JOURNAL OF PSYCHOPATHOLOGY 2026 JUL 28;
ONLINE FIRST
DOI: 10.36148/2284-0249-2001

Review

What is Computational Psychiatry? An introductory narrative review

Linda Antonucci[1], Filippo Corponi[2], Federica Folesani[3], Tommaso Jannini[4], Chiara Montemitro[3], Angela Muscettola[3], Francesco Achilli[3], Paolo Ossola[5], Alessandro Pigoni[6], Alessandra Raio[1], Rodolfo Rossi[7], Martino Belvederi Murri[3]

[1] Department of Translational Biomedicine and Neuroscience (DiBraiN), University of Bari "Aldo Moro", Bari, Italy; [2] Division of Psychiatry, Department of Brain Sciences, Imperial College London; [3] Institute of Psychiatry, Department of Neuroscience and Rehabilitation, University of Ferrara, Ferrara, Italy; [4] Department of Experimental Medicine, Tor Vergata University of Rome, Rome, Italy; [5] Department of Medicine and Surgery, University of Parma, Parma, Italy; [6] Department of Neurosciences and Mental Health, Fondazione IRCCS Ca' Granda, Ospedale Maggiore Policlinico, Milan, Italy; [7] Department of Systems Medicine, Tor Vergata University of Rome, Italy

SUMMARY

The aim of this review is to provide an introductory overview of Computational Psychiatry (CP). CP is a broad, interdisciplinary field that integrates concepts and methods from mathematics, informatics, neuroscience and related disciplines to improve the understanding, prognosis, and treatment of psychiatric disorders. CP seeks to leverage the advances of computational tools to address the intrinsic complexities of psychiatry phenomena and to overcome long-standing limitations of psychiatric research, including the lack of formal theories, the reliance on descriptive and agnostic diagnostic systems, the scarcity of valid biomarkers, the predominance of simplistic outcome measures, limited treatment specificity, and the low reproducibility of many findings.

CP encompasses two complementary strategies. Theory-driven approaches focus on the development of generative models often informed by cognitive neuroscience, that provide formal (mathematical) representation of latent mental processes, within frameworks such as belief updating, decision-making, or reinforcement learning. These models aim to explain how psychiatric phenomena emerge from altered neural computations, thereby offering mechanistic links between neurobiology, behaviour and subjective experience. Data-driven approaches, instead, leverage Machine Learning and other inferential techniques to analyse large, multimodal datasets. This enables the identification of latent structure in complex data supporting tasks such as disease subtyping, risk stratification, prediction of clinical trajectories and estimation of differential treatment response.

Together, theory-driven and data-driven approaches provide complementary tools to generate and test specific hypotheses about psychiatric conditions, while advancing precision psychiatry through the tailoring of interventions to individual etiological, clinical, and pathophysiology profiles. This review illustrates key applications of CP through selected examples, including Reinforcement Learning models of anhedonia, Drift Diffusion Models of evidence accumulation, and contemporary AI approaches such as deep learning and explainable AI. These examples highlight how CP can move psychiatry beyond descriptive symptom frameworks, toward explanatory, predictive, and personalised accounts of mental disorders. While CP holds promise for the development of more precise diagnostic and therapeutic tools, important challenges remain, including issues of generalisability, interpretability, ethical oversight, and integration into routine clinical practice.

Key words: computational psychiatry, theory-driven models, data-driven models, reinforcement learning, precision psychiatry

Martino Belvederi Murri
E-mail: martino.belvederimurri@unife.it

© *Copyright by Pacini Editore Srl*

How to cite this article: Antonucci L, Cor-poni F, Folesani F, et al. What is Computational Psychiatry? An introductory narrative review. Journal of Psychopathology 2026 Jul 28 Online first. https://doi.org/10.36148/2284-0249-2001

OPEN ACCESS

This is an open access article distributed in accordance with the CC-BY-NC-ND (Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International) license. The article can be used by giving appropriate credit and mentioning the license, but only for non-commercial purposes and only in the original version. For further information: https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en

Introduction: What is Computational Psychiatry?

This review provides an introductory overview of Computational Psychiatry (CP) in the broad sense: an interdisciplinary field that leverages the potential of mathematics, informatics and related disciplines to improve the understanding of psychiatric phenomena. Over the last years, it has become evident that CP may contribute to reshaping the way we conceptualise, diagnose and treat mental disorders.

Mathematics, advanced statistics, and informatics have been extensively applied in psychiatry for several decades (*i.e.*, statistical analyses of experimental data, neuroimaging, psychometric methods, etc.). However, clinical practice still heavily relies on subjective measures like patient interviews, questionnaires, and clinical intuition. A subject-centred perspective is the *cornerstone* of mental-health care, but its inherently qualitative nature can contribute to heterogeneous diagnoses, uncertain prognoses, and variable treatment response. CP seeks not to replace but to formalise and enrich these subjective insights by integrating quantitative approaches that, among various applications, may link lived experience with objective indicators of information processing, neurobiology, and behaviour [1]. By modelling how aberrant computations give rise to both observable actions and first-person symptoms, the field aims to strengthen diagnostic precision, tailor interventions, and ultimately improve patient outcomes-while avoiding excessive reductionist approaches to mental disorders, such as interpreting them solely in terms of isolated biological or psychological abnormalities [2].

A key challenge in the endeavour of CP is to adequately represent the inherent complexity of psychiatric phenomena using mathematical models [3,4]. Psychiatric disorders, in fact, involve multiple levels of interpretation-neurobiological, psychological, behavioural, social, environmental, among others-and are not fully reducible to a single domain. From an empirical perspective, models should aim to be "*as simple as possible, but not simpler*", and should retain the ability to make correct predictions of the phenomena they seek to describe. Mainstream classification systems, such as the DSM-5 [5] and the ICD-11 [6], have been criticized for being rooted in a descriptive medical framework, and relying on a simplistic conceptualization of "symptoms". These diagnostic systems define disorders based on the co-occurrence of aggregated phenomena, rather than attempting to develop explicit working models of mental functions and dysfunctions; hence, they prioritize *reliability* over *validity* [7]. In other words, psychiatric diagnostic categories have been developed to be consistent, not to identify patterns of causal mechanisms or to support a theoretical basis. There is general consensus that the field of psychiatry still lacks formal, testable theories that would allow to confirm or refute core hypotheses.

The need to promote theoretical and process-based models of mental illness [8] has prompted a number of initiatives, most notably the National Institute of Mental Health's (NIMH), Research Domain Criteria (RDoC). The RDoC exemplifies the drive away from descriptive, statistical models of clinical diagnostic categories, towards mechanistic, theoretical ones by organizing psychopathology around functional domains and relying on neurobehavioral constructs [8]. Despite their promise, however, dimensional frameworks still lack a theoretical formalization that effectively explains the connection between clinical phenomenology, psychological and behavioural processes and underlying neurobiological features [7,9,10]. Other recent efforts to improve psychiatric research have aimed to enhance methodological transparency through open science practices, data sharing, and other rigorous procedures [11]. While extremely important, increased methodological transparency alone may not fully address the underlying theoretical gaps: a valid, theory-based classification system joining research and clinical practice remains an ongoing goal [12].

CP can be viewed as an array of methodological frameworks for constructing and testing formal models of mental health phenomena, explicitly designed for handling complex data and to represent latent mechanisms. Rather than relying on symptom aggregation, computational models aim to build explicit mappings between unobserved processes and observable data, while making assumptions and sources of uncertainty explicit, allowing hypotheses about such mechanisms to be specified, empirically tested and progressively refined [13]. In addition, computational approaches may process high-dimensional, multimodal data without prespecifying a guiding theoretical framework, but working backwards *from* detecting patterns of data *to* discover the origin of such recurrent structures, which might be due to relevant causal mechanisms. Advances in computing technology have been a key factor for the implementation of these strategies, allowing the estimation, simulation and comparison of models that would be otherwise computationally infeasible [14].

Schematically, two main strategies have emerged: the *theory-driven* and the *data-driven* approach. Both apply mathematical models to interpret human experience and behaviour but follow a distinct pathway. The theory-driven approach first assumes mathematical models (*a priori)* that may regulate neural computations, mainly based on cognitive neuroscience [10, 11], and often informed by neural data with high spatial or temporal resolution (*e.g.*, fMRI, EEG) [15]. Whereas the data-driven approach applies techniques such as Machine Learning (ML) to large quantities of data *post-hoc,* with the

aim of detecting patterns that may facilitate diagnoses and outcome predictions [13,16]. These methods are complementary: the former seeks to provide a scientific understanding about *how* and *why* phenomena are generated, ensuring that models are causally interpretable and grounded in existing neuroscience. The latter seeks to discover patterns that may prompt, extend or challenge existing theories and models.

Considering the example of *negativity bias* in depression. Cognitive theories of depression [17] suggest that maladaptive schemata activate a negativity bias that contributes to symptoms such as pessimism and hopelessness. Within a theory-driven approach, this hypothesis would be translated into an explicit computational model, describing exactly how specific aspects of information processing may be altered and generate negatively skewed evaluations or choices. The mathematical model would be also implemented in a behavioural task, allowing to compute unequivocal, precise operational indices of negativity bias (*i.e.* parameters) for each individual. These indices would be tested for their specificity to depression, or their relation to hypothesized symptom dimensions or neurobiological measures – e.g. pessimism or brain circuit activation. Finally, the model would provide the grounds to verify whether pessimism bias can be healed, for instance, by psychological interventions, neuromodulation, or pharmacological treatments. A data-driven approach, instead, works the other way around: it does not begin from a formal model, or even from a formed theoretical concept of negativity bias. It starts from analysing large behavioural, neuroimaging or digital phenotyping datasets using machine learning techniques to identify patterns or dimensions that might, for instance, discriminate individuals or predict clinical outcomes. In this example, a model analysing large-scale online behaviour data may identify a pattern of preferential engagement with negatively-valenced content, which may specifically predict pessimism and hopelessness, even though no construct such as "negativity bias" was specified a priori. Theory- and data-driven approaches are not independent alternatives but form a circular and mutually reinforcing process, in which data-driven analyses can generate or refine hypotheses, and theory-driven models provide structured explanations that can be tested and challenged by empirical data.

What is a model?

In both theory-driven and data-driven research, models are used in different ways but share a set of common features that are worth clarifying. A model, in the broad sense, is a simplified representation of a complex system, not necessarily a mathematical one. For example, classical psychoanalytical frameworks-such as the Freudian topographic model of id, ego, and superego-were conceived as conceptual models intended to organise and interpret mental phenomena, despite lacking formal mathematical specification. In the natural sciences, a model is built to structure observed data, describe, predict, or explain the behaviour of a system [18]. It is a framework that imposes a formal, comprehensible structure on observed phenomena. This can be particularly challenging in complex systems, especially in the context of psychiatry, due to *equifinality* and *multifinality* [19]. These terms indicate that similar observable states (*e.g.*, a psychiatric symptom) can arise from different origins, e.g., different underlying mechanisms, in equifinality, or that similar initial states can evolve into diverse outcomes (multifinality). In multifinality, this means that individuals with similar underlying dysfunctions (*i.e.*, brain lesions) or risk factors (*i.e.*, childhood maltreatment) can go on to develop rather different clinical characteristics or clinical trajectories [20].

Statistical models are among the most widely recognized models, describing possible links between observed variables through mathematical formulas, such as linear regression. These models are common in psychiatry due to their simplicity and interpretability, but they also have known limitations. Primarily, they reveal *associations* rather than causation; second, they may struggle to capture complex, non-linear relationships that are specific to the phenomena at hand. These and other features may limit their generalizability or clinical usefulness [21]. For example, the correlation between striatal dopamine D2/D3 receptor occupancy and the severity of anhedonia may be analysed with a linear regression [22]. However, this analysis does not represent key processes involved in the relationship, is not necessarily causal, and assumes that the association is strictly linear, despite neurobiological evidence of the contrary. Most biological mechanisms, in fact, are inherently non-linear: for instance, the relationship between brain function and neural activity typically follows a U-shaped pattern, with moderate neural activity associated with optimal functioning and extreme levels linked to dysfunction [23].

Computational models are specifically developed to address these limitations [21]. A computational model is a (generally complex) mathematical representation of a biological or psychological phenomenon; in the case of theory-driven approaches, it is based on a formal theoretical hypothesis of the *data-generating process* [18], in the case of data-driven analyses it captures the key features of *data patterns*. One important strength of computational models lies in their ability to recreate realistic scenarios of system dynamics, enabling researchers to explore the consequences of different types of perturbations, in terms of neural circuits, behaviour, symptoms, or psychiatric syndromes[24]. For instance, a recent

computational model formalizes panic disorder as a dynamic system in which symptoms reinforce each other through positive and negative feedback loops[25]. This model was used to explain either how physiological arousal and perceived threat might generate a self-amplifying cycle leading to a panic attack, and how avoidance behaviours and individual beliefs about perceived danger might contribute to symptoms persistence and chronicity.

Computational models summarise the hypothesised process through *parameters*. These parameters are latent variables that may quantify cognitive, neural, or informational processes. They do not exist as observable entities but provide a useful way of capturing and describing underlying phenomena. For instance, in simple linear models, the slope and the intercept of the regression line offer an intuitive summary of the linear association between two variables (Fig. 1A). In theory-driven approaches, where models are more complex, parameters represent unobservable processes (e.g., how quickly someone updates their beliefs, how much information is required to make a decision, or the mathematical relationships linking different datapoints) that are inferred from behavioural or neural data [26]. Every model has an associated *error*, that is the distance between the model predictions and the observed data (Fig. 1B). *Model fitting* hence become a crucial step, where alternative competing models are tested on the same data, and model parameters are systematically adjusted to minimize the error [27,28] (Fig. 1C). This process enhances the accuracy of models, allowing for more reliable inference [29]. Advanced statistical and mathematical criteria also assist in selecting the most appropriate model by balancing its theoretical validity, complexity, and its explanatory or predictive power [30]. Once parameter values are estimated, the model should be able to predict future data in similar conditions, such as the behaviour of a subject during more iterations of a computerized task, or, for instance, the probability of developing a condition in subjects with similar characteristics. Most importantly, the model should be able to generalize to unseen data, or to behaviour *outside the laboratory*. In

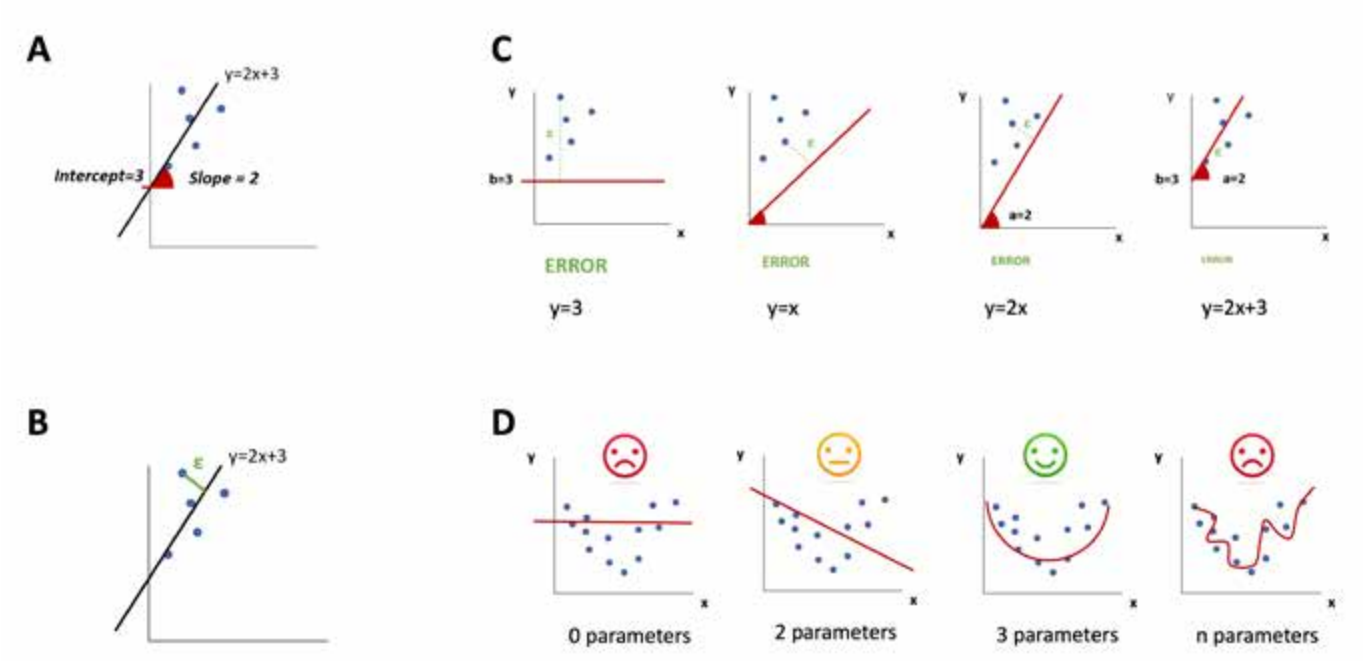


FIGURE 1. Basic principles of modelling. (A) *Illustration of a simple linear model (y=2x + 3) fitted to data points. The intercept (b=3) and the slope (a=2) summarize the linear relationship between variables x and y.* **(B)** *The green line represents the error (ε) as the distance between observed data point and the model's prediction. Model fitting aims to minimize these errors.* **(C)** *Comparison of models with increasing complexity and their associated errors. From left to right: a constant model (y=3) with one parameter (intercept only); a line through the origin (y=x) with one parameter (slope only, fixed at 1); a line with adjusted slope (y=2x) with one parameter; and a full linear model (y=2x + 3) with two parameters (slope and intercept). As model complexity increases appropriately, error decreases.* **(D)** *Models with too few parameters (0 parameters) underfit the data and fail to capture the underlying pattern. Models with moderate complexity (2-3 parameters) can capture the true relationship while maintaining generalizability. Models with excessive parameters (n parameters) overfit by matching every data point, losing the ability to generalize to new data.*

fact, a significant issue is *overfitting,* i.e., when a model becomes excessively tuned with a specific dataset, but loses its ability to generalize to new, unseen data [13] (Fig. 1D). A model must have a level of complexity appropriate to the problem at hand and be sufficiently representative across populations to support reliability and clinical use [31].

Theory-driven computational models

Theory-driven methods use existing scientific theories, principles and hypotheses to develop models that explain how and why certain behaviours or symptoms occur [15]. The neurobiological or causal mechanisms of a specific phenomenon are outlined at the theoretical level and formalised into one or more computational models. Mathematical models are used to assess whether the observed data are consistent with the theory, allowing specific hypotheses to be tested [15].

A defining characteristic of theory-driven computational models in psychiatry is that they do not merely describe associations between observable variables and behaviour but also postulate *latent* (theoretical, not directly measurable*) entities and processes* through which psychiatric phenomena emerge [26,28]. These *generative* models start by explicitly defining latent variables – typically representing cognitive or neural constructs – then describe their interactions in mathematical terms to describe how they generate the observable data-such as task choices, reaction times, or symptom ratings [27]. Models are constructed so that they can extrapolate, simulate virtual data as similar as possible to those observed in patients. This generative property is central to their scientific value [13,32] because it forces researchers to make explicit assumptions about which latent processes may be altered in a given disorder, and how such alterations are expected to manifest as observable symptoms or behavioural changes. In this way, generative models move beyond descriptive correlations and provide mechanistic accounts of psychiatric phenomena.

An important application of generative models is the exploration of counterfactual scenarios, for example by predicting how behaviour would change if a specific latent entity were altered. This allows hypotheses about potential interventions to be examined at the level of mechanism, prior to their application in patients. The iterative comparison between simulated and observed data supports progressive refinement of both models and theories, fostering cumulative advances in understanding psychiatric disorders

Finally, theory-driven models allow estimating the values of latent variables values at the individual level, providing a principled way to characterise inter-individual variability and identify distinct latent mechanisms that may give rise to similar clinical presentations (equifinality). This feature is particularly relevant for informing differential treatments strategies and supports the development of personalised approaches in precision psychiatry.

Generative models can be used to link different levels of analysis, such as behaviour, neuroimaging, and physiological measures – within a common formal framework, by specifying how observable data may arise from shared latent processes. Although theory-driven approaches can in principle be applied across biological levels, from neural circuits to large scale brain signals, in psychiatry the limited biological grounding of behaviour has meant that most applications have focused on modelling behavioural data from computerised tasks and relating the resulting parameters to subjective reports. This emphasis reflects a longstanding effort to bridge self-report measures and experimental evidence in psycho(patho)logy [33].

Within this context, decision-making has emerged as a central domain, as many psychiatric disorders can be understood as disturbances of choice under uncertainty [34,35]. Across conditions such as depression, addiction, schizophrenia, anxiety, and personality disorders, patients show systematic biases in how options are evaluated [36], beliefs are updated [37], or short- and long-term outcomes are weighed. To investigate such mechanisms, psychiatry has increasingly adopted behavioural tasks developed in neuroeconomics. These paradigms place participants in controlled, incentivised choice settings and allow the estimation of latent parameters – such as learning rates, risk preferences, or social valuation – that are difficult to capture reliably through self-report alone. Because these tasks are grounded in normative models from economics and neuroscience and can be translated across species, they provide a structured entry point for computational frameworks of decision-making, including reinforcement learning.

To further push forward theory-driven and mechanistic modelling, a promising direction is the development of biologically inspired Artificial Neural Networks (ANNs), which are emerging as tools for simulating neural dynamics across different levels of analysis [38]. Among these, Spiking Neural Networks (SNNs) are particularly valuable, as their bottom-up architecture allows them to include key neurobiological mechanisms, including synaptic plasticity and neuromodulation, reproducing neural dynamics with realistic topology [39]. Moreover, they can potentially model how cognitive, behavioural, and conscious processes may emerge from underlying neural computations, providing models that may realistically reproduce the brain's multilevel organisation[40].

While such biologically inspired approaches represent the frontier of theory-driven modelling, most current ap-

plications in psychiatry rely on more established computational frameworks. Reinforcement Learning[41] provides a formal decision-making framework for understanding how agents learn from feedback by evaluating outcomes and updating future choices based on past experiences and long-term goals [42]. Central to RL theory is the concept of the *prediction error,* defined as the discrepancy between expected and actual outcomes, which drives learning and behavioural adaptation [43]. Within this framework, behaviour is shaped by iterative cycles of prediction, feedback, minimization of surprise (or prediction error) and updating.

For example, imagine a person trying two new cafés on their way to work. At first, they alternate between options, uncertain which will provide better coffee. Each visit provides feedback-sometimes satisfying (reward), sometimes disappointing (null outcome)-leading to large prediction errors early on, when expectations are unstable. With repeated experience, learning occurs and expectations stabilise. The person comes to expect better coffee from one café and worse from the other, while prediction errors decrease, and choices increasingly favour the more rewarding option (Fig. 2A). Psychiatric disorders can entail alterations of this learning process: in addiction, choices may persist despite repeated negative feedback; in depression, reduced sensitivity to positive outcomes may attenuate learning; and in anxiety, exaggerated expectations of negative outcomes may bias behaviour toward avoidance. In this sense, RL offers a mechanistic account of how learning from experience can go awry in mental illness.

One of the most effective ways to formalise RL is the

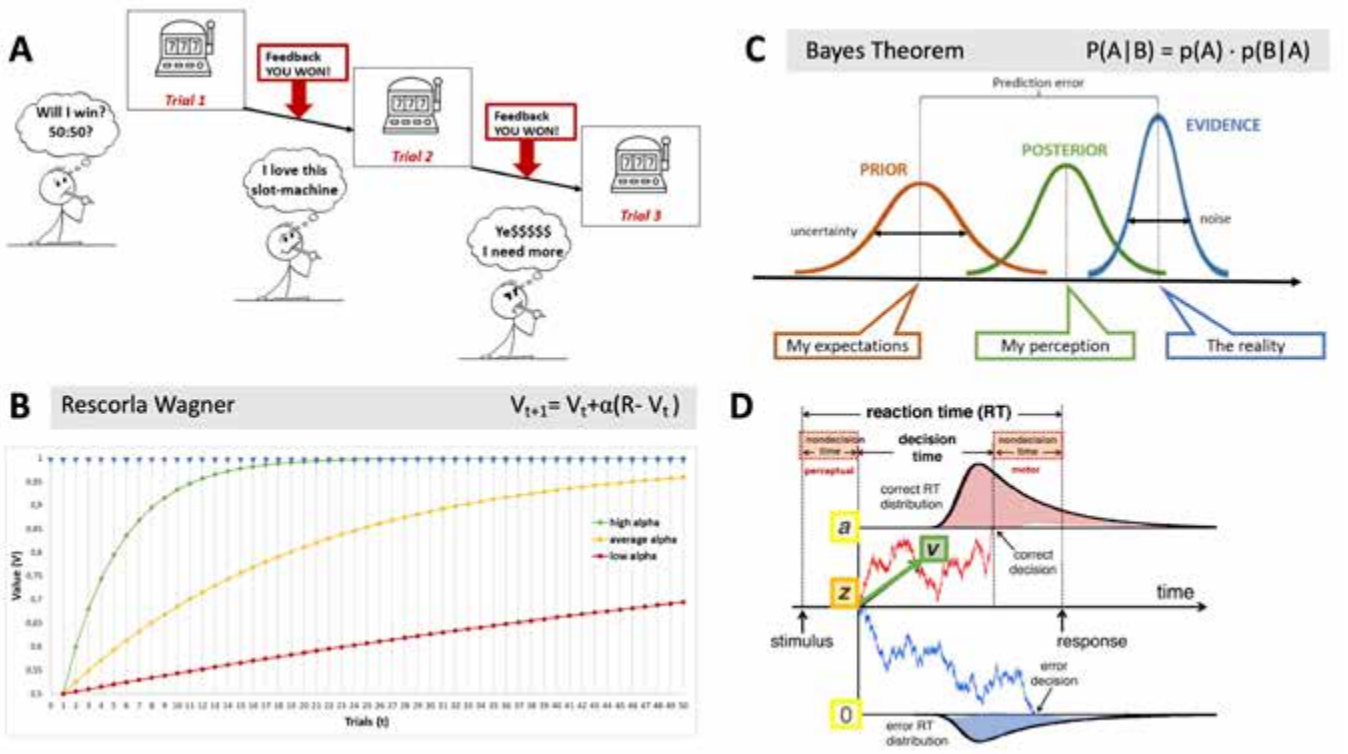


FIGURE 2. Theory-driven approaches. (A) A classic single arm bandit task. *A person interacts with a slot machine. Initially uncertain (Trial 1: "Will I win? 50:50"), positive feedback ("You won!") updates expectations, increasing confidence and value attribution (Trial 2: "I love this slot-machine"), which further strengthens with continued reward. This shows how expectations stabilise and choices become increasingly driven by learned value.* **(B) The Rescorla-Wagner model of learning.** *The equation formalises value updating, where the expected value at time t + 1 is adjusted by a fraction (α, the learning rate) of the prediction error (R - Vt). The graph shows how value estimates evolve for different learning rates.* **(C) Bayes inference as a framework for belief updating.** *Bayes' theorem describes how prior expectations (red distribution) are combined with new perceptual evidence (blue distribution) to form a posterior belief (green distribution). The prediction error represents the discrepancy between expectations and incoming evidence, with uncertainty and noise influencing the weighting of prior versus new information.* **(D) Drift Diffusion Model (DDM).** *Evidence accumulates over time from a starting point (z) toward one of two decision boundaries (a=correct, 0= error). The drift rate (v) reflects the strength of evidence. Reaction time (RT) comprises non-decision time (perceptual encoding and motor execution) plus decision time. The distribution at each boundary represents the reaction time for correct responses and error.*

Rescorla-Wagner (RW) model, which expresses learning as the incremental updating of expected value over time. In this model, the expected value at time t (Vt) is updated from the previous expectation (Vt-1) by a fraction of the prediction error (PE). PE is defined as the difference between the expected value and the actual reward (R) at each time. In this simple equation (Fig. 2B) the key parameter governing the individual differences is the learning rate (*alpha*), which determines how strongly new information influences updating in each subject: higher values reflect rapid adaptation to recent outcomes, whereas lower values reflect slower, more stable learning.

The RW equation suggests that the future value (Vt) is the product of the past (Vt-1) and the present ones (R). This updating scheme bears a conceptual resemblance to Bayesian inference, in which the prior expectations are updated in light of new evidence (likelihood) to form posterior beliefs (Fig. 2C). While RL and Bayesian models are distinct, both frameworks emphasise learning as an iterative process of expectation updating under uncertainty. This broader perspective aligns with the Bayesian Brain paradigm, which proposes that the brain – a hierarchical inference system – continuously maintains and revises interval models of the world [44], updating them more strongly when incoming information deviates from expectations [45–48]. Behaviour, in turn, can be understood as a means of the brain to sample or modify the environment, to reduce the discrepancy between predictions and observed outcomes [49-51].

If borrowing tasks from neuroeconomics and decision making deepened our understanding of how learning from reward and punishment is impaired in psychiatry, other computational models are more fit for psychophysics experiments and how evidence accumulations works and it's impaired. Drift Diffusion Models, for example, are a useful approach to simple two-choice tasks in which feedback is not given but the responses are based on the perceptual characteristic of the stimulus [52]. DDMs assume that decision-making involves the noisy accumulation of evidence over time, represented as a "drift" toward one of two boundaries corresponding to the possible choices. The speed of this drift reflects the strength of the evidence, while the boundary distance reflects the decision threshold (Fig. 2D). Random variability in the accumulation process accounts for differences in both accuracy and reaction times. Although originally developed as mathematical models of human decision-making, DDMs have also proven valid in describing neural mechanisms. In particular, the accumulation-to-boundary process in DDMs closely parallels the way populations of neurons integrate incoming evidence until a firing-rate threshold is reached, triggering a spike or action potential [53]. This convergence between behavioural modelling and neurophysiology [54] highlights the value of DDMs as a bridge between abstract decision theory and concrete brain dynamics.

Computational, generative models often reach extreme levels of complexity which may render them difficult for clinicians who don't possess a strong mathematical background. However, they represent a promising perspective to develop a true "precision pathophysiology", by providing insights into the genesis of phenomena that were previously only tackled at the descriptive level [51]. Applications include overarching models of psychopathology, of depression [55,56], schizophrenia [57] interoceptive disorders [58], hallucinations [59], and delusions [60].

Reinforcement Learning Models and Reward-Based Decision Making: an example of theory-based computational models

Among the most used tasks, the Probabilistic Reward Task (PRT, [61]) has been largely applied to assess RL in the context of mood and substance use disorders, and has been proposed as a biomarker for anhedonia. Briefly, the PRT consists of three blocks of 100 trials each with a simple line-drawing face presented with either a short (11.5 mm) or long (13 mm) mouth for 100ms. Participants are required to indicate the mouth length and received positive feedback for only 40% of correct responses ("Correct!! You won 20pts") with an asymmetrical reinforcement schedule (correct responses to one stimulus were rewarded three times more than correct responses to the other). This asymmetry in the reinforcement schedule is responsible for systematic preference towards the most rewarded stimulus, namely, *response bias*. Importantly, subjects have no information about the asymmetry, they are only instructed to win as much as possible.

Although some model-free signal-detection analysis methods are available [61,62], RL models recently emerged as useful methods to dissect RL behaviour emerging during the PRT [63] and have been applied in the context of mood [63] and substance use disorders [64]. In RL models participants' choices are modelled with a decision function that expresses selection probabilities [63]. Different RL models may include different combinations of behavioural parameters, but key parameters are usually the learning rate (α), reflecting the speed at which the system is updated based on previous trials outcomes, and the reward sensitivity (β), the responsiveness to the reward bias. Those who rapidly modify behaviour based on recent outcomes have higher learning rates, adapting quickly to environmental changes, while a low learning rate indicates slower, more rigid learning influenced by long-term trends.

Being the PRT a two-alternative forced-choice task with objective correct responses and measurable errors, it also lends itself naturally to the application of drift dif-

fusion models (DDMs). Applied to the PRT [65], DDMs can capture not only the overall accuracy and reaction times, but also trial-by-trial dynamics of how biased reinforcement learning influences evidence accumulation.

Data-driven modelling

Data-driven approaches focus on analysing large datasets to uncover patterns and relationships, not necessarily relying on predefined hypotheses[13,66,67]. These models are often built using data collected from multiple sources – such as neuroimaging, electrophysiological, clinical, and ecological data – and employ algorithms to identify recurring motifs, trends, or structures within the data. The goal is generally to use the identified patterns in observed data to make predictions about unseen data, e.g., clinical outcomes, treatment responses, or future disease progression[68]. Typically, these models are trained on a subset of data and validated on separate, unseen datasets to assess their predictive accuracy. A key feature of data-driven approaches, especially those utilizing AI, is their ability to detect complex structures within large datasets and distinguish meaningful signals from background noise.

Artificial Intelligence and Machine Learning

Artificial Intelligence (AI) refers to the broad field of developing systems capable of performing tasks that typically require human intelligence, such as reasoning, learning, or problem-solving (Fig. 3A). Within AI, Machine Learning (ML) focuses on algorithms that can learn patterns from data rather than being explicitly programmed. ML is broadly divided into supervised and unsupervised learning, each serving distinct roles in analysing complex clinical data. In supervised learning the model is trained on labelled data (*i.e.*, patient features paired with a specific diagnosis or outcome) to predict clinical categories or treatment responses (Fig. 3B). In contrast, unsupervised learning does not rely on labels but instead uncovers hidden structures in data such as subtypes or patterns [66,69] (Fig. 3B). Deep learning has emerged as a successful approach in ML, where multilayer neural networks automatically extract complex features from raw data [66,69]. Neural networks are computational models inspired by the brain, composed of interconnected layers of "neurons" that process information (Fig. 3C). Neural networks come in different "architectures", i.e. designs optimised to automatically extract features from different types of raw data. Convolutional Neural Networks (CNNs) are designed to detect spatial patterns, making them ideal for analysing images or brain scans by progressively extracting features as some visual neurons do [70]. Recurrent Neural Networks (RNNs), on the other hand, specialize in sequential data (*e.g.*, language, speech, physiological signals), as they maintain a "memory" of previous inputs to model temporal dependencies. Building on these foundations, transformer architectures introduced a new way of handling sequences by using attention mechanisms that capture long-range dependencies more efficiently than RNNs. Transformers form the backbone of today's Large Language Models (LLMs), which are trained on massive text corpora and can generate, summarize, or classify language with human-like flexibility. LLMs may also enable AI-driven chatbots to provide supportive interventions or guidance [71]. While these developments require careful ethical consideration [72], they may offer several benefits.

In addition to classification and prediction, data-driven models are used to represent psychopathology as dynamic symptom networks [3] (Fig. 3D). This approach views symptoms as interacting elements rather than mere indicators of an underlying disorder, enabling the study of symptom associations and dynamics [73–76]. These psychopathological networks aim to explain individual differences and symptom fluctuations within psychiatric syndromes[74].

A primary application of ML is diagnostic classification using neurobiological and neurocognitive data [66,77]. For instance, ML models can distinguish schizophrenia patients from healthy controls with approximately 80-90% accuracy [78], though these methods are currently less effective for depression classification [79]. Supervised ML has also proved good reliability in differential diagnosis frameworks, such as between schizophrenia and bipolar disorder [80]. Apart from classification of diagnosis, that might find only limited application in clinical practice, ML models show promise in predicting prognosis and clinical outcomes, including treatment response, personal functioning [81], and suicidal behaviours [82]. Network-based functional connectivity can predict the development of psychiatric disorders [13], and early diagnostic tools are being implemented to identify individuals at risk for psychosis [83] or depression [73,84] and to predict persistent psychotic-like experiences in longitudinal framework s[85]. Early-stage identification is perhaps the area in which ML has already achieved genuine clinical deployment: models are now used to estimate the resources needed for managing first-episode psychosis [86], and to detect emerging cases of depression [87].

Integrating Computational Psychiatry into Clinical Practice: Promises, Challenges and Future Directions

In both inpatient and outpatient care settings, CP might have notable implications [13,66,67]. Predictive models can identify individuals at heightened risk for agitation, self-

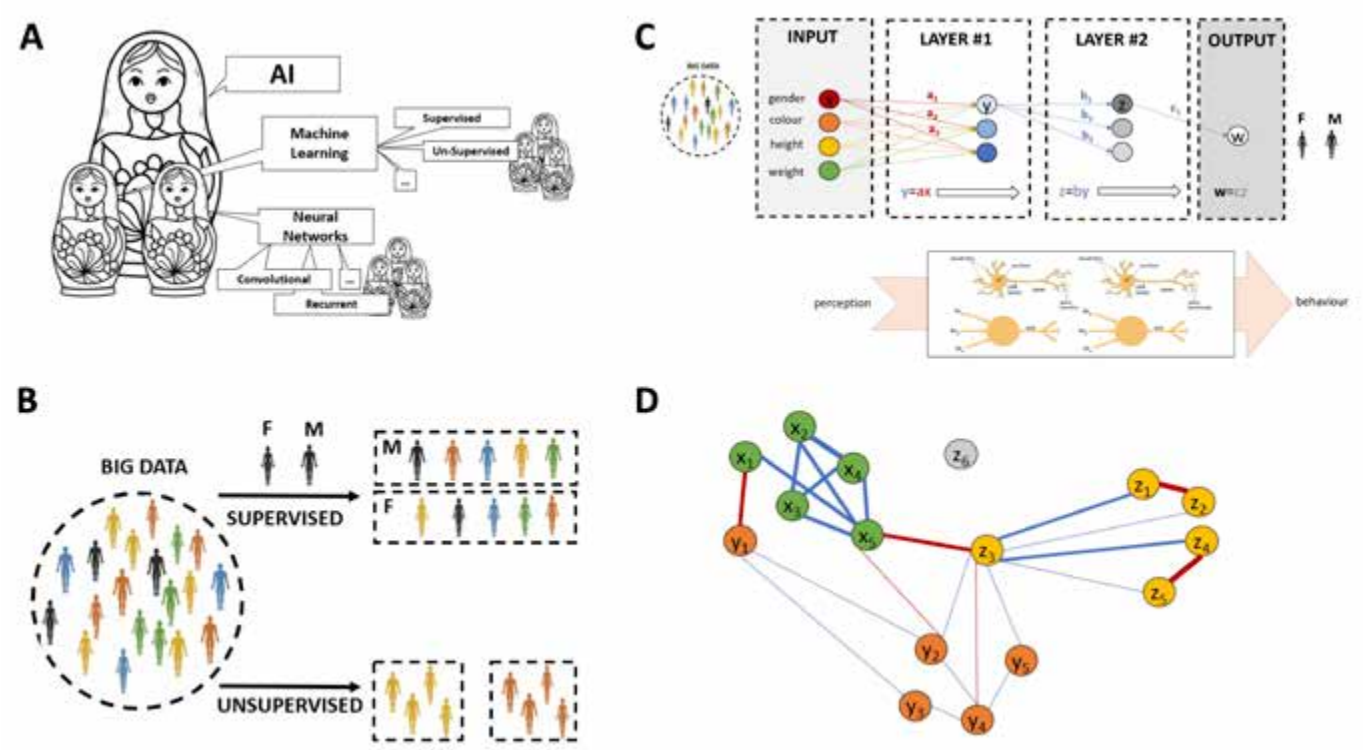


FIGURE 3. Data-drive approaches. (A) Hierarchical structure of Artificial Intelligence (AI). *Within AI, Machine Learning (ML) includes supervised and unsupervised approaches. Neural Networks, subsets of ML, comprise specialised architectures for data analysis, such as Convolutional (for spatial/image data) and Recurrent (for sequential data) networks.* **(B) Comparison of supervised and unsupervised machine learning.** *In supervised learning, models are trained on labelled data. In unsupervised learning, algorithms identify hidden structures within unlabelled data without predefined categories.* **(C) Architecture of a neural network for classification.** *Input features are processed through layers of nodes. Each layer transforms the input through weighted connections, extracting features until the output layer produces a classification. The lower panel illustrates parallels between neural networks' architecture and the neuron's structure.* **(D) Network representation of psychopathology.** *Symptoms or variables are represented as interconnected elements. Edges represent associations between symptoms, with line thickness and colours indicating the strength and direction of relationships.*

harm, or aggression, facilitating early interventions in hospitals [88]. Models derived from available data can be developed and released as public tools to identify people with increased probability of developing psychopathology (*e.g.*, https://manto.unife.it). AI-assisted pharmacogenomics has shown effectiveness in tailoring medication adjustments for optimized treatment and reduced side effects [89]. Wearable sensors and digital monitoring tools offer real-time patient data, which can help inform dynamically adjusted treatment plans [90]. Outpatient services can benefit from tools for early relapse detection in mood disorders and psychosis. Furthermore, AI-driven mental health applications can deliver personalized cognitive-behavioural therapy (CBT), while Bayesian modelling techniques enable clinicians to refine therapy approaches based on patient progress.

In general, computational approaches allow clinicians to simultaneously characterise patients across multiple levels, from molecular and neurophysiological processes to network dynamics and observable symptom dimensions, ultimately supporting more precise and individually tailored intervention strategies [91].

However, enthusiasm should be tempered by some limitations [92]. First, computational models face the same challenges as traditional statistical models in terms of

generalizability [93]. Clinical implementation needs continuous validation and up-to-date studies, as well as digital knowledge to effectively exploit these tools [94]. Particularly, AI and ML algorithms are not necessarily causally valid [95]: they are often regarded as less interpretable ("black boxes"), and this aspect might prove crucial for further applications of such techniques in psychiatry. Especially in the field of mental health, the usability and reliability of AI tools need to be carefully and continuously assessed: they should assist clinicians and not introduce further complexity; in order to be trustworthy, their limits should be rigorously verified and clarified[96]. In recent years, the concept of explainable AI ("XAI") emerged as an additional possibility to narrow the gap between the algorithms and the clinicians. In fact, clinicians need to understand how AI models reach their conclusions in order to trust and effectively use them in practice, especially given the probabilistic and multifactorial nature of psychiatric data [97]. XAI focuses on making models transparent (how the model works) and interpretable (how to understand its decisions), which together support human understandability [97]. By enhancing transparency and interpretability, XAI supports safer, more effective, and clinician-friendly AI tools in mental health care.

Finally, the implementation of these technologies often comes with significant costs, requiring specialized infrastructure and training. Clinicians may need additional knowledge and skills to interpret the outputs of computational models effectively, and integrating these tools into existing clinical workflows poses logistical challenges. For example, collecting a huge amount of data for AI could be initially feasible only in the context of university centres, thereby reducing the potential for translational application of the results, and for their access. Similarly, ensuring equitable access to these advanced technologies and addressing potential biases in algorithms are crucial considerations for responsible implementation.

Conclusions

Computational psychiatry is a transformative approach that seeks to revolutionize mental health care by integrating neuroscience, mathematical models, and clinical practice. Theory-driven and data-driven methodologies offer innovative solutions for a better understanding of psychiatric disorders across diverse settings, research and clinical questions. Theory-driven approaches, grounded in established scientific principles, build models to explain the mechanisms underlying phenomena related to psychiatry, including mental processes and behaviours. Data-driven methods leverage vast datasets and machine learning algorithms to uncover patterns and predict clinical risks and outcomes.

The advancement and responsible implementation of computational psychiatry require careful consideration of several key areas; in particular, ethical, infrastructure, and generalizability frameworks must be developed to address concerns related to data privacy, algorithmic bias, and the potential for over-reliance on automated systems. Beyond technical challenges, some inherent limitations should not be ignored: the parameters are mathematical proxies that may inadequately capture the subjective dimensions of mental disorders, and even sophisticated models risk oversimplifying phenomena that may be irreducibly biological, psychological, and social in nature [98,99]. In addition, it is crucial to invest in training initiatives to provide clinicians with the necessary background to use computational models effectively. Continued research, rigorous validation, and ongoing ethical oversight will be essential in realizing the full potential of computational psychiatry, ultimately improving patient outcomes and advancing the field of mental health treatment.

Acknowledgments

Conflict of interest

The authors have no conflict of interest to declare.

Funding

L.A. and A.R. were supported by Complementary National Plan PNC-I.1 "Research initiatives for innovative technologies and pathways in the health and welfare sector" (D.D. 931 of 06/06/2022), through the DARE - DigitAl lifelong pRevEntion initiative (code PNC0000002, CUP B53C22006420001).

Author contributions

All authors (L.A., F.C., F.F., T.J., C.M., A.M., F.A., P.O., A.P., A.R., R.R., M.B.M.) reviewed the literature and contributed to the manuscript writing and revisions. All authors approved the final manuscript.

Ethical consideration

Not applicable.

References

1 Adams RA, Huys QJM, Roiser JP. Computational Psychiatry: towards a mathematically informed understanding of mental illness. J Neurol Neurosurg Psychiatry 2016;87:53-63 . https://doi.org/10.1136/jnnp-2015-310737

2 Vasilchenko KF, Федорович ВК, Chumakov EM, et al. Current status, challenges and future prospects in computational psychiatry: a narrative review. Consortium PSYCHIATRICUM 2023;4:33-2 . https://doi.org/10.17816/CP11244

3 Borsboom D. A network theory of mental disorders. World Psychiatry: Official Journal of the World Psychiatric Association

(WPA) 2017;16:5-13.

4 Wichers M, Schreuder MJ, Goekoop R, et al. Can we predict the direction of sudden shifts in symptoms? Transdiagnostic implications from a complex systems perspective on psychopathology. Psychological Medicine 2019;49:380-7. https://doi.org/10.1017/S0033291718002064

5 American Psychiatric Association, editor. Diagnostic and statistical manual of mental disorders: DSM-5-TR™. Fifth edition, text revision. Washington, DC: American Psychiatric Association Publishing 2022.

6 World Health Organization, editor. International statistical classification of diseases and related health problems. 10th revision, 2nd edition. Geneva: World Health Organization 2004.

7 Eaton NR, Bringmann LF, Elmer T, et al. A review of approaches and models in psychopathology conceptualization research. Nature Reviews Psychology 2023;2:622-36.

8 Ghaemi SN. After the failure of DSM: clinical research on psychiatric diagnosis. World Psychiatry 2018;17:301-2. https://doi.org/10.1002/wps.20563

9 Krueger RF, Kotov R, Watson D, et al. Progress in achieving quantitative classification of psychopathology. World Psychiatry 2018;17:282-93. https://doi.org/10.1002/wps.20566

10 Leucht S, Os J, Jäger M, et al. Prioritization of Psychopathological Symptoms and Clinical Characterization in Psychiatric Diagnoses: A Narrative Review. JAMA Psychiatry 2024;81:1149-58. https://doi.org/10.1001/jamapsychiatry.2024.2652

11 Ng JY, Santoro LJ, Cobey KD, et al. alternative, and integrative medicine researchers' practices and perceived barriers related to open science: An international, cross-sectional survey. PLOS ONE 2024;19:e0301251. https://doi.org/10.1371/journal.pone.0301251

12 Ossola P, Pike AC. Editorial: What is computational psychopathology, and why do we need it? Neuroscience & Biobehavioral Reviews 2023;152:105170. https://doi.org/10.1016/j.neubiorev.2023.105170

13 Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. Nat Neurosci 2016;19:404-13. https://doi.org/10.1038/nn.4238

14 Castro Martínez JC, Santamaría-García H. Understanding mental health through computers: An introduction to computational psychiatry. Front Psychiatry 2023;14:1092471. https://doi.org/10.3389/fpsyt.2023.1092471

15 Khaleghi A, Mohammadi MR, Shahi K, et al. Computational Neuroscience Approach to Psychiatry: A Review on Theory-driven Approaches. Clin Psychopharmacol Neurosci 2022;20:26-36. https://doi.org/10.9758/cpn.2022.20.1.26

16 Montague PR, Dolan RJ, Friston KJ, et al. Computational psychiatry. Trends in Cognitive Sciences 2012;16:72-80. https://doi.org/10.1016/j.tics.2011.11.018

17 Disner SG, Beevers CG, Haigh EAP, et al. Neural mechanisms of the cognitive model of depression. Nat Rev Neurosci 2011;12:467-77. https://doi.org/10.1038/nrn3027

18 Mujica-Parodi LR, Strey HH. Making Sense of Computational Psychiatry. International Journal of Neuropsychopharmacology 2020;23:339-47. https://doi.org/10.1093/ijnp/pyaa013

19 Cicchetti D, Rogosch FA. Equifinality and multifinality in developmental psychopathology. Development and Psychopathology 1996;8:597-600. https://doi.org/10.1017/S0954579400007318

20 Shaw AK, Bisesi AT, Wojan C, et al. Six personas to adopt when framing theoretical research questions in biology. Proc Biol Sci 2024;291:20240803. https://doi.org/10.1098/rspb.2024.0803

21 Krishnadas R, Leighton SP, Jones PB. Precision psychiatry: thinking beyond simple prediction models – enhancing causal predictions. The British Journal of Psychiatry 2025;226:184-8. https://doi.org/10.1192/bjp.2024.258

22 Pizzagalli DA. Toward a Better Understanding of the Mechanisms and Pathophysiology of Anhedonia: Are We Ready for Translation? Am J Psychiatry 2022;179:458-69. https://doi.org/10.1176/appi.ajp.20220423.

23 Northoff G, Tumati S. Average is good, extremes are bad" – Non-linear inverted U-shaped relationship between neural mechanisms and functionality of mental features. Neuroscience & Biobehavioral Reviews 2019;104:11-25. https://doi.org/10.1016/j.neubiorev.2019.06.030

24 Frässle S, Yao Y, Schöbi D, et al. Generative models for clinical applications in computational psychiatry. WIREs Cognitive Science 2018;9:e1460. https://doi.org/10.1002/wcs.1460

25 Robinaugh DJ, Haslbeck JMB, Waldorp LJ, et al. Advancing the network theory of mental disorders: A computational model of panic disorder. Psychological Review 2024;131:1482-508. https://doi.org/10.1037/rev0000515

26 Friston K. Computational psychiatry: from synapses to sentience. Molecular Psychiatry 2023;28:256-68 . https://doi.org/10.1038/S41380-022-01743-Z

27 Schwartenbeck P, Friston K. Computational Phenotyping in Psychiatry: A Worked Example. Eneuro 2016;3:ENEURO.0049-16.2016. https://doi.org/10.1523/ENEURO.0049-16.2016

28 Wiecki TV, Poland J, Frank MJ. Model-Based Cognitive Neuroscience Approaches to Computational Psychiatry: Clustering and Classification. Clinical Psychological Science 2015;3:378-99. https://doi.org/10.1177/2167702614565359

29 Możaryn JF, Szczegielniak AR. COMPUTATIONAL PSYCHIATRY: A BRIDGE BETWEEN TRANSLATION AND PRECISION. International Journal of Neuropsychopharmacology 2025;28:i276-7. https://doi.org/10.1093/ijnp/pyae059.489

30 Cavanaugh JE, Neath AA. The Akaike information criterion: Background, derivation, properties, application, interpretation, and refinements. WIREs Computational Statistics 2019;11. https://doi.org/10.1002/wics.1460

31 Bennett D, Silverstein SM, Niv Y. The Two Cultures of Computational Psychiatry. JAMA Psychiatry 2019;76:563-4. https://doi.org/10.1001/jamapsychiatry.2019.0231

32 Huys QJM, Browning M, Paulus MP, et al. Advances in the computational understanding of mental illness. Neuropsychopharmacology 2021;46:3-19. https://doi.org/10.1038/s41386-020-0746-4

33 Dang J, King KM, Inzlicht M. Why Are Self-Report and Behavioral Measures Weakly Correlated? Trends Cogn Sci 2020;24:267-9 . https://doi.org/10.1016/j.tics.2020.01.007.

34 Huys Q, Vogelstein J, Dayan P. Psychiatry: Insights into depression through normative decision-making models. In: Advances in Neural Information Processing Systems. Curran Associates, Inc. 2008.

35 Huys QJM, Guitart-Masip M, Dolan RJ, et al. Decision-theoretic psychiatry. Clinical Psychological Science 2015;3:400-21. https://doi.org/10.1177/2167702614562040

36 Ossola P, Garrett N, Biso L, et al. Anhedonia and sensitivity to punishment in schizophrenia, depression and opiate use disorder. J Affect Disord 2023;330:319-28. https://doi.org/10.1016/j.jad.2023.02.120

37 Ossola P, Garrett N, Sharot T, et al. Belief updating in bipolar disorder predicts time of recurrence. eLife 2020;9:e58891. https://doi.org/10.7554/eLife.58891

38 Jeon I, Kim T. Distinctive properties of biological neural networks and recent advances in bottom-up approaches toward a better biologically plausible neural network. Front Comput Neurosci 2023;17. https://doi.org/10.3389/fncom.2023.1092185

39 Achterberg J, Akarca D, Strouse DJ, et al. Spatially embedded recurrent neural net-

works reveal widespread links between structural and functional neuroscience findings. Nat Mach Intell 2023;5:1369-81. https://doi.org/10.1038/s42256-023-00748-9

40 Volzhenin K, Changeux J-P, Dumas G. Multilevel development of cognitive abilities in an artificial neural network. Proceedings of the National Academy of Sciences 2022;119:e2201304119. https://doi.org/10.1073/pnas.2201304119

41 Sutton RS, Barto AG. Reinforcement learning: An introduction. MIT press 2018.

42 Rangel A, Camerer C, Montague PR. A framework for studying the neurobiology of value-based decision making. Nature Reviews Neuroscience 2008;9:545-56. https://doi.org/10.1038/nrn2357

43 Schultz W, Dayan P, Montague PR. A Neural Substrate of Prediction and Reward. Science 1997;275:1593-9 . https://doi.org/10.1126/science.275.5306.1593

44 Friston KJ, Stephan KE, Montague R, et al. Computational psychiatry: the brain as a phantastic organ. The Lancet Psychiatry 2014;1:148-58. https://doi.org/10.1016/S2215-0366(14)70275-5

45 Corlett PR, Fletcher PC. Computational psychiatry: a Rosetta Stone linking the brain to mental illness. The Lancet Psychiatry 2014;1:399-402. https://doi.org/10.1016/S2215-0366(14)70298-6

46 Qela B, Damiani S, Santis S, et al. Predictive coding in neuropsychiatric disorders: A systematic transdiagnostic review. Neuroscience and Biobehavioral Reviews 2025;169:106020. https://doi.org/10.1016/j.neubiorev.2025.106020

47 Smith R, Badcock P, Friston KJ. Recent advances in the application of predictive coding and active inference models within clinical neuroscience. Psychiatry and Clinical Neurosciences 2021;75:3-13. https://doi.org/10.1111/pcn.13138

48 Spratling MW. A review of predictive coding algorithms. Brain and Cognition 2017;112:92-7. https://doi.org/10.1016/j.bandc.2015.11.003

49 Friston K. The free-energy principle: a unified brain theory? Nature Reviews Neuroscience 2010;11:127-38. https://doi.org/10.1038/nrn2787

50 Friston K, FitzGerald T, Rigoli F, et al. Active Inference: A Process Theory. Neural Computation 2017;29:1-49. https://doi.org/10.1162/NECO_a_00912

51 Friston KJ. Precision Psychiatry. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging 2017;2:640-3. https://doi.org/10.1016/j.bpsc.2017.08.007

52 Ratcliff R, McKoon G. The diffusion decision model: theory and data for two-choice decision tasks. Neural Comput 2008;20:873-922. https://doi.org/10.1162/neco.2008.12-06-420

53 Shadlen MN, Newsome WT. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. J Neurophysiol 2001;86:1916-36. https://doi.org/10.1152/jn.2001.86.4.1916

54 Ossola P, Antonucci C, Meehan KB, et al. Effortful control is associated with executive attention: A computational study. Journal of Personality 2021;89:774-85. https://doi.org/10.1111/jopy.12614

55 Berg M, Feldmann M, Kirchner L, et al. Oversampled and undersolved: Depressive rumination from an active inference perspective. Neuroscience and Biobehavioral Reviews 2022;142:104873. https://doi.org/10.1016/j.neubiorev.2022.104873

56 Smith R, Lavalley CA, Taylor S, et al. Elevated decision uncertainty and reduced avoidance drives in depression, anxiety and substance use disorders during approach-avoidance conflict: a replication study. Journal of Psychiatry & Neuroscience: JPN 2023;48:E217-E231. https://doi.org/10.1503/jpn.220226

57 Friston K, Brown HR, Siemerkus J, et al. The dysconnection hypothesis. Schizophrenia Research 2016;176:83-94. https://doi.org/10.1016/j.schres.2016.07.014

58 Schoeller F, Horowitz AH, Jain A, et al. Interoceptive technologies for psychiatric interventions: From diagnosis to clinical applications. Neuroscience and Biobehavioral Reviews 2024;156:105478. https://doi.org/10.1016/j.neubiorev.2023.105478

59 Collerton D, Barnes J, Diederich NJ, et al. Understanding visual hallucinations: A new synthesis. Neuroscience and Biobehavioral Reviews 2023;150. https://doi.org/10.1016/j.neubiorev.2023.105208

60 Ashinoff BK, Singletary NM, Baker SC, et al. Rethinking delusions: A selective review of delusion research through a computational lens. Schizophr Res 2022;245:23-41. https://doi.org/10.1016/j.schres.2021.01.023

61 Pizzagalli DA, Jahn AL, O'Shea JP. Toward an objective characterization of an anhedonic phenotype: A signal-detection approach. Biological Psychiatry 2005;57:319-27. https://doi.org/10.1016/j.biopsych.2004.11.026

62 Macmillan NA, Creelman CD. Detection theory: A user's guide. 2nd Ed. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers 2005.

63 Huys QJ, Pizzagalli DA, Bogdan R, et al. Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. Biology of Mood & Anxiety Disorders 2013;3:12. https://doi.org/10.1186/2045-5380-3-12

64 Montemitro C, Ossola P, Ross TJ, et al. Longitudinal changes in reinforcement learning during smoking cessation: a computational analysis using a probabilistic reward task. Sci Rep 2024;14:32171. https://doi.org/10.1038/s41598-024-84091-y

65 Dillon DG, Belleau EL, Origlio J, et al. Using Drift Diffusion and RL Models to Disentangle Effects of Depression On Decision-Making vs. Learning in the Probabilistic Reward Task. Computational Psychiatry 2024;8:46-69. https://doi.org/10.5334/cpsy.108

66 Rutledge RB, Chekroud AM, Huys QJM. Machine learning and big data in psychiatry: toward clinical applications. Current Opinion in Neurobiology 2019;55:152-9. https://doi.org/10.1016/j.conb.2019.02.006

67 Yip SW, Barch DM, Chase HW, et al. From Computation to Clinic. Biological Psychiatry Global Open Science 2023;3:319-28. https://doi.org/10.1016/j.bpsgos.2022.03.011

68 Ressler KJ, Williams LM. Big data in psychiatry: multiomics, neuroimaging, computational modeling, and digital phenotyping. Neuropsychopharmacology 2021;46:1-2.

69 Qi GJ, Luo J. Small Data Challenges in Big Data Era: A Survey of Recent Progress on Unsupervised and Semi-Supervised Methods. IEEE Transactions on Pattern Analysis and Machine Intelligence 2022;44:2168-87. https://doi.org/10.48550/arXiv.1903.11260

70 Hubel DH, Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology (London) 1962;160:106-54 . https://doi.org/10.1113/jphysiol.1962.sp006837

71 Volkmer S, Meyer-Lindenberg A, Schwarz E. Large language models in psychiatry: Opportunities and challenges. Psychiatry Research 2024;339:116026. https://doi.org/10.1016/j.psychres.2024.116026

72 Mökander J, Schuett J, Kirk HR, et al. Auditing large language models: a three-layered approach. AI and Ethics 2024;4:1085-115. https://doi.org/10.48550/arXiv.2302.08500

73 Belvederi Murri M, Grassi L, Caruso R, et al. Depressive symptom complexes of community-dwelling older adults: a latent network model. Mol Psychiatry 2022;27:1075-82. https://doi.org/10.1038/s41380-021-01310-y

74 Briganti G, Scutari M, Epskamp S, et al. Network analysis: An overview for mental health research. International Journal of Methods in Psychiatric Research 2024;33:e2034. https://doi.org/10.1002/

mpr.2034

75 Triolo F, Belvederi Murri M, Calderón-Larrañaga A, et al. Bridging late-life depression and chronic somatic diseases: a network analysis. Translational Psychiatry 2021;11:e6. https://doi.org/10.1017/S2045796024000064

76 Corponi F, Anmella G, Verdolini N, et al. Symptom networks in acute depression across bipolar and major depressive disorders: A network analysis on a large, international, observational study. European Neuropsychopharmacology 2020;35:49-60. https://doi.org/10.1016/j.euroneuro.2020.03.017

77 Antonucci LA, Pergola G, Pigoni A, et al. A Pattern of Cognitive Deficits Stratified for Genetic and Environmental Risk Reliably Classifies Patients With Schizophrenia From Healthy Control Subjects. Biol Psychiatry 2020;87:697-707. https://doi.org/10.1016/j.biopsych.2019.11.007

78 Di Camillo F, Grimaldi DA, Cattarinussi G, et al. Magnetic resonance imaging–based machine learning classification of schizophrenia spectrum disorders: a meta-analysis. Psychiatry and Clinical Neurosciences 2024;78:732-43. https://doi.org/10.1111/pcn.13736

79 Belov V, Erwin-Grabner T, Aghajani M, et al. Multi-site benchmark classification of major depressive disorder using machine learning on cortical and subcortical measures. Scientific Reports 2024;14:1084. https://doi.org/10.1038/s41598-023-47934-8

80 Raio A, Pergola G, Rampino A, et al. Similarities and differences between multivariate patterns of cognitive and socio-cognitive deficits in schizophrenia, bipolar disorder and related risk. Schizophr 2023;9:11. https://doi.org/10.1038/s41537-023-00337-0

81 Antonucci LA, Penzel N, Sanfelici R, et al. Using combined environmental–clinical classification models to predict role functioning outcome in clinical high-risk states for psychosis and recent-onset depression. The British Journal of Psychiatry 2022;220:229-45. https://doi.org/10.1192/bjp.2022.16

82 Pigoni A, Tesic I, Pini C, et al. Multimodal Machine Learning Prediction of 12-Month Suicide Attempts in Bipolar Disorder. Bipolar Disord 2025;27:217-31. https://doi.org/10.1111/bdi.70011

83 Sanfelici R, Dwyer DB, Antonucci LA, et al. Individualized Diagnostic and Prognostic Models for Patients With Psychosis Risk Syndromes: A Meta-analytic View on the State of the Art. Biological Psychiatry 2020;88:349-60. https://doi.org/10.1016/j.biopsych.2020.02.009

84 Cattelani L, Murri MB, Chesani F, et al. Risk Prediction Model for Late Life Depression: Development and Validation on Three Large European Datasets. IEEE Journal of Biomedical and Health Informatics 2019;23:2196-204. https://doi.org/10.1109/JBHI.2018.2884079

85 Antonucci LA, Raio A, Kikidis GC, et al. Personality changes during adolescence predict young adult psychosis proneness and mediate gene-environment interplays of schizophrenia risk. Psychol Med 2024;54:1-11. https://doi.org/10.1017/S0033291724002198

86 McDonald K, Ding T, Ker H, et al. Using epidemiological evidence to forecast population need for early treatment programmes in mental health: a generalisable Bayesian prediction methodology applied to and validated for first-episode psychosis in England. The British Journal of Psychiatry 2021;219:383-91. https://doi.org/10.1192/bjp.2021.18

87 Belvederi Murri M, Cattelani L, Chesani F, et al. Risk prediction models for depression in community-dwelling older adults. The American Journal of Geriatric Psychiatry 2022;30:949-60. https://doi.org/10.1016/j.jagp.2022.05.017

88 Tay JL, Li Z, Sim K. Effectiveness of Artificial Intelligence Methods in Personalized Aggression Risk Prediction within Inpatient Psychiatric Treatment Settings—A Systematic Review. Journal of Personalized Medicine 2022;12:1470. https://doi.org/10.3390/jpm12091470

89 Kim Y, Suescun J, Schiess MC, et al. Computational medication regimen for Parkinson's disease using reinforcement learning. Scientific Reports 2021;11:9313. https://doi.org/10.1038/s41598-021-88619-4

90 Corponi F, Li BM, Anmella G, et al. Automated mood disorder symptoms monitoring from multivariate time-series sensory data: getting the full picture beyond a single number. Transl Psychiatry 2024;14:161. https://doi.org/10.1038/s41398-024-02876-1

91 Constant A, Badcock P, Friston K, et al. Integrating Evolutionary, Cultural, and Computational Psychiatry: A Multilevel Systemic Approach. Front Psychiatry 2022;13:763380. https://doi.org/10.3389/fpsyt.2022.763380

92 Karvelis P, Paulus MP, Diaconescu AO. Individual differences in computational psychiatry: A review of current challenges. Neurosci Biobehav Rev 2023;148:105137. https://doi.org/10.1016/j.neubiorev.2023.105137

93 Chekroud AM, Bondar J, Delgadillo J, et al. The promise of machine learning in predicting treatment outcomes in psychiatry. World Psychiatry 2021;20:154-70. https://doi.org/10.1002/wps.20882

94 Meehan AJ, Lewis SJ, Fazel S, et al. Clinical prediction models in psychiatry: a systematic review of two decades of progress and challenges. Molecular Psychiatry 2022;27:2700-8. https://doi.org/10.1038/s41380-022-01528-4

95 Nielsen AN, Barch DM, Petersen SE, et al. Machine Learning With Neuroimaging: Evaluating Its Applications in Psychiatry. Biological Psychiatry Cognitive Neuroscience and Neuroimaging 2020;5:791-8. https://doi.org/10.1016/j.bpsc.2019.11.007

96 Joyce S, Umney C, Whittaker X, et al. New social relations of digital technology and the future of work: Beyond technological determinism. New Technology, Work and Employment 2023;38:145-61. https://doi.org/10.1111/ntwe.12276

97 Joyce DW, Kormilitzin A, Smith KA, et al. Explainable artificial intelligence for mental health through transparency and interpretability for understandability. Npj Digital Medicine 2023;6:6. https://doi.org/10.1038/s41746-023-00751-9

98 Zavlis O, Story G, Friedrich C, et al. A systematic review of computational modeling of interpersonal dynamics in psychopathology. Nat Mental Health 2025;3:932-42. https://doi.org/10.1038/s44220-025-00465-9

99 Fried EI, Robinaugh DJ. Systems all the way down: embracing complexity in mental health research. BMC Med 2020;18:205. https://doi.org/10.1186/s12916-020-01668-w