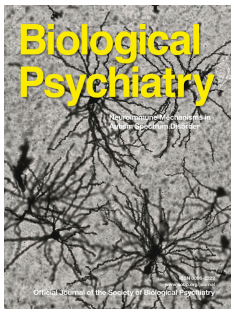
Journal Pre-proof



Spatiotemporal precision of neuroimaging in psychiatry

Jessica McFadyen, Raymond J. Dolan

PII: S0006-3223(22)01525-6

DOI: https://doi.org/10.1016/j.biopsych.2022.08.016

Reference: BPS 14965

To appear in: *Biological Psychiatry*

Received Date: 9 May 2022

Revised Date: 20 July 2022

Accepted Date: 12 August 2022

Please cite this article as: McFadyen J. & Dolan R.J, Spatiotemporal precision of neuroimaging in psychiatry, *Biological Psychiatry* (2022), doi: https://doi.org/10.1016/j.biopsych.2022.08.016.

Spatiotemporal precision of neuroimaging in psychiatry

Jessica McFadyen [*1, 2] & Raymond J Dolan [1, 2, 3]

[1] The UCL Max Planck Centre for Computational Psychiatry and Ageing Research, University College London, London, UK

[2] Wellcome Centre for Human Neuroimaging, University College London, London, UK

[3] State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, 100875, China

* corresponding author: j.mcfadyen@ucl.ac.uk

Keywords: MEG, decoding, replay, machine learning, representations, neuroimaging

Short title: Spatiotemporal precision of neuroimaging in psychiatry

Abstract

Aberrant patterns of cognition, perception, and behaviour seen in psychiatric disorders are thought to be driven by a complex interplay of neural processes that evolve at a rapid temporal scale. Understanding these dynamic processes in vivo in humans has been hampered by a trade-off between the spatial and temporal resolution inherent to current neuroimaging technology. A recent trend in psychiatric research has been the use of high temporal resolution imaging, particularly magnetoencephalography (MEG), often in conjunction with sophisticated machine learning decoding techniques. Developments here promise novel insights into the spatiotemporal dynamics of cognitive phenomena, including domains relevant to psychiatric illness such as reward and avoidance learning, memory, and planning. This review considers recent advances afforded by exploiting this increased spatiotemporal precision, with specific reference to applications the seek to drive a mechanistic understanding of psychopathology and the realisation of preclinical translation.

An important goal within cognitive neuroscience is to determine the precise neurophysiological features that contribute to the expression of psychiatric phenomena, with an ultimate goal to inform psychiatric diagnosis and treatment. Given the multitude of neuroimaging tools accessible to researchers today, it may seem surprising that neuroimaging research has had scant impact on clinical psychiatry (1,2). Several non-competing explanations have been put forward (3), pointing to either the historical limitations of neuroimaging analyses and their interpretation (4–9), or to the restrictive, subjective, and arbitrary nature of clinical diagnosis (6,8,10). Here, we focus on the former. We argue that the utility of neuroimaging in psychiatry has reached an inflection point upon which recent methodological advancements can now dramatically improve the spatiotemporal precision of functional brain mapping, opening new approaches to elucidating the neurocognitive dynamics underlying complex human behaviour and psychopathology.

Our ability to precisely capture spatiotemporal patterns of neural activity has, until recently, been limited by two primary obstacles. One relates to a trade-off between spatial and temporal resolution that is inherent to a reliance on non-invasive neuroimaging approaches. This limits the ability of any single methodology to provide a complete picture of both the “where” and “when” of the neural processes that underlie complex human cognition and behaviour, potentially obscuring core aspects of neural dynamics that play causal roles in the genesis of psychiatric illnesses.

A second obstacle is the extent to which it is possible to ascribe precise mechanistic significance to *in vivo* recorded brain activity; in other words, the “what” and “how” of a neural process. For example, increased blood-oxygen-level dependent (BOLD) signal in the striatum after receipt of a reward is interpreted as indicating a functional role for this structure in reward processing, but this observation lacks specificity as to what that functional role actually is (11). Mechanistic specificity can be gained from designing highly controlled experiments that attempt to isolate a precise cognitive function, usually informed by a computational model, though this often entails reduced ecological validity and generalisability (12,13).

The dynamic nature and real-world relevance of features that characterise psychiatric disorders mean that both spatiotemporal and functional precision are crucial to improving our understanding and, ultimately, guiding development of targeted treatments (14). In this review, we outline current trends in human neuroimaging that advance a quest for increased spatiotemporal precision. First, we provide an overview of the current spatiotemporal resolution achievable in neuroimaging. Second, we illustrate how to enhance spatiotemporal precision by extracting meaningful state representations from neuroimaging data, as well as how to track the dynamic reinstatement of these processes in the brain, taking recent breakthroughs in

the detection of hippocampal replay using magnetoencephalography (MEG) as a case example. Finally, we explore how uncovering the spatiotemporal dynamics of mechanistically-relevant neural activity can be combined with generative modelling of pathological behaviour and cognition, with specific relevance to the burgeoning field of computational psychiatry (15).

Spatiotemporal precision of neuroimaging

Non-invasive neuroimaging methods range from modern ultra-high-field MRI that delivers a spatial resolution as fine as 0.5 millimetres (16), to older technologies such as electroencephalography (EEG) and MEG that provide measurements of mass neural activity at a millisecond resolution (17,18). Each of these modalities have strengths and weaknesses with regards to spatial and temporal resolution, in addition to factors such as tolerance in freedom of movement (19) and the precise physiological processes used to index neural activity.

In psychiatry, it can be conjectured that processes underlying psychopathology encompass rapidly-evolving *and* spatially-specific neural dynamics. For example, disordered belief formation in schizophrenia has been ascribed to aberrant activity in prefrontal cortex and hippocampus related to reduced synaptic gain, causing an imprecise coding of prior beliefs which, in turn, influences neural responses to surprising stimuli as early as 50 ms post-stimulus onset (11). Similarly, depression has been thought of as a “disconnection” syndrome, where connectivity between anatomically-discrete brain regions is reduced (20,21) but where the rapid, dynamic evolution of this connectivity (i.e., sub-second transient changes in distinct spatial neuronal populations) differ between clinical subtypes (22,23), providing a potential biomarker for the efficacy of electroconvulsive therapy (24). Thus, despite apparent progress using conventional approaches it is nevertheless the case that fundamental research questions related to neural dynamics likely require a level of spatiotemporal precision that has historically been extremely difficult to realise (25).

Multimodal imaging

Considerable effort has been invested in attaining higher spatiotemporal precision by deriving converging results from separate neuroimaging methodologies with complementary spatial and temporal resolutions, either recorded simultaneously (e.g., simultaneous EEG-fMRI) or in separate sessions (e.g., MEG, followed by fMRI) (26). In many cases, this multimodal approach to neuroimaging has been informative about brain dynamics underlying psychopathology (27). For instance, the amplitude of a fast-latency signature of

reward processing detected with EEG correlates with BOLD signal in striatum and, together, this fast striatal reward responsivity is reported as blunted in a subtype of depression characterised by impaired mood reactivity (28). Thus, multimodal imaging has the potential to enhance detectability of subtle, neurobiological effects that would otherwise be difficult to detect through reliance on a single modality (26,29). Multimodal imaging studies, however, impose a significantly higher demand on resources, and a lack of a unifying model can lead to difficulties with interpreting convergent or discrepant multimodal findings (27,30).

Increasing granularity using statistical learning

A recently developed approach to enhancing spatiotemporal precision of a single neuroimaging modality involves the exploitation of machine (or "statistical") learning, which harnesses a range of statistical techniques to distinguish between neural or behavioural states. This approach has demonstrated that even the most nuanced fluctuations in spatiotemporal neural data may contain relevant information (31). These nuances, such as small differences in the angle of neighbouring dipoles in MEG data, create statistically-separable patterns that are identifiable using multivariate pattern classification algorithms.

An early example of a machine learning approach to neuroimaging data involved decoding visual orientation from human visual cortex using multi-voxel pattern analysis (MVPA) of functional MRI data (32). Although orientation-selective cortical columns are much smaller than the spatial resolution of functional MRI (3 mm^3), orientation selectivity can be reliably estimated from signals generated by entire ensembles of voxels. Remarkably, orientation selectivity (33) and retinotopic maps in primary visual cortex (34) have now been reliably estimated from MEG data using support vector machine (SVM) classifiers, despite source-reconstructed MEG having a resolution in the order of several millimetres at the cortical surface. This example demonstrates that modern analytic approaches can exploit subtle variation in coarse spatial or temporal information to detect, and classify, neural processes that unfold at a finer scale than the resolution of the imaging modality itself. Such a feat can be achieved by biology-agnostic machine learning methods that distil spatiotemporal information from rich sources of neuroimaging data (as just described), and also by biophysically-realistic models that utilise prior knowledge of neurophysiological activity (provided by other modalities; e.g., invasive electrophysiological recordings in animals), to capture traces of such processes present in non-invasive human neuroimaging data (e.g., dynamic causal modelling of fMRI and M/EEG; (35)). Thus, both biologically-informed models and biology-agnostic machine learning methods can be used to offset spatiotemporal constraints of current neuroimaging methodologies.

Hippocampal replay as a case example

A striking example of the use of statistical learning to extract precise spatiotemporal information from MEG data comes from pioneering studies demonstrating hippocampal replay in humans (43). A central tenet of this review is that non-invasive measurement of hippocampal replay in humans is likely to represent a major advance not only for cognitive neuroscience but also biological psychiatry. The approach indicates that neuroimaging data can provide a sufficiently rich source of spatiotemporal information to signal rapid, dynamic, shifts in mental states, thereby allowing for a more precise estimate of when and where cognitive processes unfold in the brain. Below, we detail this approach and discuss how it has been, and can be, exploited to further the field of biological psychiatry.

The methodological challenge of replay

Replay was first observed in rodents in the 1990s where, during post-task rest, hippocampal place cells indexing the trajectory of an animal through an environment rapidly reactivated in the same order in which these locations were experienced, albeit with a pronounced temporal compression (44–46). This spontaneous and rapid unfolding activity pattern was subsequently shown to play a causal role in memory consolidation (47–50), and has been linked to higher-order cognitive functions such reward learning (51–57) and planning (58–62).

In humans, measuring hippocampal replay non-invasively presents a considerable methodological challenge, as one of its putative source (the hippocampus) is located deep within the brain, and the speed with which replay events unfold is extremely fast (in animals, the sequential reactivation of place cells indexing discrete locations is typically separated by tens of milliseconds). This challenge is shared by neuroimaging research in psychiatry, where there is often a need for both spatial and temporal precision. For example, in mood disorders, fast latency activity in deep brain structures, such as the amygdala, allegedly play a pivotal role in the genesis and maintenance of symptoms but is notoriously difficult to measure in vivo (25). Moreover, replay by its very nature involves reactivation of anatomically-specific neural populations (e.g., specific place cells) that represent specific mental states (e.g., different locations in space). Thus, measuring replay in humans from non-invasive neuroimaging data necessitates innovative approaches, such as the exploitation of statistical learning to extract fast sequential reactivation of state representations (37,63).

Measuring hippocampal replay

An approach to quantifying replay from non-invasive neuroimaging data is Temporally Delayed Linear Modelling (TDLM) (37), which estimates evidence for sequential state reactivation. TDLM capitalises on the fact that reactivation of a particular state within the hippocampus causes a cascade of related activity across a distributed network that includes the entorhinal cortex (64), medial temporal cortex (65), visual cortex (66), and prefrontal cortex (67–70). Thus, while hippocampal activity can be challenging to identify from MEG recordings (but far from impossible: see 71,72), information related to a specific memory or state can be decoded from unique spatial patterns of neural activity to uncover rapid, sequential reactivation of prior experiences (63,73–79). This ability to detect subtle but relevant spatial information increases both temporal and representational precision (e.g., specific memories) even at relatively low spatial resolution. Importantly, in psychiatry research, representational precision might often be considered more valuable than spatial precision, such as when investigating whether a therapeutic intervention instantiates a change in cognitive processes.

How can specific states be isolated and captured? Investigators commonly use visual stimuli presented in a particular order to represent distinct “states”. A key idea here is that the brain organises information — spatial or otherwise — into “cognitive maps” constructed from information like conceptual associations or temporal-order relationships (39). By using visually- and conceptually-unique images, machine learning algorithms can accurately and reliably classify spatial patterns of neural activity associated with viewing each image (**Fig. 1A**). The sheer size of the visual system in the human brain means that visual stimuli can be classified from distributed spatiotemporal activity generated primarily from occipital and temporal cortices, with classification accuracy typically in the range of 37% to 50%, which is 3 to 8 times higher than what would be expected by chance (74,76,78,80). Classifiers are typically trained on neural activity patterns recorded during an initial functional localiser, when participants view images before learning about task-related temporal-order relationships (37). Hence, this constitutes a “supervised” machine learning approach, where identity labels are known (e.g., whether participants were viewing image A or image B). The associated MEG sensor patterns then provide a reliable estimate of activity when these states are subsequently reactivated, for example during a cognitive task such as planning (online) or during a rest period (offline) (**Fig. 2**). Both hippocampus and medial temporal lobe, as well as visual cortex, have been identified as likely sources of such replay events in humans (74,75,78).

Overall, investigating replay in the human brain exemplifies how a rapidly-evolving neurophysiological signal can be detected and characterised at an extremely fine temporal resolution. More importantly, these studies provide a representational specificity (e.g., states in a cognitive map) that is not easily obtained using traditional neuroimaging analyses. This implies that a "representation-rich" characterisation of neuroimaging data can greatly enhance the granularity of observable neural dynamics (43), allowing exploration of more abstract neural processes underlying complex cognition.

Mechanistic specificity

Computational modelling of behaviour

The ability to uncover hidden spatiotemporal dynamics of cognition from neuroimaging data has the potential to unlock crucial information about psychiatric disorders that might otherwise be undetectable from behaviour alone. As an example, consider the cognitive processes that contribute to planning. These include an ability to learn and retrieve a cognitive model of the environment that captures how states are connected, the consequences of taking different actions at different states, and the effective appraisal of prospective reward and loss (81). Computations such as these evolve dynamically over time, where one type of processing (e.g., the accessibility of an aversive memory) may influence another (e.g., the perceived probability of being punished) (82). These dynamics are pervasive in existing computational psychiatry models of behaviour, which reveal information about how specific cognitive mechanisms operate differently in psychiatric disorders (83).

Spatiotemporally-precise neuroimaging can bestow cognitive models with biological plausibility, revealing how modelled dynamics of cognition (where cognition is either a construct, as in algorithmic models like reinforcement learning, or a biophysically-realistic process, as in synthetic models like attractor network models) are supported by the temporal profile of network activity (84). Therefore, it seems reasonable to conjecture that clinical translation of computational psychiatry may be catalysed by approaches to neuroimaging analysis that enhance spatiotemporal precision by: a) validating the dynamics of theory-driven cognitive processes through convergent biological evidence, b) assigning a neurophysiological basis to modelled cognitive mechanisms, potentially revealing targets for treatment, and c) enhancing the informational content of models by revealing hidden states. Below, we describe recent studies that pair spatiotemporally-precise neuroimaging, such as sequential state reactivation during replay, with computational psychiatry models, with a particular focus on structural inference and reward learning.

Inferring environment structure

Decoded state representations shed light on how we learn, store, and retrieve structured representations of our environment. The spontaneous reactivation of sequences — both experienced and imagined — is implicated in constructing and utilising internal representations of the environment. For instance, an ordered reactivation of previously-experienced states during a post-task rest period has been shown to correspond not to an experienced structure, but instead to an *inferred* structure that participants abstracted based on a learned task rule (76,78). This sensitivity of reactivated state representations to inferred structural features implies that MEG-decoded replay can provide a neurobiological signature of an ability to structurally reorganise our model of the world.

A breakdown in structural inference has been conjectured to underlie psychiatric symptoms that indicate inflexible or repetitive thinking, including compulsive behaviour in obsessive-compulsive disorder, detrimental drug consumption in addiction disorders, and incoherent thought in schizophrenia (85–89). This accords with findings of relatively stronger evidence for model-free decision-making (i.e., habitual behaviour that disregards environment structure), compared to model-based control (i.e., deliberate behaviour that grants flexibility and accuracy at the cost of increased cognitive load) (89), in these clinical populations.

In schizophrenia, we can ask whether a putative deficit in structural inference is reflected in spontaneous neural replay. After completing a task in which the temporal order of a stimulus sequence needs to be inferred, even though the “true” order is never experienced, patients with schizophrenia show weaker evidence for reorganisation of ordered state reactivation during rest compared with healthy controls, an effect that localises to hippocampus and corresponds with behaviour (78). This finding is consistent with a theory of reduced synaptic gain in schizophrenia, which is thought to significantly impact synaptic plasticity and attractor dynamics within hippocampus (90–92). This points to a link between an observable cognitive process (impaired structural inference, possibly manifesting as incoherent thought) and a previously unobservable neurophysiological process (replay of an inferred cognitive map in hippocampus) that might guide prognosis, as well as pharmacological and therapeutic treatment (90).

Making inferences under uncertainty

A feature of several psychiatric disorders is an impaired ability to update beliefs about the structure of an environment when something changes unexpectedly. For instance, behavioural modelling of decision-making has shown that paranoia and delusions can be explained by having a general expectation that

stimulus-outcome contingencies will change more frequently, resulting in poorer learning in volatile environments (93–97). This translates to an overweighting of unlikely explanations (i.e., paranoid delusions), the quality of which depends on a complex interplay of other parameters such as mood, prior habits, and whether beliefs pertain to social interaction (95).

Dysfunctional belief updating is a target of cognitive behavioural therapy (CBT), which reports success in correcting beliefs about risk and uncertainty in the context of obsessive-compulsive (OCD) disorder (98), as well as in reducing negative beliefs in depression through “cognitive restructuring” methods (99). There are, however, instances where CBT inexplicably fails, such as with the long-term persistence of paranoid delusions (100) and with treatment resistance in specific subtypes of OCD (101), even when administered alongside pharmacotherapy. The ability to derive a precise neural signature of how beliefs evolve over time, much in the same way that state representations are decoded to indicate neural replay (37), can in principle help reveal whether cognitive restructuring in CBT is having a significant impact on generative processes throughout the course of treatment, potentially serving also as a post-treatment predictor of relapse.

Research on healthy participants has demonstrated that dynamic belief updating can indeed be detected via spatiotemporal decoding of MEG data. Weiss et al. (2021) investigated the computational and neural mechanisms underlying structural inference in uncertain environments with and without an ability to control how information was sampled (102). They found that being able to choose which information to sample made environments appear more stable, echoing beliefs people with OCD hold about compulsive and repetitive behaviours (103). Moreover, MEG pattern classification revealed crucial temporal and spatial dynamics of how evidence was evaluated against current beliefs during information gathering. Specifically, activity in temporal and visual cortex encoded how consistent each piece of evidence was with current beliefs, revealing changes of mind that occurred throughout a trial prior to making a response. These changes of mind were delayed when participants had control over information sampling, consistent with participants reportedly viewing these environments as being more stable. This work elegantly demonstrates how neural pattern classification can reveal temporally-precise trajectories of beliefs with a neuroanatomical grounding, which could provide novel information about such cognitive processes in conditions such as OCD (102,104).

Tracking the dynamics of reward learning

Disordered belief updating leads to dysfunctional decision-making, which is a cause of disruption to everyday life in people with certain psychiatric disorders (88). In mood disorders, a bias towards using

negative information to update beliefs (which we can consider analogous to "learning") (105) can be computationally deduced (e.g., by reinforcement learning models) from patterns of dysfunctional decision-making, such as increased risk aversion in anxiety and reduced reward-seeking behaviour in depression (88). Neuroimaging can complement such computational models of decision-making in psychopathology by measuring a "reward prediction error" signal (i.e., the difference between the reward that was received and the reward that was expected), a key computational component in reinforcement learning and active inference models (106). Reward prediction error signals localise to specific neurochemical circuitry (e.g., dopaminergic pathways) and are observable in both M/EEG (107,108) and fMRI (109).

Reward prediction error signals, detected with fMRI, accurately predict response to CBT in depression, where an increased responsivity of amygdala and striatum to unexpected rewards has been interpreted as indicating a susceptibility to subsequent belief updating during cognitive restructuring during CBT (110). In contrast, reward prediction errors derived from computational modelling of behaviour alone have not yet been shown to predict treatment response, highlighting the power of mechanism-focused neuroimaging analysis for detecting subtle but clinically relevant effects. Extending this, we might consider that belief updating occurs not only at outcome receipt (when reward prediction errors occur) but also in anticipation of an event (e.g., worrying about the future in anxiety) (82) and when recollecting and re-interpreting past events (e.g., rumination in depression or "post-event processing" in social anxiety) (82,111). Uncovering hidden *temporal* dynamics of belief updating could broaden our understanding of how events are evaluated and deliberated upon before and after decision-making, potentially enabling a closer mapping to specific symptoms such as rumination and worry.

In animals, understanding the temporal dynamics of reward learning has benefited from machine learning. An elegant example is that of Rich and Wallis (2016), who used linear discriminant analysis (LDA) to capture patterns of neural firing in OFC corresponding to four potential choice options, each represented by unique images. While the animals deliberated on their choice, neural activity patterns in OFC alternated approximately every 230 ms between the chosen and unchosen option at each trial, with the chosen option becoming increasingly decodable across deliberation time. This also corresponded to fewer switches towards an unchosen option, as well as faster decision-making and less deliberation (112). Building on this, recent studies have classified patterns of activity in OFC that represent not only the dynamics of outcome representations, but also features such as task structure (e.g., preconditioned associations between states, predictions of upcoming states) and the expected reward value of each state (113–115).

Tracking representations of reward over time provide added value to computational models of decision-making. For example, Eldar et al. (2018) investigated whether a person's mood relates to differences in receptivity to reward, a process thought to play a significant role in the onset of depression and bipolar disorder (116–118). Here, reinforcement learning models suggested two underlying mechanisms of reward learning: a fast learning process that rapidly forgot, and a slower learning process that persisted across multiple days. This model then formed the basis for a parameterised data set containing trial-by-trial estimates of the prediction errors produced by fast and slow learning rates and where a statistical learning analysis showed these two types of prediction errors were decodable from heart rate and EEG data (recorded from a single wearable electrode) collected over the course of the experiment. Crucially, these physiological representations of prediction error accurately predicted self-reported mood at short and long timescales, revealing a relationship not evident from behaviour alone (119).

An increasing number of studies now use decoded state representations to investigate how reward is algorithmically processed, with considerable potential for understanding mood disorders such as depression and anxiety (120). One formulation of value-guided decision-making is the "successor representation"(121), which describes how we build a predictive map of state values. Recent decoding of functional MRI data has shown that, during decision-making, the successor representation predicts which states are reactivated in the brain more accurately than other behavioural models (122). In a similar vein, MEG investigations have shown that neural reactivation of outcomes during choice deliberation is modulated by both the subjective value and probability of an outcome (123), and predicts subsequent choice behaviour (124).

Conclusion

We highlight a recent trend in the application of statistical learning to neuroimaging data, particularly MEG, where the goal has been to uncover rapid reactivation of state representations that might otherwise go undetected, either due to spatiotemporal limitations of neuroimaging modalities or the complexity of the evolving state representation. These decoded representations can serve as rich and dynamic support for, or latent variables within, computational models of complex cognitive processes, allowing investigation of a range of candidate processes that may go awry in psychiatric disorders. When combined with neurophysiological recordings, such as MEG, pattern classification provides a level of spatiotemporal precision that is virtually impossible to gain from behaviour-only models or from conventional neuroimaging analyses. In turn, combining neural decoding of states with computational models of behaviour or cognition provides a level of representational precision not easily attained using conventional neuroimaging analysis alone. Moreover, by classifying holistic mental states, researchers can access highly temporally-resolved signatures of disorder-related representations, opening new avenues for examining cognition and behaviour in ecological contexts that involve a high degree of representational complexity, including indexing the impact of treatments.

Acknowledgments

This work was carried out within the framework of a Max Planck UCL Center for Computational Psychiatry and Ageing Research, supported by the Max Planck Society (MPS). RJD and JM are supported by a Wellcome Investigator Award, 098362/Z/12/Z. The Wellcome Centre for Human Neuroimaging (WCHN) is supported by core funding from the Wellcome Trust (203147/Z/16/Z).

Disclosures

The authors report no biomedical financial interests or potential conflicts of interest.

References

1. Kapur S, Phillips AG, Insel TR (2012): Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Mol Psychiatry* 17: 1174–1179.
2. Aryutova K, Paunova R, Kandilarova S, Todeva-Radneva A, Stoyanov D (2021): Implications from translational cross-validation of clinical assessment tools for diagnosis and treatment in psychiatry. *World J Psychiatry* 11: 169–180.
3. Abramovitch, Schweiger (n.d.): Misuse of cognitive neuropsychology in psychiatry research: the intoxicating appeal of neo-reductionism. *Behav Ther* . Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.710.7869&rep=rep1&type=pdf
4. Specht K (2019): Current Challenges in Translational and Clinical fMRI and Future Directions. *Front Psychiatry* 10: 924.
5. First MB, Drevets WC, Carter C, Dickstein DP, Kasoff L, Kim KL, *et al.* (2018): Clinical Applications of Neuroimaging in Psychiatric Disorders. *Am J Psychiatry* 175: 915–916.
6. Fried EI (2021, August): Studying mental health problems as systems, not syndromes. https://doi.org/10.31234/osf.io/k4mhv
7. Neuner I, Veselinović T, Ramkiran S, Rajkumar R, Schnellbaecher GJ, Shah NJ (2022): 7T ultra-high-field neuroimaging for mental health: an emerging tool for precision psychiatry? *Transl Psychiatry* 12: 36.
8. Etkin A (2019): A Reckoning and Research Agenda for Neuroimaging in Psychiatry. *American Journal of Psychiatry*, vol. 176. pp 507–511.
9. Saggar M, Uddin LQ (2019): Pushing the Boundaries of Psychiatric Neuroimaging to Ground Diagnosis in Biology. *eNeuro* 6. https://doi.org/10.1523/ENEURO.0384-19.2019
10. Henderson TA, van Lierop MJ, McLean M, Uszler JM, Thornton JF, Siow Y-H, *et al.* (2020): Functional Neuroimaging in Psychiatry—Aiding in Diagnosis and Guiding Treatment. What the American Psychiatric Association Does Not Know. *Front Psychiatry* 11. https://doi.org/10.3389/fpsyt.2020.00276
11. Adams RA, Huys QJM, Roiser JP (2016): Computational Psychiatry: towards a mathematically informed understanding of mental illness. *J Neurol Neurosurg Psychiatry* 87: 53–63.
12. Mobbs D, Wise T, Suthana N, Guzmán N, Kriegeskorte N, Leibo JZ (2021): Promises and challenges of human computational ethology. *Neuron* 109: 2224–2238.
13. Zaki J, Ochsner K (2009): The need for a cognitive neuroscience of naturalistic social cognition. *Ann N Y Acad Sci* 1167: 16–30.

14. Hitchcock PF, Fried EI, Frank MJ (2022): Computational Psychiatry Needs Time and Context. *Annu Rev Psychol* 73: 243–270.

15. Huys QJM, Maia TV, Paulus MP (2016): Computational Psychiatry: From Mechanistic Insights to the Development of New Treatments. *Biol Psychiatry Cogn Neurosci Neuroimaging* 1: 382–385.

16. De Martino F, Yacoub E, Kemper V, Moerel M, Uludağ K, De Weerd P, *et al.* (2018): The impact of ultra-high field MRI on cognitive and computational neuroimaging. *Neuroimage* 168: 366–382.

17. Cohen D (1968): Magnetoencephalography: evidence of magnetic fields produced by alpha-rhythm currents. *Science* 161: 784–786.

18. Pravdich-Neminsky, W (1912): Ein Versuch der Registrierung der Elektrischen Gehirnerscheinungen. *Zentralbl Physiol* 27: 951–960.

19. Boto E, Holmes N, Leggett J, Roberts G, Shah V, Meyer SS, Brookes MJ (2018): Moving brain imaging towards real-world applications using a wearable MEG system. *Nature* 17–19.

20. Liao Y, Huang X, Wu Q, Yang C, Kuang W, Du M, *et al.* (2013): Is depression a disconnection syndrome? Meta-analysis of diffusion tensor imaging studies in patients with MDD. *J Psychiatry Neurosci* 38: 49–56.

21. Kaiser RH, Andrews-Hanna JR, Wager TD, Pizzagalli DA (2015): Large-Scale Network Dysfunction in Major Depressive Disorder: A Meta-analysis of Resting-State Functional Connectivity. *JAMA Psychiatry* 72: 603–611.

22. Li J, Li N, Shao X, Chen J, Hao Y, Li X, Hu B (2021): Altered Brain Dynamics and Their Ability for Major Depression Detection using EEG Microstates Analysis. *IEEE Transactions on Affective Computing* 1–1.

23. Murphy M, Whitton AE, Deccy S, Ironside ML, Rutherford A, Beltzer M, *et al.* (2020): Abnormalities in electroencephalographic microstates are state and trait markers of major depressive disorder. *Neuropsychopharmacology* 45: 2030–2037.

24. Xin Y, Bai T, Zhang T, Chen Y, Wang K, Yu S, *et al.* (2022): Electroconvulsive therapy modulates critical brain dynamics in major depressive disorder patients. *Brain Stimul* 15: 214–225.

25. McFadyen J, Dolan RJ, Garrido MI (2020): The influence of subcortical shortcuts on disordered sensory and cognitive processing. *Nat Rev Neurosci* 21: 264–276.

26. Uludağ K, Roebroeck A (2014): General overview on the merits of multimodal neuroimaging data fusion. *Neuroimage* 102 Pt 1: 3–10.

27. Calhoun VD, Sui J (2016): Multimodal Fusion of Brain Imaging Data: A Key to Finding the Missing Link(s) in Complex Mental Illness. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, vol. 1. pp 230–244.

28. Foti D, Carlson JM, Sauder CL, Proudfit GH (2014): Reward dysfunction in major depression: multimodal neuroimaging evidence for refining the melancholic phenotype. *Neuroimage* 101: 50–58.

29. Keren H, O'Callaghan G, Vidal-Ribas P, Buzzell GA, Brotman MA, Leibenluft E, *et al.* (2018): Reward Processing in Depression: A Conceptual and Meta-Analytic Review Across fMRI and EEG Studies. *Am J Psychiatry* 175: 1111–1120.

30. Zhang Y-D, Dong Z, Wang S-H, Yu X, Yao X, Zhou Q, *et al.* (2020): Advances in multimodal data fusion in neuroimaging: Overview, challenges, and novel orientation. *Inf Fusion* 64: 149–187.

31. Stokes MG, Wolff MJ, Spaak E (2015): Decoding Rich Spatial Information with High Temporal Resolution. *Trends Cogn Sci* 19: 636–638.

32. Kamitani Y, Tong F (2005): Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8: 679–685.

33. Cichy RM, Ramirez FM, Pantazis D (2015): Can visual information encoded in cortical columns be decoded from magnetoencephalography data in humans? *Neuroimage* 121: 193–204.

34. Nasiotis K, Clavagnier S, Baillet S, Pack CC (2017): High-resolution retinotopic maps estimated with magnetoencephalography. *Neuroimage* 145: 107–117.

35. Friston KJ, Harrison L, Penny W (2003): Dynamic causal modelling. *Neuroimage* 19: 1273–1302.

36. Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001): Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293: 2425–2430.

37. Liu Y, Dolan RJ, Higgins C, Penagos H, Woolrich MW, Freyja Ólafsdóttir H, *et al.* (2021): Temporally delayed linear modelling (TDLM) measures replay in both animals and humans. *eLife*, vol. 10. https://doi.org/10.7554/elife.66917

38. deCharms RC, Zador A (2000): Neural representation and the cortical code. *Annu Rev Neurosci* 23: 613–647.

39. Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, Kurth-Nelson Z (2018): What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior. *Neuron* 100: 490–509.

40. Momennejad I (2020): Learning Structures: Predictive Representations, Replay, and Generalization. *Curr Opin Behav Sci* 32: 155–166.

41. Sutton RS, Barto AG, Others (1999): Reinforcement learning. *J Cogn Neurosci* 11: 126–134.

42. Nair A, Rutledge RB, Mason L (2020): Under the Hood: Using Computational Psychiatry to Make Psychological Therapies More Mechanism-Focused. *Front Psychiatry* 11: 140.

43. Liu Y, Nour MM, Schuck NW, Behrens TEJ, Dolan RJ (2022): Decoding cognition from spontaneous neural activity. *Nat Rev Neurosci* 23: 204–214.

44. Skaggs WE, McNaughton BL (1996): Replay of neuronal firing sequences in rat hippocampus during sleep following spatial experience. *Science* 271: 1870–1873.

45. Nádasdy Z, Hirase H, Czurkó A, Csicsvari J, Buzsáki G (1999): Replay and time compression of recurring spike sequences in the hippocampus. *J Neurosci* 19: 9497–9507.

46. Wilson MA, McNaughton BL (1994): Reactivation of hippocampal ensemble memories during sleep. *Science* 265: 676–679.

47. Jadhav SP, Kemere C, German PW, Frank LM (2012): Awake hippocampal sharp-wave ripples support spatial memory. *Science* 336: 1454–1458.

48. Dupret D, O'Neill J, Pleydell-Bouverie B, Csicsvari J (2010): The reorganization and reactivation of hippocampal maps predict spatial memory performance. *Nat Neurosci* 13: 995–1002.

49. Ego-Stengel V, Wilson MA (2010): Disruption of ripple-associated hippocampal activity during rest impairs spatial learning in the rat. *Hippocampus* 20: 1–10.

50. Girardeau G, Benchenane K, Wiener SI, Buzsáki G, Zugaro MB (2009): Selective suppression of hippocampal ripples impairs spatial memory. *Nat Neurosci* 12: 1222–1223.

51. Diba K, Buzsáki G (2007): Forward and reverse hippocampal place-cell sequences during ripples. *Nat Neurosci* 10: 1241–1242.

52. Foster DJ, Wilson MA (2006): Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440: 680–683.

53. Singer AC, Frank LM (2009): Rewarded outcomes enhance reactivation of experience in the hippocampus. *Neuron* 64: 910–921.

54. Ambrose RE, Pfeiffer BE, Foster DJ (2016): Reverse Replay of Hippocampal Place Cells Is Uniquely Modulated by Changing Reward. *Neuron* 91: 1124–1136.

55. Carey AA, Tanaka Y, van der Meer MAA (2019): Reward revaluation biases hippocampal replay content away from the preferred outcome. *Nat Neurosci* 22: 1450–1459.

56. Michon F, Sun J-J, Kim CY, Ciliberti D, Kloosterman F (2019): Post-learning Hippocampal Replay Selectively Reinforces Spatial Memory for Highly Rewarded Locations. *Curr Biol* 29: 1436–1444.e5.

57. Mou X, Pokhrel A, Suresh P, Ji D (2021): Observational learning promotes hippocampal remote awake replay toward future reward locations. *Neuron*. https://doi.org/10.1016/j.neuron.2021.12.005

58. Gupta AS, van der Meer MAA, Touretzky DS, Redish AD (2010): Hippocampal replay is not a simple function of experience. *Neuron* 65: 695–705.

59. Mattar MG, Daw ND (2018): Prioritized memory access explains planning and hippocampal replay. *Nature Neuroscience*, vol. 21. pp 1609–1617.

60. Ólafsdóttir HF, Bush D, Barry C (2018): The Role of Hippocampal Replay in Memory and Planning. *Curr Biol* 28:

R37–R50.

61. Ólafsdóttir HF, Carpenter F, Barry C (2017): Task Demands Predict a Dynamic Switch in the Content of Awake Hippocampal Replay. *Neuron* 96: 925–935.e6.

62. Shin JD, Tang W, Jadhav SP (2019): Dynamics of Awake Hippocampal-Prefrontal Replay for Spatial Learning and Memory-Guided Decision Making. *Neuron* 104: 1110–1125.e7.

63. Kurth-Nelson Z, Economides M, Dolan RJ, Dayan P (2016): Fast Sequences of Non-spatial State Representations in Humans. *Neuron* 91: 194–204.

64. Chenani A, Sabariego M, Schlesiger MI, Leutgeb JK, Leutgeb S, Leibold C (2019): Hippocampal CA1 replay becomes less prominent but more rigid without inputs from medial entorhinal cortex. *Nat Commun* 10: 1341.

65. Vaz AP, Wittig JH Jr, Inati SK, Zaghloul KA (2020): Replay of cortical spiking sequences during human memory retrieval. *Science* 367: 1131–1134.

66. Ji D, Wilson MA (2007): Coordinated memory replay in the visual cortex and hippocampus during sleep. *Nat Neurosci* 10: 100–107.

67. Peyrache A, Khamassi M, Benchenane K, Wiener SI, Battaglia FP (2009): Replay of rule-learning related neural patterns in the prefrontal cortex during sleep. *Nat Neurosci* 12: 919–926.

68. Kaefer K, Nardin M, Blahna K, Csicsvari J (2020): Replay of Behavioral Sequences in the Medial Prefrontal Cortex during Rule Switching. *Neuron* 106: 154–165.e6.

69. Yu JY, Liu DF, Loback A, Grossrubatscher I, Frank LM (2018): Specific hippocampal representations are linked to generalized cortical representations in memory. *Nat Commun* 9: 2209.

70. Berners-Lee A, Wu X, Foster DJ (2021): Prefrontal cortical neurons are selective for non-local hippocampal representations during replay and behavior. *J Neurosci*. https://doi.org/10.1523/JNEUROSCI.1158-20.2021

71. Tierney TM, Levy A, Barry DN, Meyer SS, Shigihara Y, Everatt M, *et al.* (2021): Mouth magnetoencephalography: A unique perspective on the human hippocampus. *Neuroimage* 225: 117443.

72. Ruzich E, Crespo-García M, Dalal SS, Schneiderman JF (2019): Characterizing hippocampal dynamics with MEG: A systematic review and evidence-based guidelines. *Hum Brain Mapp* 40: 1353–1375.

73. Jafarpour A, Fuentemilla L, Horner AJ, Penny W, Duzel E (2014): Replay of very early encoding representations during recollection. *J Neurosci* 34: 242–248.

74. Liu Y, Mattar MG, Behrens TEJ, Daw ND, Dolan RJ (2021): Experience replay is associated with efficient nonlocal learning. *Science* 372. https://doi.org/10.1126/science.abf1357

75. Wimmer GE, Liu Y, Vehar N, Behrens TEJ, Dolan RJ (2020): Episodic memory retrieval success is associated with rapid replay of episode content. *Nat Neurosci* 23: 1025–1033.

76. Liu Y, Dolan RJ, Kurth-Nelson Z, Behrens TEJ (2019): Human Replay Spontaneously Reorganizes Experience. *Cell* 178: 640–652.e14.

77. Michelmann S, Staresina BP, Bowman H, Hanslmayr S (2019): Speed of time-compressed forward replay flexibly changes in human episodic memory. *Nat Hum Behav* 3: 143–154.

78. Nour MM, Liu Y, Arumuham A, Kurth-Nelson Z, Dolan RJ (2021): Impaired neural replay of inferred relationships in schizophrenia. *Cell* 184: 4315–4328.e17.

79. Eldar E, Lièvre G, Dayan P, Dolan RJ (2020): The roles of online and offline replay in planning. *Elife* 9. https://doi.org/10.7554/eLife.56911

80. McFadyen J, Liu Y, Dolan RJ (2021, November 19): Differential replay for reward and punishment paths predicts approach and avoidance. *bioRxiv*. p 2021.11.18.468950.

81. Mattar MG, Lengyel M (2022): Planning in the brain. *Neuron* 110: 914–934.

82. Gagne C, Dayan P, Bishop SJ (2018): When planning to survive goes wrong: predicting the future and replaying the past in anxiety and PTSD. *Current Opinion in Behavioral Sciences* 24: 89–95.

83. Huys QJM, Browning M, Paulus MP, Frank MJ (2021): Advances in the computational understanding of mental illness. *Neuropsychopharmacology* 46: 3–19.

84. Huys QJM, Maia TV, Frank MJ (2016): Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci* 19: 404–413.

85. Groman SM, Massi B, Mathias SR, Lee D, Taylor JR (2019): Model-Free and Model-Based Influences in Addiction-Related Behaviors. *Biol Psychiatry* 85: 936–945.

86. Chen C, Takahashi T, Nakagawa S, Inoue T, Kusumi I (2015): Reinforcement learning in depression: A review of computational research. *Neurosci Biobehav Rev* 55: 247–267.

87. Daw N (2020): Model-Based and Model-Free Learning in Anorexia Nervosa and Other Disorders. *Biol Psychiatry* 87: S20.

88. Bishop SJ, Gagne C (2018): Anxiety, Depression, and Decision Making: A Computational Perspective. *Annu Rev Neurosci* 41: 371–388.

89. Voon V, Reiter A, Sebold M, Groman S (2017): Model-Based Control in Dimensional Psychiatry. *Biol Psychiatry* 82: 391–400.

90. Musa A, Khan S, Mujahid M, El-Gaby M (2022): The shallow cognitive map hypothesis: A hippocampal framework for thought disorder in schizophrenia. *Schizophrenia* 8: 1–11.

91. Nour MM, Dolan RJ (2022, January 15): Synaptic Gain Abnormalities in Schizophrenia and the Potential Relevance for Cognition. *Biological Psychiatry*, vol. 91. pp 167–169.

503 92. Adams RA, Pinotsis D, Tsirlis K, Unruh L, Mahajan A, Horas AM, *et al.* (2022): Computational Modeling of
504 Electroencephalography and Functional Magnetic Resonance Imaging Paradigms Indicates a Consistent Loss of
505 Pyramidal Cell Synaptic Gain in Schizophrenia. *Biol Psychiatry* 91: 202–215.
506 93. Reed EJ, Uddenberg S, Suthaharan P, Mathys CD, Taylor JR, Groman SM, Corlett PR (2020): Paranoia as a
507 deficit in non-social belief updating. *eLife*, vol. 9. https://doi.org/10.7554/elife.56345
508 94. Suthaharan P, Reed EJ, Leptourgos P, Kenney JG, Uddenberg S, Mathys CD, *et al.* (2021): Paranoia and belief
509 updating during the COVID-19 crisis. *Nat Hum Behav* 5: 1190–1202.
510 95. Adams RA, Vincent P, Benrimoh D, Friston KJ, Parr T (2022): Everything is connected: Inference and attractors
511 in delusions. *Schizophr Res* 245: 5–22.
512 96. Sheffield JM, Suthaharan P, Leptourgos P, Corlett PR (2022): Belief Updating and Paranoia in Individuals with
513 Schizophrenia. *Biol Psychiatry Cogn Neurosci Neuroimaging*. https://doi.org/10.1016/j.bpsc.2022.03.013
514 97. Katthagen T, Fromm S, Wieland L, Schlagenhauf F (2022): Models of Dynamic Belief Updating in Psychosis-A
515 Review Across Different Computational Approaches. *Front Psychiatry* 13: 814111.
516 98. McKay D, Sookman D, Neziroglu F, Wilhelm S, Stein DJ, Kyrios M, *et al.* (2015): Efficacy of cognitive-behavioral
517 therapy for obsessive–compulsive disorder. *Psychiatry Res* 225: 236–246.
518 99. Cuijpers P, Berking M, Andersson G, Quigley L, Kleiboer A, Dobson KS (2013): A meta-analysis of cognitive-
519 behavioural therapy for adult depression, alone and in comparison with other treatments. *Can J Psychiatry* 58:
520 376–385.
521 100. Fried EI, Koenders MA, Blom JD (2021): Bleuler revisited: on persecutory delusions and their resistance to
522 therapy. *Lancet Psychiatry* 8: 644–646.
523 101. Sookman D, Steketee G (2007): Directions in Specialized Cognitive Behavior Therapy for Resistant Obsessive-
524 Compulsive Disorder: Theory and Practice of Two Approaches. *Cogn Behav Pract* 14: 1–17.
525 102. Weiss A, Chambon V, Lee JK, Drugowitsch J, Wyart V (2021): Interacting with volatile environments stabilizes
526 hidden-state inference and its brain signatures. *Nat Commun* 12: 2228.
527 103. Fradkin I, Adams RA, Parr T, Roiser JP, Huppert JD (2020): Searching for an anchor in an unpredictable world:
528 A computational model of obsessive compulsive disorder. *Psychol Rev* 127: 672–699.
529 104. Rouault M, Weiss A, Lee JK, Drugowitsch J, Chambon V, Wyart V (2021): Controllability reveals defining
530 features of information seeking. Retrieved from https://europepmc.org/article/ppr/ppr260674
531 105. Pike AC, Robinson OJ (2022): Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control
532 Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry* 79: 313–322.
533 106. FitzGerald THB, Dolan RJ, Friston K (2015): Dopamine, reward learning, and active inference. *Front Comput*

Neurosci 9: 136.

107. Liuzzi L, Chang KK, Zheng C, Keren H, Saha D, Nielson DM, Stringaris A (2021): Magnetoencephalographic Correlates of Mood and Reward Dynamics in Human Adolescents. *Cereb Cortex*. https://doi.org/10.1093/cercor/bhab417

108. Sambrook TD, Hardwick B, Wills AJ, Goslin J (2018): Model-free and model-based reward prediction errors in EEG. *Neuroimage* 178: 162–171.

109. Kahnt T (2018): A decade of decoding reward-related fMRI signals and where we go from here. *NeuroImage*, vol. 180. pp 324–333.

110. Queirazza F, Fouragnan E, Steele JD, Cavanagh J, Philiastides MG (2019): Neural correlates of weighted reward prediction error during reinforcement learning classify response to cognitive behavioral therapy in depression. *Sci Adv* 5: eaav4962.

111. Hitchcock P, Forman E, Rothstein N, Zhang F, Kounios J, Niv Y, Sims C (2021): Rumination Derails Reinforcement Learning With Possible Implications for Ineffective Behavior. *Clin Psychol Sci* 21677026211051324.

112. Rich EL, Wallis JD (2016): Decoding subjective decisions from orbitofrontal cortex. *Nat Neurosci* 19: 973–980.

113. Zhou J, Gardner MPH, Stalnaker TA, Ramus SJ, Wikenheiser AM, Niv Y, Schoenbaum G (2019): Rat Orbitofrontal Ensemble Activity Contains Multiplexed but Dissociable Representations of Value and Task Structure in an Odor Sequence Task. *Curr Biol* 29: 897–907.e3.

114. Wang F, Schoenbaum G, Kahnt T (2020): Interactions between human orbitofrontal cortex and hippocampus support model-based inference. *PLoS Biol* 18: e3000578.

115. Elliott Wimmer G, Büchel C (2019): Learning of distant state predictions by the orbitofrontal cortex in humans. *Nat Commun* 10: 2554.

116. Alloy LB, Olino T, Freed RD, Nusslock R (2016): Role of Reward Sensitivity and Processing in Major Depressive and Bipolar Spectrum Disorders. *Behav Ther* 47: 600–621.

117. Eldar E, Niv Y (2015): Interaction between emotional state and learning underlies mood instability. *Nat Commun* 6: 6149.

118. Huys QJ, Pizzagalli DA, Bogdan R, Dayan P (2013): Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol Mood Anxiety Disord* 3: 12.

119. Eldar E, Roth C, Dayan P, Dolan RJ (2018): Decodability of Reward Learning Signals Predicts Mood Fluctuations. *Curr Biol* 28: 1433–1439.e7.

120. Huys QJM, Daw ND, Dayan P (2015): Depression: a decision-theoretic analysis. *Annu Rev Neurosci* 38: 1–23.

121. Dayan P (1993): Improving Generalization for Temporal Difference Learning: The Successor Representation. *Neural Comput* 5: 613–624.

122. Russek EM, Momennejad I, Botvinick MM, Gershman SJ, Daw ND (2021, August 31): Neural evidence for the successor representation in choice evaluation. *bioRxiv*. p 2021.08.29.458114.

123. Russek E, Moran R, Liu Y, Dolan RJ, Huys QJM (2021, October 10): Selective outcome reinstatement during evaluation drives heuristics in risky choice. *PsyArXiv*. https://doi.org/10.31234/osf.io/kb6ew

124. Castegnetti G, Tzovara A, Khemka S, Melinščak F, Barnes GR, Dolan RJ, Bach DR (2020): Representation of probabilistic outcomes during risky decision-making. *Nat Commun* 11: 2419.

125. Nour MM, Dahoun T, Schwartenbeck P, Adams RA, FitzGerald THB, Coello C, *et al.* (2018): Dopaminergic basis for signaling belief updates, but not surprise, and the link to paranoia. *Proc Natl Acad Sci U S A* 115: E10167–E10176.

126. Wise T, Liu Y, Chowdhury F, Dolan RJ (2021): Model-based aversive learning in humans is supported by preferential task state reactivation. *Sci Adv* 7. https://doi.org/10.1126/sciadv.abf9616

Figure legends

Figure 1. Capturing mental states using statistical learning. (A) Mental states, such as viewing an image, can be differentiated by the unique patterns of evoked spatiotemporal brain activity, captured with MEG. These spatiotemporal state classifiers can then be applied to MEG data during a task of interest (e.g., decision-making), revealing the time-course of state reactivation associated with specific aspects of cognition and behaviour. **(B)** Visual orientation can be classified from MEG and EEG sensor data due to unique configurations of angled dipoles along the cortical surface. Adapted from Stokes et al. (2015). **(C)** Different mental states may also evoke different neural network configurations, producing unique patterns of activity across MEG sensors.

587

Term	Definition
Machine learning	A methodological approach in which an algorithm (e.g., a support vector machine) is iteratively improved to capture relationships between variables in a training data set (43). The optimised algorithm is then applied to a test data set to predict the same relationships. Machine learning may be supervised or unsupervised, and is generally model-agnostic.
Statistical learning	A branch of machine learning in which a suitable statistical model (e.g., logistic regression) is deliberately selected and fit to a training data set in order to infer relationships between variables, in accordance with the assumptions of the selected model (44). The optimised model may then be used to predict relationships in a test data set.
Multi-voxel pattern analysis (MVPA)	A supervised classification problem that captures the relationship between spatial patterns of BOLD signal across voxels and a particular experimental condition in a training data set (43). These spatial patterns can then be detected by applying classifiers to a test data set.
Neural representation	A spatiotemporal pattern of neural activity that is reliably evoked by a specific mental or physical state, indicating that the pattern "encodes" the state (45).
Cognitive map	A neural representation of how different states relate to each other (46).
Structural inference	The ability to infer how an environment is structured, given previous experience of state-to-state transitions, as well as any higher-order information (46). In other words, the ability to construct, utilise, and update a cognitive map.
Replay	A neurophysiological phenomenon whereby neural representations of states are reactivated in a specific order, indicating their relationships within a cognitive map (47).
Computational psychiatry	A field of research in which generative mathematical models are constructed to explain the relationships between behaviour, cognition, environment, and underlying neurobiology of psychiatric disorders (17).
Reinforcement learning	A computational model describing how decision-making is influenced by past experiences of reward (48).
Cognitive behavioural therapy (CBT)	A talking therapy that aims to reduce symptoms of mental disorders by challenging dysfunctional beliefs (cognition) and their associated behaviours (49).

588 **Table 1. Key terms and definitions**

589

Research question	Existing data	Potential use cases
What are the fine-grained neurobiological causes of psychiatric symptoms, and can knowledge of this assist with prognosis and/or treatment?	**Schizophrenia**: Disorganised replay suggests a neurophysiological basis for impaired structural inference, implying abnormal NMDA receptor function in hippocampus (85,98). **Schizophrenia**: Multimodal imaging shows a coupling of computationally-derived belief updates with BOLD signal in striatum that relates to dopamine receptor functionality measured with positron-emission tomography (PET) (132). **Depression**: Functional connectivity measured with fMRI in depression is markedly reduced at rest (27,28). Sub-second transient changes in microstates of functional connectivity detected with EEG is significantly different between clinical subtypes of depression (29,30).	**Schizophrenia:** Replay of reorganised state sequences may be used as an indicator of the efficacy of dopaminergic antagonists on increasing synaptic gain in hippocampus, supporting structural inference capabilities. **Depression**: MEG may be used as a more spatially-precise measure of rapid changes in microstates of functional connectivity, a measure that could help to predict patient-specific efficacy of electroconvulsive therapy (31).
How can we better estimate the efficacy of CBT in restructuring dysfunctional beliefs?	**Depression**: Reward prediction error signals related to learning in amygdala and striatum (measured with fMRI) predict response of depressed patients to CBT (117). **General**: The perceived congruence between current evidence and prior beliefs can be decoded from MEG activity and used to indicate the time course of belief updating and subsequent decision-making (109).	**Depression**: By using decoding to track how rewarding outcomes are neurally represented during choice deliberation, we could assess the efficacy of CBT in increasing representation of reward in a manner that relates to improved choice behaviour. **OCD**: Neural signatures of belief updating could indicate how acting on an environment to sample information (as is the case in compulsive behaviour) influences beliefs about uncertain environments, and whether this is influenced by CBT (109).
How do thought patterns (conscious or unconscious) differ between clinical subtypes, and can this guide personalised therapy?	**Anxiety**: Replay supports flexible avoidance of potential threat by simulating inferred trajectories to threat (133). **General**: Replay reflects an ability to infer trajectories that lead to future reward in changing environments (81).	**Anxiety**: Patients with anxiety may differ in whether they anxiously anticipate the future or ruminate on the past, which could reflect different magnitudes of forwards replay of paths leading to threat versus backwards replay after outcome receipt. These signatures, if present, could therefore serve as biological markers of anxiety subtypes.

590 **Table 2. Outstanding questions in psychiatry that may be addressed by using increasing spatiotemporal resolution of neuroimaging data**

591