).

When maladaptive models are active, expectations (predictions) of positive experiences are low and, when encountered, such positive events will generate prediction-errors with low precision and consequent low capacity to modify the maladaptive prediction models[16]. In contrast, predictions of negative experiences are relatively high, details of these prediction-congruent events will be selectively processed (cognitive bias) as the prediction models **self-evidence**. These prediction-aligned events will be encoded as an experience-near record of the unfolding present, their **consolidation** and reconsolidation (Box 1) further sharpening the precision of the maladaptive prediction model hierarchy.

For individuals not currently experiencing an episode of mental ill health but at elevated risk, vulnerability is represented by a hierarchy of active prediction models coding a fragile 'good enough' mindset. These active prediction models have low precision, relative to the more functional and precise models of low-vulnerability individuals. Consequently, not only will relatively more

negative events be predicted and subsequently encoded, but prediction-errors generated by unexpected negative events will have relatively higher precision, a greater capacity to modify the prior fragile active prediction models, and will be encoded more strongly as a record of model-incongruent experience. When vulnerability is particularly high, and/or when negative events are encountered that mesh closely with latent maladaptive prediction models, then the active but fragile hierarchy of 'good enough' models will be usurped by a dysfunctional model structure and mental health will spiral downwards.

Some newly encountered negative events – traumas – are highly unpredicted and represent significant danger. Even for those with no history of mental health problems, for whom prior 'all is basically fine' prediction models have high precision, the very large prediction-errors generated by these events mandate that prior models be revised. Trauma memories are thus encoded into long-term storage with little modification as extreme model-incongruent experiences[27]. The psychological struggle to assimilate traumatic events into existing models is reflected in the oscillating intrusion and avoidance symptoms of posttraumatic stress disorder (PTSD). In the worst cases, traumatic events rapidly generate a new 'trauma-centric' hierarchy of models coding the world as dangerous, uncontrollable and unpredictable, and self as broken or vulnerable[28] that is resistant to therapeutic change.

Box 1: Encoding, consolidation and reconsolidation of AM

Research elucidating consolidation and reconsolidation exemplifies how AM can be linked from animal models detailing molecular and cellular mechanisms[29], to pharmacological and behavioural manipulation of such mechanisms[30, 31], and development of mechanistic clinical interventions[32, 33]. Extant theory[see 34] proposes that new AMs consolidate slowly over time, offering the potential for the reactivation, alteration, and reconsolidation of information. At the cellular level, consolidation requires protein synthesis in the hippocampus, amygdala and associated structures, such that protein synthesis inhibitors may destabilise the consolidation process[35, 36].

This has important implications not only for the retention of new learning, but for psychological presentations that are driven by negative event or fear memories.

Reactivation and reconsolidation of trauma memories to edit maladaptive content is proposed to update the self- and world-related prediction models that an individual carries forward[37], and is a key mechanism underlying recommended psychological interventions for PTSD[NICE; 38]; Trauma-focussed **cognitive behaviour therapy** (CBT) and Eye Movement Desensitization and Reprocessing (EMDR). Similarly, fearful event memories play a causal role in anxiety responses[30], such that weakening a fear memory reduces behavioural and physiological fear expression[31, 39]. Various proposals have outlined how AM reconsolidation techniques could be behaviourally and pharmacologically enhanced for treatment purposes[40, 41], with some evidence to support the role of N-methyl-D-aspartate receptor (NMDAR) activation in enhancing the effects of exposure therapy for specific phobia, panic, social anxiety, and obsessive compulsive disorder[42, 43]. Similarly, simplified interventions designed to target underlying AM mechanisms are being trialled to prevent the development of symptoms (e.g., playing Tetris after a traumatic event to reduce later involuntary memories[44]).

A key next step is to evaluate how memory reconsolidation relates to alteration of spatio-temporally higher-order self-world prediction models (or schema in **Beckian cognitive theory**). AM representations are destabilised and subsequently reconsolidated upon each retrieval. While this offers the ability to adaptively update AMs via therapeutic intervention, it may conversely pose a problem for mental health in naturalistic settings in which AMs consistent with maladaptive higher-order prediction models are repeatedly retrieved and thus reconsolidated. Elucidating the dynamic interplay between currently active prediction models and the encoding, consolidation, and reconsolidation of AMs will offer important insights into mechanisms promoting symptom maintenance and change.

Retrieval of autobiographical memories

Voluntary retrieval

Deliberate recollection of prior autobiographical experiences underlies a number of daily cognitive processes, including problem solving[45], future planning[46, 47], social interactions[48], judgements[49], and developing and maintaining a coherent identity across time. Within the prediction framework advocated here, voluntary retrieval represents a form of 'psychological active inference'[20] where action in the mind – searching the AM database[3, 4] – acts to confirm model predictions by selectively accessing congruent content (self-evidencing). Retrieved AMs will vary in their degree of generality from categorical themes reflecting regularities across the lifespan to specific autonoetic recollections of relevant prior events. The modal level of generality will be determined by the individual's phenomenological centre. For those struggling with their mental health and whose psychological experience is currently focused on broad higher-level self-world prediction models (e.g., 'their sense of worthlessness'; see Figure 1), retrieved AMs will be more generalised and aligned with the contextual breadth of the models-in-place. In such circumstances, specific AM retrieval will require more effort, as evidenced in countless research studies[50-52], with detrimental effects on mental health [28; 55-59], and mental life will be populated with ruminations and worries centred around these generalised autobiographical tropes[53].

Involuntary retrieval

Intrusive and involuntary AMs are ubiquitous[54], ranging from affectively-benign recollections to intrusive reliving of life-changing past traumas. Repeated involuntary retrieval of highly emotive AMs characterizes PTSD, mood-, eating-, and anxiety-disorders, maintains distress, and links to poor prognosis[55-57].

Within the predictive processing framework advocated here, we posit that two forms of involuntary AMs predominate in those struggling with mental health problems. First, as with self-directed voluntary retrieval, the hierarchy of in-place self-world prediction models will automatically self-evidence by involuntarily accessing congruent AMs at varying degrees of

generality, from specific prototypical episodes that reinforce predictions to generalised autobiographical themes that fuel ruminations and worries. Second, we propose that AMs of past episodes that significantly diverge from the content of active prediction models, are also differentially activated and accessible for recollection[66]. These discrepant AMs will generate prediction-errors, varying in their precision, that continuously calibrate the models-in-place. One function of these specific AMs therefore appears to be to delineate the 'boundary conditions' of the active self-world models – these AMs essentially act as a record of events from the past when the predictions did not apply and act as a brake on the overgeneralisation of predictions (Box 2).

The most potent and disabling experience of involuntary retrieval involves intrusive reliving of trauma memories. Intrusions can occur spontaneously or be cued by congruent triggers[58]. These intrusions are a recurring reminder of the violated boundary conditions of pre-trauma prediction models that are no longer plausibly tenable. Traumatic intrusions are a central symptom and maintaining factor of PTSD[71,72] and at the extreme contribute to the experience of dissociation[59], and auditory and visual hallucinations[60].

Box 2: AMs as markers of boundary conditions

Within the framework we have described, one function of specific AMs which are inconsistent with prediction models may be to provide exemplars of situations in which current self-models are invalid. That is, incongruent specific memories may mark the boundaries of self-world models. In a prior experiment[61], we demonstrated that when rating how well a negative personality trait (e.g., *boring*) described themselves (relative to the control condition of providing the dictionary definition of that same word), healthy individuals were primed (see Figure I) to recall specific memories that were *inconsistent* with the negative trait (e.g., *I made a colleague laugh yesterday*). In contrast, in depressed individuals, this priming effect disappeared, consistent with

relatively unbounded, negative self-world models in depression. For positive personality characteristics, the opposite occurred. When making a positive generalisation about themselves (e.g., *successful*), depressed participants were primed to recall an inconsistent specific memory (e.g., *I failed a final year exam*). This time, no such priming effect was evident for healthy individuals, consistent with the overly-generous, rose-tinted self-world models that characterise good mental health. Interestingly, for individuals in depression remission, there were no priming effects for positive or negative traits. This potentially suggests that the fragile 'good enough' self-models in-place during remission lack the same clear boundaries of the in-place models that respectively characterise mental health or acute depression.

Figure I. *Mean (square-root transformed) response time (ms; y-axis) for recall of an inconsistent specific memory after rating how well a negative (top) or positive (bottom) personality trait describes the self (trait self-rating), relative to a dictionary definition control condition.* Taken from [61]. Note. *=p<.05. Error-bars= standard error. Shorter response times for trait self-rating relative to definition represents the priming effect.

At a behavioural, self-report level, it can be difficult to disentangle a generalised AM (e.g., *People told me I was boring all through high school*) from a core belief characterising a schema or self-world model (e.g., *I am boring*). Recent representational similarity analysis of neuroimaging data demonstrated that specific episodic information (e.g., a person/place) yields replicable activity patterns [62], potentially allowing future research to isolate specific episodic information during more complex cognitive tasks (e.g., belief ratings). Neuroimaging techniques to isolate specific AM episodes may therefore help elucidate the role of AMs in forming and maintaining the content of higher-order precision models, allowing assessment of central tenets of cognitive theories of mental ill-health.

Autobiographical memory and therapeutic practice

The prediction framework advocated here suggests three routes to enacting therapeutic change for common mental health problems. First, the precision of the maladaptive hierarchy of

self-world models and the products of their self-evidencing (e.g., congruent negative thoughts, AMs, perceptions etc.) can be reduced. Second, the salience of prediction-discrepant positive and/or self-affirming experiences can be enhanced, thus augmenting the size of ascending prediction-errors and their resultant capacity to modify maladaptive models. Finally, the precision of alternative – 'Theory B'[63] – sets of prediction models can be augmented to foster and reinforce a competing set of hierarchical prediction models that operationalise mentally-healthy functioning.

Traditional CBT seeks to do all of these things[22, 64]. The precision of the dysfunctional prediction model hierarchy and its products (maladaptive schemas and associated core beliefs, dysfunctional assumptions and negative cognitions/perceptions, in CBT terms) is reduced through either a) direct cognitive challenging, or b) increasing the frequency and resultant prediction-error precision of discrepant self-affirming experiences via behavioural activation[65], positive event diaries, savouring approaches[66], imagery work, or behavioural experiments[67]. The collated information can then be used not only to 'challenge' the content of maladaptive self-world models (schemas) and thus reduce their precision, but also to build alternative, more functional (Theory B[63]) models grounded in regularities across multiple self-affirming experiences.

As an alternative to seeking to change the content of maladaptive prediction models, third-wave therapies [68] such as mindfulness-based approaches[69] or Acceptance and Commitment Therapy[70] seek to change our relationship to the maladaptive models and their products, viewing them simply as an unwanted dysfunctional mindset that produces congruent mental events and interpretations, rather than as a set of fundamental personal truths. This 'decentering'[71] thus reduces the precision of these maladaptive models by nesting them within supra-ordinate models that devalue their veridicality. Third-wave approaches also seek to enhance the precision of prediction-errors arising from any model-discrepant new events by shifting the phenomenological centre down the model hierarchy, away from generalised context-wide mental content to specific, local experiences; for instance, by fostering a sense of being 'in-the-moment' during mindfulness[17, 69].

We have advocated elsewhere for a science of 'memory therapeutics'[72] whereby AM-based interventions, built around these three proposed therapeutic pathways, can provide accessible, low-intensity transdiagnostic alternatives to higher-intensity interventions such as CBT. The essence of these different mnemonic approaches (Box 3) is firstly to enhance the accessibility and salience of self-affirming and/or positive specific AMs through regimes of repeated practice, thus augmenting the precision of the prediction-errors they produce to modify maladaptive self-world prediction models. Second, this focus on bolstering AM specificity shifts the individual's phenomenological centre away from higher-order maladaptive prediction models and their congruent generalised autobiographical memories/themes toward specific, localised information. These 'specificity induction' techniques[73] also have the potential to augment established therapies such as CBT, where accessing model-disconfirming specific autobiographical material is critical. Finally, the recalled specific self-affirming and/or positive AMs can be collated together to generate novel self-affirming categorical themes to bolster alternative, adaptive higher-order self-world models (Theory B).

Our most established AM-focused interventions are trauma-focused therapies for posttraumatic stress[74]. Here, the salience of trauma memories and the impact of the resultant trauma-related prediction-errors to challenge pre-trauma adaptive self-world models are mitigated either behaviourally (through repeated imaginal exposure to memories) or cognitively by restructuring the trauma memory through the introduction of details which dilute the 'current threat' posed by the memory[37].

Box 3: Therapeutic efforts to target memory specificity

Low-intensity AM-based interventions aim to improve the specificity and salience of AMs, which in the proposed framework, would increase the precision weighting of their associated prediction-errors. Our prior meta-analysis[75] indicated that mnemonic interventions to enhance the ability to isolate specific, single incident events, and counter the tendency to retrieve

generalized negative summaries of the past, improved symptoms of low mood, posttraumatic stress, and anxiety (Figure II). In addition to generalised memory retrieval, mentally unwell individuals experience a reduction in the salience of positive AMs, indexed by lower self-report ratings of vividness[76] and underscored by blunted amygdala haemodynamic activity during retrieval[77]. Memory interventions targeted at improving the vividness and emotional intensity of positive AMs also improved symptoms, through either behavioural interventions (in which the participant is provided with repeated, cued-recall practice)[75] or using neurofeedback[77].

Figure II. *Forest plot of Cohen's ds for the between-group difference in change in depressive symptoms from pre- to post-intervention for a range of AM-based therapies versus control interventions.* MEST=Memory Specificity Training; COMET=Competitive Memory Training, Positive Imagery-CBM=Positive Imagery Cognitive Bias Modification. Adapted from[75].

More recently-developed interventions have also explored whether repeated, deliberate AM retrieval may be a way to foster alternate 'Theory B' positive self-world models or schemas (e.g.,[78]). For example, the Memory Flexibility (MemFlex) intervention trains flexible shifting between specific versus generalised AMs, and between AMs with positive versus negative emotional valence. MemFlex shows therapeutic promise for depressive symptoms[78] and posttraumatic stress[79]. Such interventions may potentially enhance the precision of more adaptive higher-order self-world models, in a simple, cost-effective manner, while also being less confrontational for patients who may feel that CBT tells them they are wrong to think the way they do.

Recent pharmacological and neuroimaging findings have advanced these interventions beyond behavioural techniques. A low dose of d-cycloserine appears to enhance specific-memory retrieval up to 24 hours later (proposedly by acting as a NMDAR glycine site agonist) in healthy individuals[80]. NMDAR glycine site antagonism has been previously proposed as a cost-effective target for mechanistically novel antidepressants[81] and these recent findings suggest that AM may

be one system through which such intervention may exert its effect. Similarly, amygdala neurofeedback-training during AM retrieval has been combined with CBT, yielding positive effects for depressive symptoms[77]. Such interventions represent exciting avenues for further research.

Proposed inclusion of AM in the National Institute of Mental Health (NIMH) Research Domain Criteria

Given this primary role in driving mental health, AM is a notable omission in research frameworks such as the Research Domain Criteria (RDoC). The RDoC framework identifies dimensional neurocognitive constructs that represent putative universal, transdiagnostic targets for mental wellbeing[7], and aims to bridge the basic and clinical science divide[82] to lead advances in treatment[83]. For inclusion, a proposed new construct (AM could be considered as a construct within the Cognitive Systems Domain) must, first, be measurable from molecular to self-report level. AM meets these criteria. AM processes can be indexed via molecules (NMDA), cells (grid and place cells), brain circuits (hippocampal, middle temporal and parahippocampal activation; amygdala connectivity with the salience network), physiology (protein synthesis; NMDA-related synaptic plasticity), behaviour (retrieval), self-report (diaries; experience sampling), and experimental paradigms (Autobiographical Memory Test or Interview[19]). Second, a proposed construct must predict functioning across the trajectory from health to illness. Third, there must be demonstrable interactions between the construct and the individual's environment. We have highlighted the multi-unit measurement of AM throughout this paper (with specific examples in Box 1), we therefore focus on the latter two criteria here.

In addition to supporting self-evidencing prediction models, AM also shapes interaction with the world during both mental illness and health by guiding problem solving[47], providing a heuristic for planning and imagining future events[45, 46], enabling us to foster intimacy in social relationships [48], and supporting emotion regulation [84-86]. AM disruptions are evident prior to the transition to illness – reduced access to specific AMs, and disrupted associated neural

connectivity during retrieval, are observed in at-risk samples[87, 88], and AM disruptions longitudinally predict the first onset of mental illness[89, 90].

AM function is also impacted by the external environment, via social influences such as Western versus collectivist culture [91, 92], the style with which our mothers recall their own autobiographical memories[93, 94], and exposure to stress[95] or trauma[96, 97]. For example, children exposed to maltreatment display reduced hippocampal and increased middle temporal and parahippocampal brain activation during recall of positive AMs, but increased amygdala activation and connectivity with the salience network during recall of negative AMs[98].

In sum, we submit there is now adequate evidence to warrant formal evaluation of AM for potential inclusion in the RDoC matrix. Thorough evaluation of AM via a workshop of interdisciplinary experts will aid advancement of RDoC as it moves into its second decade, and shape science-driven improvements of truly transdiagnostic psychiatric and psychological interventions.

Concluding remarks

We have outlined how Beckian cognitive theory and predictive processing frameworks could be integrated to provide an account of the primary role of AM in mental health. Formal mathematical and computational modelling, and further extensions of these proposals, are now warranted (see Outstanding Questions). This approach generates many interesting hypotheses amenable to testing within cognitive psychology, computational psychiatry[99], psychopharmacology[100], and cognitive neuroscience[12], as well as in the clinical sciences[71]. An integrated science of AM may offer a truly transdiagnostic approach to improving mental health.

1. Seligman, M.E.P., et al., *Navigating Into the Future or Driven by the Past.* Perspectives on Psychological Science, 2013. **8**(2): p. 119-141.
2. Killingsworth, M.A. and D.T. Gilbert, *A Wandering Mind Is an Unhappy Mind.* Science, 2010. **330**(6006): p. 932-932.
3. Conway, M.A. and C.W. Pleydell-Pearce, *The construction of autobiographical memories in the self-memory system.* Psychological Review, 2000. **107**(2): p. 261-288.
4. Conway, M.A., J.A. Singer, and A. Tagini, *The Self and Autobiographical Memory: Correspondence and Coherence.* Social Cognition, 2004. **22**(5): p. 491-529.
5. Clark, A., *Whatever next? Predictive brains, situated agents, and the future of cognitive science.* Behavioral and Brain Sciences, 2013. **36**(3): p. 181-204.
6. Holland, A.C. and E.A. Kensinger, *Emotion and autobiographical memory.* Physics of Life Reviews, 2010. **7**(1): p. 88-131.
7. Insel, T., et al., *Research domain criteria (RDoC): toward a new classification framework for research on mental disorders*. 2010, Am Psychiatric Assoc.
8. Von Helmholtz, H., *Handbuch der physiologischen Optik: mit 213 in den Text eingedruckten Holzschnitten und 11 Tafeln*. Vol. 9. 1867: Voss.
9. Pickering, A., *The cybernetic brain: Sketches of another future*. 2010: University of Chicago Press.
10. Powers, W.T. and W.T. Powers, *Behavior: The control of perception*. 1973: Aldine Chicago.
11. Friston, K., *Does predictive coding have a future?* Nature Neuroscience, 2018. **21**(8): p. 1019-1021.
12. Barrett, L.F., *The theory of constructed emotion: an active inference account of interoception and categorization.* Social Cognitive and Affective Neuroscience, 2017. **12**(1): p. 1-23.
13. Clark, J.E., S. Watson, and K.J. Friston, *What is mood? A computational perspective.* Psychological Medicine, 2018. **48**(14): p. 2277-2284.
14. Barrett, L.F., K.S. Quigley, and P. Hamilton, *An active inference theory of allostasis and interoception in depression.* Philosophical Transactions of the Royal Society B: Biological Sciences, 2016. **371**(1708): p. 20160011.
15. Badcock, P.B., et al., *The Depressed Brain: An Evolutionary Systems Theory.* Trends in Cognitive Sciences, 2017. **21**(3): p. 182-194.
16. Kube, T., et al., *Distorted Cognitive Processes in Major Depression: A Predictive Processing Perspective.* Biological Psychiatry, 2020. **87**(5): p. 388-398.
17. Manjaly, Z.-M. and S. Iglesias, *A Computational Theory of Mindfulness Based Cognitive Therapy from the "Bayesian Brain" Perspective.* Frontiers in Psychiatry, 2020. **11**(404).
18. Seth, A.K., *Interoceptive inference, emotion, and the embodied self.* Trends in Cognitive Sciences, 2013. **17**(11): p. 565-573.
19. Dalgleish, T. and M.J. Power, *The I of the storm—Relations between self and conscious emotion experience: Comment on Lambie and Marcel (2002).* Psychological Review, 2004. **111**(3): p. 812-818.
20. Pezzulo, G., F. Rigoli, and K.J. Friston, *Hierarchical Active Inference: A Theory of Motivated Control.* Trends in Cognitive Sciences, 2018. **22**(4): p. 294-306.
21. Bouizegarene, N., et al., *Narrative as active inference.* 2020.
22. Beck, A.T. and E.A.P. Haigh, *Advances in Cognitive Theory and Therapy: The Generic Cognitive Model.* Annual Review of Clinical Psychology, 2014. **10**(1): p. 1-24.
23. Matt, G.E., C. Vázquez, and W.K. Campbell, *Mood-congruent recall of affectively toned stimuli: A meta-analytic review.* Clinical Psychology Review, 1992. **12**(2): p. 227-255.
24. Williams, J., et al., *Autobiographical memory specificity and emotional disorder.* Psychol Bull, 2007. **133**(1): p. 122.
25. Monroe, S.M. and K.L. Harkness, *Life stress, the" kindling" hypothesis, and the recurrence of depression: considerations from a life stress perspective.* Psychological review, 2005. **112**(2): p. 417.

26. Post, R.M., D.R. Rubinow, and J.C. Ballenger, *Conditioning and Sensitisation in the Longitudinal Course of Affective Illness.* British Journal of Psychiatry, 2018. **149**(2): p. 191-201.
27. Brewin, C.R., T. Dalgleish, and S. Joseph, *A dual representation theory of posttraumatic stress disorder.* Psychological Review, 1996. **103**(4): p. 670-686.
28. Janoff-Bulman, R., *Shattered assumptions*. 2010: Simon and Schuster.
29. Haubrich, J., M. Bernabo, and K. Nader, *Noradrenergic projections from the locus coeruleus to the amygdala constrain fear memory reconsolidation.* eLife, 2020. **9**: p. e57010.
30. Vervliet, B., M.G. Craske, and D. Hermans, *Fear Extinction and Relapse: State of the Art.* Annual Review of Clinical Psychology, 2013. **9**(1): p. 215-248.
31. Kindt, M., M. Soeter, and B. Vervliet, *Beyond extinction: erasing human fear responses and preventing the return of fear.* Nature Neuroscience, 2009. **12**: p. 256.
32. Telch, M.J., et al., *Use of a Brief Fear Memory Reactivation Procedure for Enhancing Exposure Therapy.* Clinical Psychological Science, 2017. **5**(2): p. 367-378.
33. Kroes, M.C.W., et al., *Translational Approaches Targeting Reconsolidation*, in *Translational Neuropsychopharmacology*, T.W. Robbins and B.J. Sahakian, Editors. 2016, Springer International Publishing: Cham. p. 197-230.
34. McGaugh, J.L., *Memory--a Century of Consolidation.* Science, 2000. **287**(5451): p. 248-251.
35. Nader, K., G.E. Schafe, and J.E. Le Doux, *Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval.* Nature, 2000. **406**: p. 722.
36. Nader, K. and O. Hardt, *A single standard for memory: the case for reconsolidation.* Nature Reviews Neuroscience, 2009. **10**: p. 224.
37. Ehlers, A. and D.M. Clark, *A cognitive model of posttraumatic stress disorder.* Behaviour Research and Therapy, 2000. **38**(4): p. 319-345.
38. National Institute for Health and Care Excellence. *Recommendations for management of PTSD in children, young people and adults*. 2018; Available from: https://www.nice.org.uk/guidance/ng116/chapter/Recommendations#management-of-ptsd-in-children-young-people-and-adults.
39. Agren, T., et al., *Disruption of Reconsolidation Erases a Fear Memory Trace in the Human Amygdala.* Science, 2012. **337**(6101): p. 1550-1552.
40. Phelps, E.A. and S.G. Hofmann, *Memory editing from science fiction to clinical practice.* Nature, 2019. **572**(7767): p. 43-50.
41. Weisman, J.S. and T.L. Rodebaugh, *Exposure therapy augmentation: A review and extension of techniques informed by an inhibitory learning approach.* Clinical Psychology Review, 2018. **59**: p. 41-51.
42. Norberg, M.M., J.H. Krystal, and D.F. Tolin, *A Meta-Analysis of D-Cycloserine and the Facilitation of Fear Extinction and Exposure Therapy.* Biological Psychiatry, 2008. **63**(12): p. 1118-1126.
43. Otto, M.W., et al., *Enhancement of Psychosocial Treatment With D-Cycloserine: Models, Moderators, and Future Directions.* Biological Psychiatry, 2016. **80**(4): p. 274-283.
44. Iyadurai, L., et al., *Preventing intrusive memories after trauma via a brief intervention involving Tetris computer game play in the emergency department: a proof-of-concept randomized controlled trial.* Molecular Psychiatry, 2018. **23**(3): p. 674-682.
45. Jing, H.G., K.P. Madore, and D.L. Schacter, *Worrying About the Future: An Episodic Specificity Induction Impacts Problem Solving, Reappraisal, and Well-Being.* Journal of Experimental Psychology: General, 2016. **145**: p. 402-418.
46. Jing, H.G., K.P. Madore, and D.L. Schacter, *Preparing for what might happen: An episodic specificity induction impacts the generation of alternative future events.* Cognition, 2017. **169**: p. 118-128.
47. Goddard, L., B. Dritschel, and A. Burton, *Role of autobiographical memory in social problem solving and depression.* J Abnorm Psychol, 1996. **105**.

48. Beike, D.R., N.R. Brandon, and H.E. Cole, *Is sharing specific autobiographical memories a distinct form of self-disclosure?* Journal of Experimental Psychology: General, 2016. **145**(4): p. 434-450.
49. Klein, S.B., et al., *Priming exceptions: A test of the scope hypothesis in naturalistic trait judgments.* Social Cognition, 2001. **19**(4): p. 443-468.
50. Williams, J.M.G., et al., *Autobiographical memory specificity and emotional disorder.* Psychol Bull, 2007. **133**.
51. Barry, T.J., et al., *The Neurobiology of Reduced Autobiographical Memory Specificity.* Trends in Cognitive Sciences, 2018. **22**(11): p. 1038-1049.
52. Ono, M., G.J. Devilly, and D.H.K. Shum, *A meta-analytic review of overgeneral memory: The role of trauma history, mood, and the presence of posttraumatic stress disorder.* Psychological Trauma: Theory, Research, Practice, and Policy, 2016. **8**(2): p. 157-164.
53. Chiu, C.P.Y., et al., *Meta-analysis of the association between rumination and reduced autobiographical memory specificity.* Memory, 2018. **26**(10): p. 1323-1334.
54. Berntsen, D., *Involuntary autobiographical memories: An introduction to the unbidden past*. 2009: Cambridge University Press.
55. Brewin, C.R., et al., *Memory processes and the course of anxiety and depression in cancer patients.* Psychological Medicine, 1998. **28**(1): p. 219-224.
56. Brewin, C.R., et al., *Intrusive images in psychological disorders: characteristics, neural mechanisms, and treatment implications.* Psychological review, 2010. **117**(1): p. 210.
57. Brewin, C.R., M. Reynolds, and P. Tata, *Autobiographical memory processes and the course of depression.* Journal of Abnormal Psychology, 1999. **108**(3): p. 511-517.
58. Marks, E.H., A.R. Franklin, and L.A. Zoellner, *Can't get it out of my mind: A systematic review of predictors of intrusive memories of distressing events.* Psychological Bulletin, 2018. **144**(6): p. 584-640.
59. Reynolds, M. and C.R. Brewin, *Intrusive memories in depression and posttraumatic stress disorder.* Behaviour Research and Therapy, 1999. **37**(3): p. 201-215.
60. Steel, C., D. Fowler, and E.A. Holmes, *Trauma-Related Intrusions and Psychosis: An Information Processing Account.* Behavioural and Cognitive Psychotherapy, 2004. **33**(2): p. 139-152.
61. Hitchcock, C., C. Rees, and T. Dalgleish, *The devil's in the detail: Accessibility of specific personal memories supports rose-tinted self-generalizations in mental health and toxic self-generalizations in clinical depression.* Journal of Experimental Psychology: General, 2017. **146**(9): p. 1286-1295.
62. Benoit, R.G., P.C. Paulus, and D.L. Schacter, *Forming attitudes via neural activity supporting affective episodic simulations.* Nature Communications, 2019. **10**(1): p. 2215.
63. Salkovskis, P.M., *Understanding and treating obsessive—compulsive disorder.* Behaviour Research and Therapy, 1999. **37**: p. S29-S52.
64. Wenzel, A., K.S. Dobson, and P.A. Hays, *Cognitive behavioral therapy techniques and strategies*. 2016: American Psychological Association.
65. Uphoff, E., et al., *Behavioural activation therapy for depression in adults.* Cochrane Database of Systematic Reviews, 2020(7).
66. Dunn, B.D., et al., *Bringing light into darkness: A multiple baseline mixed methods case series evaluation of Augmented Depression Therapy (ADepT).* Behaviour Research and Therapy, 2019. **120**: p. 103418.
67. Bennett-Levy, J.E., et al., *Oxford guide to behavioural experiments in cognitive therapy*. 2004: Oxford University Press.
68. Dimidjian, S., et al., *Considering Meta-Analysis, Meaning, and Metaphor: A Systematic Review and Critical Examination of "Third Wave" Cognitive and Behavioral Therapies.* Behavior Therapy, 2016. **47**(6): p. 886-905.

69. Hofmann, S.G. and A.F. Gómez, *Mindfulness-Based Interventions for Anxiety and Depression.* Psychiatric Clinics, 2017. **40**(4): p. 739-749.
70. Twohig, M.P. and M.E. Levin, *Acceptance and Commitment Therapy as a Treatment for Anxiety and Depression: A Review.* Psychiatric Clinics, 2017. **40**(4): p. 751-770.
71. Bernstein, A., Y. Hadash, and D.M. Fresco, *Metacognitive processes model of decentering: emerging methods and insights.* Current Opinion in Psychology, 2019. **28**: p. 245-251.
72. Dalgleish, T. and A. Werner-Seidler, *Disruptions in autobiographical memory processing in depression and the emergence of memory therapeutics.* Trends in cognitive sciences, 2014. **18**(11): p. 596-604.
73. Madore, K.P. and D.L. Schacter, *Remembering the past and imagining the future: Selective effects of an episodic specificity induction on detail generation.* Quarterly Journal of Experimental Psychology, 2016. **69**(2): p. 285-298.
74. Lenz, A.S., A. Haktanir, and K. Callender, *Meta-Analysis of Trauma-Focused Therapies for Treating the Symptoms of Posttraumatic Stress Disorder.* Journal of Counseling & Development, 2017. **95**(3): p. 339-353.
75. Hitchcock, C., et al., *Autobiographical episodic memory-based training for the treatment of mood, anxiety and stress-related disorders: A systematic review and meta-analysis.* Clinical Psychology Review, 2017. **52**: p. 92-107.
76. Werner-Seidler, A. and M.L. Moulds, *Autobiographical memory characteristics in depression vulnerability: Formerly depressed individuals recall less vivid positive memories.* Cognition & emotion, 2011. **25**(6): p. 1087-1103.
77. Kymberly D. Young, Ph.D. ,, et al., *Randomized Clinical Trial of Real-Time fMRI Amygdala Neurofeedback for Major Depressive Disorder: Effects on Symptoms and Autobiographical Memory Recall.* American Journal of Psychiatry, 2017. **174**(8): p. 748-755.
78. Hitchcock, C., et al., *A randomised controlled trial of memory flexibility training (MemFlex) to enhance memory flexibility and reduce depressive symptomatology in individuals with major depressive disorder.* Behaviour Research and Therapy, 2018. **110**: p. 22-30.
79. Moradi, A.R., et al., *Proof-of-concept for the autobiographical Memory Flexibility (MemFlex) intervention for Posttraumatic Stress Disorder* 2020.
80. Chen, R., et al., *Effect of the NMDA receptor partial agonist, d-cycloserine, on emotional processing and autobiographical memory.* Psychological Medicine, 2020: p. 1-9.
81. Heresco-Levy, U., et al., *A randomized add-on trial of high-dose d-cycloserine for treatment-resistant depression.* International Journal of Neuropsychopharmacology, 2013. **16**(3): p. 501-506.
82. Cosgrove, V.E., J.R. Kelsoe, and T. Suppes, *Toward a Valid Animal Model of Bipolar Disorder: How the Research Domain Criteria Help Bridge the Clinical-Basic Science Divide.* Biological Psychiatry, 2016. **79**(1): p. 62-70.
83. Yee, C.M., D.C. Javitt, and G.A. Miller, *Replacing DSM Categorical Analyses With Dimensional Analyses in Psychiatry Research: The Research Domain Criteria Initiative.* JAMA Psychiatry, 2015. **72**(12): p. 1159-1160.
84. Joormann, J. and M. Siemer, *Memory accessibility, mood regulation, and dysphoria: Difficulties in repairing sad mood with happy memories?* Journal of abnormal psychology, 2004. **113**(2): p. 179.
85. Speer, M.E. and M.R. Delgado, *Reminiscing about positive memories buffers acute stress responses.* Nature Human Behaviour, 2017. **1**: p. 0093.
86. Joormann, J., M. Siemer, and I. Gotlib, *Mood regulation in depression: differential effects of distraction and recall of happy memories on sad mood.* J Abnorm Psychol, 2007. **116**(3): p. 484.
87. Young, K.D., et al., *Behavioral and neurophysiological correlates of autobiographical memory deficits in patients with depression and individuals at high risk for depression.* JAMA Psychiatry, 2013. **70**: p. 698-708.

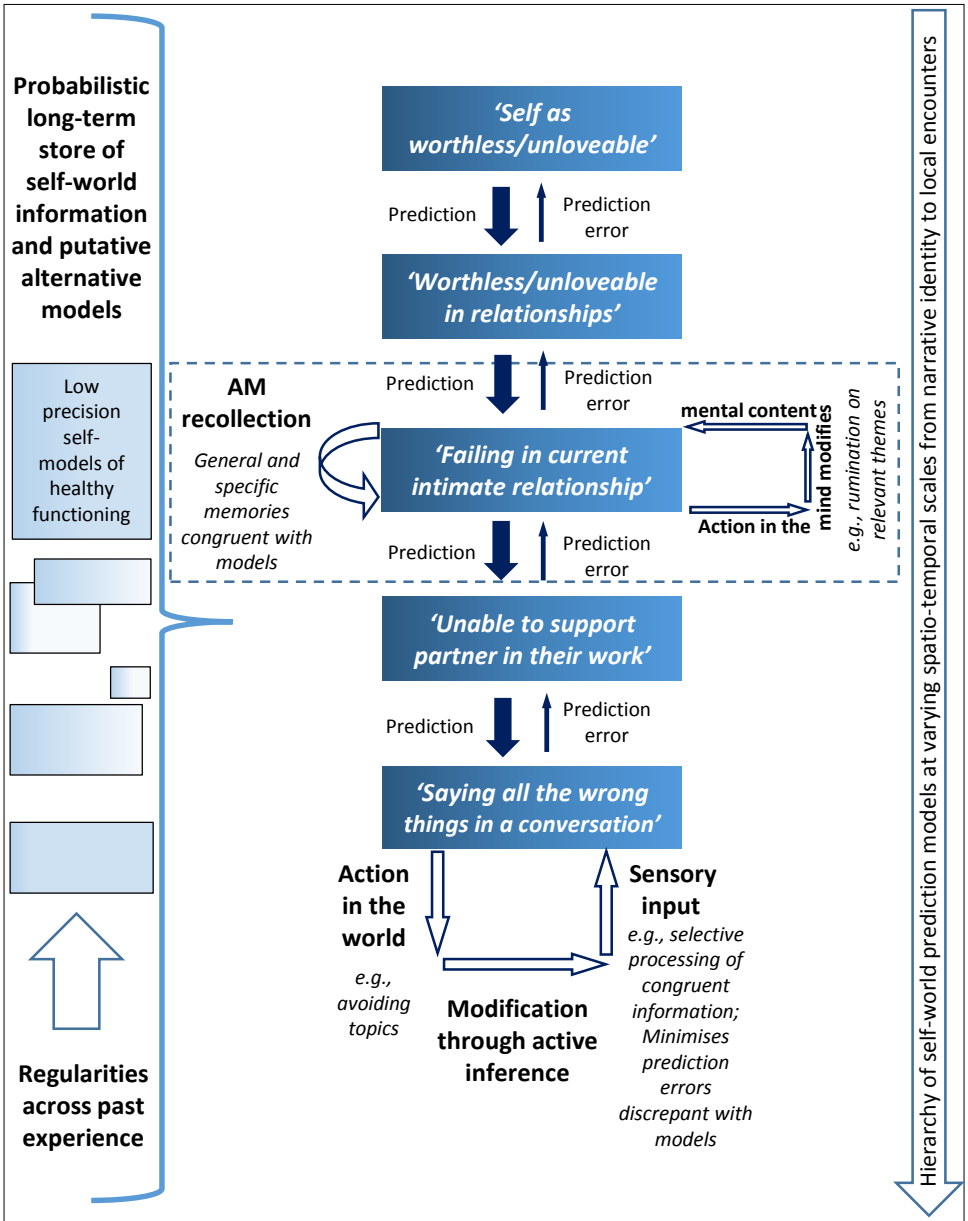
88. Macdonald, B., et al., *Altered engagement of autobiographical memory networks in adult offspring of postnatally depressed mothers.* Biological Psychology, 2016. **118**: p. 147-153.
89. Gutenbrunner, C., K. Salmon, and P.E. Jose, *Do Overgeneral Autobiographical Memories Predict Increased Psychopathological Symptoms in Community Youth? A 3-Year Longitudinal Investigation.* Journal of Abnormal Child Psychology, 2018. **46**(2): p. 197-208.
90. Kuyken, W. and T. Dalgleish, *Overgeneral autobiographical memory in adolescents at risk for depression.* Memory, 2011. **19**(3): p. 241-250.
91. Zaragoza Scherman, A., et al., *Life Script Events and Autobiographical Memories of Important Life Story Events in Mexico, Greenland, China, and Denmark.* Journal of Applied Research in Memory and Cognition, 2017. **6**(1): p. 60-73.
92. Jobson, L., et al., *Impact of culture on autobiographical life structure in depression.* British Journal of Clinical Psychology, 2018. **57**(3): p. 382-396.
93. Fivush, R., *Maternal reminiscing style: The sociocultural construction of autobiographical memory across childhood and adolescence.* The Wiley handbook on the development of children's memory, 2013. **1**: p. 568-585.
94. Reese, E. and R. Newcombe, *Training mothers in elaborative reminiscing enhances children's autobiographical memory and narrative.* Child development, 2007. **78**(4): p. 1153-1170.
95. Askelund, A.D., et al., *Positive memory specificity is associated with reduced vulnerability to depression.* Nature Human Behaviour, 2019.
96. Hitchcock, C., R.D.V. Nixon, and N. Weber, *A review of overgeneral memory in child psychopathology.* British Journal of Clinical Psychology, 2014. **53**(2): p. 170-193.
97. Moore, S.A. and L.A. Zoellner, *Overgeneral autobiographical memory and traumatic events: an evaluative review.* Psychological bulletin, 2007. **133**(3): p. 419.
98. McCrory, E.J., et al., *Autobiographical memory: A candidate latent vulnerability mechanism for psychiatric disorder following childhood maltreatment.* British Journal of Psychiatry, 2018. **211**(4): p. 216-222.
99. Jang, A.I., et al., *Positive reward prediction errors during decision-making strengthen memory encoding.* Nature Human Behaviour, 2019. **3**(7): p. 719-732.
100. Papalini, S., T. Beckers, and B. Vervliet, *Dopamine: from prediction error to psychotherapy.* Translational Psychiatry, 2020. **10**(1): p. 164.

Figure 1 Legend

Figure depicts a hierarchy of active self-world prediction models across a range of spatio-temporal scales, with the broadest contextual models at the top, capturing stable aspects of identity, down to local models reflecting specific self-event relationships. The models are generative, coding the likely causes of inputs from the internal and external worlds, through a system of Bayesian inference. The precision of the models – an inverse variance weighting based on their predictive reliability – is signified by *darker shading* relative to competing models in long-term storage or probability space (on the left). Discrepancies between predictions and inputs from the world (salience) are coded as prediction-errors. Prediction-errors ascend through the hierarchy signalling potential modifications to the models. Prediction-errors also vary in their precision with more precise errors have a greater capacity to modify the prediction models. Relative precision of descending predictions and ascending errors is coded by the *thickness of the arrows*. Prediction models are self-evidencing so cognition (perception, attention, memory etc.) and action in the internal and external worlds (active inference) act to confirm prior predictions. The phenomenological centre is represented by the *dashed box*, and reflects the level within the hierarchy at which the active models are driving the contents of current mental experience. The positioning of the phenomenological centre determines the granularity of mental experience from generic, narrative personal themes at higher levels down to specific memories or cognitions. The *italicised notations* represent a worked example of a depressogenic mindset with exemplar content for different components of the framework.

Figure 1 (Key Figure)

Figure 1. *A hierarchical predictive processing framework for understanding autobiographical memory and mental health.*



Legend

Figure depicts a hierarchy of active self-world prediction models across a range of spatio-temporal scales, with the broadest contextual models at the top, capturing stable aspects of identity, down to local models reflecting specific self-event relationships. The models are generative, coding the likely causes of inputs from the internal and external worlds, through a system of Bayesian inference. The precision of the models – an inverse variance weighting based on their predictive reliability – is signified by *darker shading* relative to competing models in long-term storage or probability space (on the left). Discrepancies between predictions and inputs from the world (salience) are coded as prediction-errors. Prediction-errors ascend through the hierarchy signalling potential modifications to the models. Prediction-errors also vary in their precision with more precise errors have a greater capacity to modify the prediction models. Relative precision of descending predictions and ascending errors is coded by the *thickness of the arrows*. Prediction models are self-evidencing so cognition (perception, attention, memory etc.) and action in the internal and external worlds (active inference) act to confirm prior predictions. The phenomenological centre is represented by the *dashed box*, and reflects the level within the hierarchy at which the active models are driving the contents of current mental experience. The positioning of the phenomenological centre determines the granularity of mental experience from generic, narrative personal themes at higher levels down to specific memories or cognitions. The *italicised notations* represent a worked example of a depressogenic mindset with exemplar content for different components of the framework.

Figure I to be included in Box 2

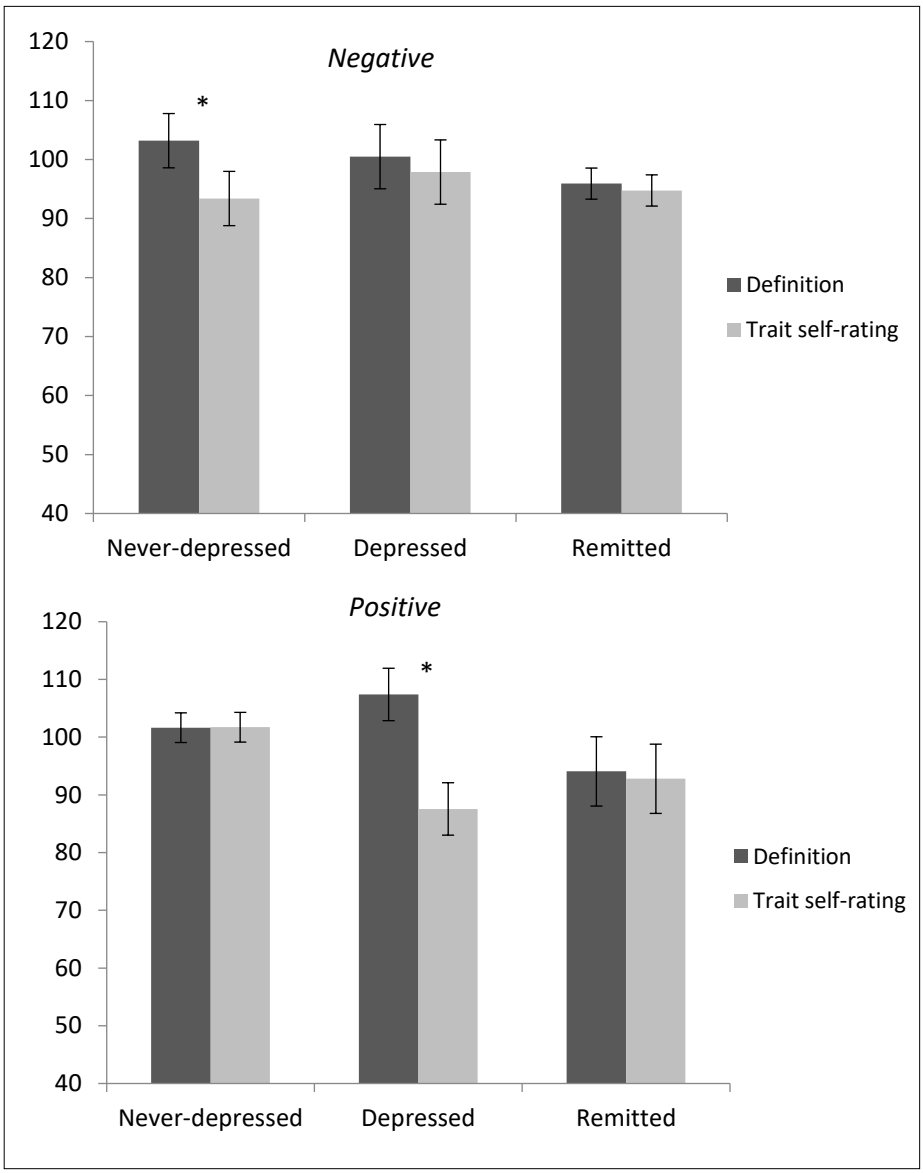


Figure I. *Mean (square-root transformed) response time (ms; y-axis) for recall of an inconsistent specific memory after rating how well a negative (top) or positive (bottom) personality trait describes the self (trait self-rating), relative to a dictionary definition control condition.* Taken from [62]. Note. *=p<.05. Error-bars= standard error. Shorter response times for trait self-rating relative to definition represents the priming effect.

Figure II to be included in Box 3

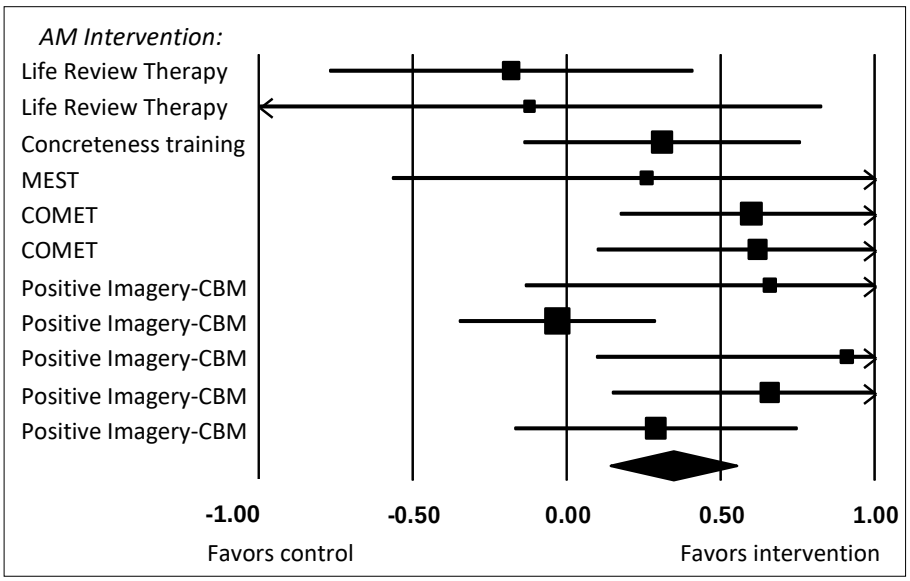


Figure II. *Forest plot of Cohen's ds for the between-group difference in change in depressive symptoms from pre- to post-intervention for a range of AM-based therapies versus control interventions.*
MEST=Memory Specificity Training; COMET=Competitive Memory Training, Positive Imagery-CBM=Positive Imagery Cognitive Bias Modification. Adapted from[76].

Outstanding Questions Box

- The current predictive processing approach provides a framework to examine relationships between autobiographical memory, mental health and therapy. How can this approach be further integrated with predictive processing models of core components of affect (mood, emotion, interoception) that are relevant to mental health?
- Can this broad integrated approach be computationally formalised?
- What are the neural networks in the brain that support these diverse aspects of autobiographical memory functioning?
- How is the encoding, consolidation, and reconsolidation of autobiographical memories impacted by the nature of current self-world models? And does the reconsolidation of memories modify the model priors?
- What are the optimal ways to augment existing evidence-based therapies such as cognitive behaviour therapy with lower-intensity autobiographical memory interventions, to maximize therapeutic gain?
- Is there now a sufficiently compelling case for the inclusion of AM within the NIMH Research Domain Criteria matrix?

Highlights box

- Autobiographical memory is essential to healthy mental functioning, such that disruptions to autobiographical memory predict the onset, duration, and relapse of a range of common mental health issues.
- We propose a transdiagnostic framework for approaching autobiographical memory that integrates a predictive processing framework with therapetic models within clinical psychology and psychiatry.
- Autobiographical memory has been largely overlooked by leading research frameworks such as the NIMH Research Domain Criteria. We submit that there is now sufficient evidence to warrant formal evaluation of autobiographical memory for inclusion in such frameworks.
- As such, we advocate the integration of neurobiological, computational, pharmacological, and psychological approaches to the cognitive science of autobiographical memory.