Trends in Cognitive Sciences

CellPress

Review

Cognitive modeling of real-world behavior for understanding mental health

Dan-Mircea Mirea[1,*], Erik C. Nook[1], and Yael Niv[1,2]

A core strength of computational psychiatry is its focus on theory-driven research, in which cognitive processes are precisely quantified using computational models that formalize specific theoretical mechanisms. However, the data used in these studies often come from traditional laboratory-based cognitive tasks, which have unclear ecological validity. In this review we propose that the same theoretical frameworks and computational models can be applied to real-world data such as experience sampling, passive data, and digital-behavior data (e.g., online activity such as on social media). In turn, modeling real-world data can benefit from a theory-driven computational approach to move from purely predictive to explanatory power. We illustrate these points using emerging studies and discuss the challenges and opportunities of using real-world data in computational psychiatry.

Highlights

Computational psychiatry is most often theory-driven, aiming to identify mechanisms underlying mental health using computational cognitive models applied to behavior from laboratory-based cognitive tasks.

Cognitive tasks have unclear ecological validity, and the increasing availability of smartphone-collected, passive, and/or digital data represents an opportunity to test the generalizability of computational psychiatry findings to real-world behaviors.

Recent studies have begun to use theory-driven approaches and cognitive modeling on real-world data, sometimes uncovering previously unobserved associations with mental health symptoms.

Computational psychiatry and its reliance on cognitive tasks

In the past couple of decades, researchers have increasingly relied on computational, cognitive, and neuroscientific insights and approaches in mental health research to address the complexity of psychiatric phenomena [1]. This has germinated the field of computational psychiatry [2], which is often divided into two different but complementary families of approaches: theory-driven and data-driven [3]. Theory-driven approaches apply formal theoretical frameworks and mathematical/computational models to help explain the behavioral and neural bases of mental health [4]. By contrast, data-driven approaches leverage the power of machine learning models to predict mental health outcomes using many types of data, including self-reports sampled throughout the day or data collected passively from digital sensors or smartphone apps (sometimes referred to as 'digital phenotyping' [5,6]).

Theory-driven approaches, borrowed largely from computational neuroscience and computational cognitive science, rely mainly on analyzing how people behave in cognitive tasks. These are computerized games that are carefully designed to capture cognitive processes such as learning, decision-making, and categorization. Data from these tasks are then analyzed using computational models that formalize the studied processes [2,7], based on frameworks such as Bayesian inference [8] and reinforcement learning [9]. In computational psychiatry, there is a particular focus on estimating parameters of these computational models for each subject to precisely quantify individual differences in the underlying cognitive processes. Initial efforts in this direction have been fruitful, revealing associations between model parameters and mental health dimensions [10]. Recently, this approach has also been used to identify predictors of both pharmacological [11] and psychological [12] treatment outcomes, as well as to disentangle the cognitive mechanisms underlying different forms of psychotherapy [13].

Despite the promise of this approach, several limitations – such as low reliability and convergent validity – have recently been highlighted [14]. Task batteries and longitudinal designs that allow

[1]Department of Psychology, Princeton University, Princeton, NJ, USA
[2]Princeton Neuroscience Institute, Princeton, NJ, USA

*Correspondence: dmirea@princeton.edu (D.-M. Mirea).

modeling within-subject variation have been proposed as solutions to these problems [15,16]. Smartphones are key to this project, as they facilitate the collection of large amounts of behavioral data using experience-sampling approaches [17]. Repeated sampling of cognitive-task behavior on smartphones has uncovered both longitudinal variation in some cognitive processes (e.g., people are less risk averse for losses later in the day [18]; sensitivity to rewards fluctuates with motivation from day to day [19]) and stability in others (e.g., metacognitive biases [20]).

However, while progress has been made towards measuring psychological processes more accurately and reliably using cognitive tasks, a significant gap in ecological validity remains. Although tasks can isolate specific processes in a controlled experimental setting, they often contain decontextualized, artificial environments [21,22]. Laboratory-based cognitive tasks differ from those in the real world in at least three important ways. First, cognitive tasks have low complexity, with usually only a few stimuli and behavioral degrees of freedom. Second, the stimuli in these tasks are often not emotionally engaging or evocative; this is a particular concern in the study of mental health, which is intrinsically related to affect and emotions [23]. Third, most laboratory-based cognitive tasks lack a social component, despite social interactions constituting a large part of the stimuli people experience outside the laboratory. Social processes are also central to many mental health conditions [24], even those traditionally not considered social, such as depression [25,26] and addiction [27]. Some solutions to these issues that have been proposed [28] include gamification of tasks to increase complexity and participant engagement [29], or leveraging virtual reality for creating and delivering the tasks [30]. However, it is unclear how much a person's behavior in an artificially created environment generalizes to the real world. Moreover, most of these studies are correlational, and even those that are longitudinal estimate cognitive processes at a few timepoints rather than continuously over time. Capturing within-person fluctuations is crucial to understanding how these processes unfold in the highly variable contexts of human life.

In this review article, we advocate for using real-world data in theory-driven computational psychiatry, which infuses theory-driven approaches with the enhanced ecological validity of data-driven approaches. We explore how real-world data can be examined through the same theoretical lenses and modeled using the same cognitive computational models as cognitive task data. After introducing some types of real-world data and their benefits, we illustrate the use of such data with a few recent studies. We conclude by highlighting some opportunities for studying real-world cognition as well as challenges related to data analysis and ethical issues that accompany real-world data.

Real-world data and what they have to offer

We use the label 'real-world data' to refer to any data collected as people go about their daily activities. This includes data collected both actively, using experience sampling methodology, and passively, collected through smartphone sensors and wearable devices. We do not restrict the category to 'natural' environments, as people increasingly spend time in digital environments such as smartphone apps and social media platforms. Some of these environments involve social interactions with real people and therefore have high ecological validity.

Experience sampling data

The first class of real-world data we examine relies on experience sampling and related approaches, such as ecological momentary assessment and ambulatory assessment [31]. These approaches repeatedly measure psychological states as they occur in daily life, usually through self-report collected digitally a few times a day [32]. Researchers also commonly measure the contexts in which these states were sampled and participants' evaluations of these contexts. For example, a participant could be prompted three times a day to report their overall mood,

what they are doing at that moment, and other contextual factors such as their location. Experience sampling has been used for many years in the study of affect and mental health, revealing how symptoms dynamically unfold over time, how they can vary within the same person, and what contextual factors can cause or influence them [32]. Recently, similar approaches have been integrated into computational psychiatry, where participants are asked to perform short tasks repeatedly, with the goal of capturing within-subject variance [17]. However, this use abandons the ecological-validity benefit of experience sampling [31] due to their reliance on artificial environments. Later in this article, we will explore how traditional experience sampling, using self-report and focusing on lived experience, can also be used within theory-driven computational psychiatry.

Passive data

The second type of data is data collected passively as a person goes about their daily life. Collection is done through various devices (e.g., smartphones, smartwatches, and custom sensors embedded in clothing [33]); here, we focus on data collected through phones and watches. These data can quantify spatial location (GPS), physical activity (accelerometer), physiology (e.g., temperature, skin conductance, and heart rate), and social information (e.g., proximity to other people, recorded using Bluetooth). Such passive data have recently gained massive popularity within psychiatry, with uses ranging from monitoring symptoms [34,35] to predicting psychiatric diagnoses [36].

Digital-behavior data

Finally, we include in our taxonomy passive data that track a person's behavior as they navigate digital spaces. This includes smartphone/app-use behavior, conversation data (from texting apps), and social media data. Social media data have received particular attention in mental health research, with multiple studies identifying digital markers of depression or attention deficit hyperactivity disorder (ADHD) in social media posts [37–39,40]. The study of social media is of great public health interest, as it is still unclear whether and how social media use affects mental health. Evidence on this topic is mixed and inconclusive, leading experts to call for more research that uses a variety of methods to focus on psychological mechanisms [41,42].

Benefits of real-world data

Real-world data address most, if not all, of the limitations of task data. They are intrinsically ecologically valid and longitudinal, allowing analyses of time-varying processes as they unfold. They capture high-complexity environments, both natural and digital, which contain the emotionally meaningful stimuli that people experience in their daily life. Finally, real-world data often involve social behavior, from both in-person and virtual social interactions, the study of which can help enrich our mechanistic understanding of how cognition in social settings relates to mental health. These data are therefore promising for use within the frameworks of theory-driven computational psychiatry.

Examples of using real-world data within theory-driven computational psychiatry

We next present three emerging bodies of research that have examined different types of real-world data through a computational cognitive lens. The examples focus on reinforcement learning, the primary cognitive modeling framework used in computational psychiatry (Box 1). Some of the studies explicitly use reinforcement-learning models, while others simply use the framework as theoretical grounding. Wherever relevant, we make suggestions for further exploiting the potential of applying cognitive modeling to real-world data.

Experience sampling

Experience sampling is a promising tool for studying reinforcement learning, as it unlocks the ability to query participants regarding reinforcement-learning-related quantities, such as rewards and prediction errors, as they occur in real life. In one study, researchers leveraged this method in a

Box 1. Reinforcement learning: a primer

Reinforcement learning is a theoretical and computational framework for understanding how agents (human, animal, or AI) learn how to make optimal decisions through trial-and-error [9]. The basic idea is that agents aim to maximize reward (e.g., money) and minimize punishment (e.g., pain) and/or cost (e.g., effort). By trying out actions and observing the outcomes, agents can learn which behaviors (actions/decisions) were rewarding. Learning is achieved by updating reward expectations (usually represented as V for 'expected value', with punishment/cost treated as negative reward) using (reward) prediction errors (PE), which represent the difference between the observed and expected reward:

$$PE_t = r_t - V(S_t) \quad \text{I}$$

where r_t is the reward and $V(S_t)$ is the expected value for the state S at timepoint t. Prediction errors are then used to update the expected value for the next time this state will be encountered:

$$V(S) \leftarrow V(S_t) + \eta \cdot PE_t \quad \text{II}$$

where η is a parameter representing the learning rate or step size – the extent to which expectations are changed based on prediction errors.

The popularity of this framework is due in large part to the sizable literature on the neural basis of reinforcement learning in humans and animals [94], beginning in the 1990s with the seminal discovery that dopamine neurons in the midbrain encode reward prediction errors [95,96]. Since then, the framework has been expanded to study a range of reward-driven cognitive processes, from decision-making under risk or uncertainty to motivation [97]. Reinforcement learning was one of the first frameworks to be used in computational psychiatry, and the past decade has seen a burst of studies showing atypical reinforcement learning in people with various mental health conditions and psychiatric symptoms [98]. For example, using this framework, anhedonia (the lack of pleasure/interest in daily activities that is frequently seen in depression) has been linked to blunted sensitivity to rewards [78], whereas compulsivity – the tendency to engage in repetitive behaviors, characteristic of obsessive–compulsive disorder (OCD) and other conditions such as eating disorders – has been associated with reduced 'model-based' control, reflecting a lower tendency to use cognitive models of the environment to flexibly plan new actions at the expense of habitual behavior [99]. Given the wealth of research on the link between reinforcement learning and mental health, there is a great opportunity to test the generalizability of these findings 'in the wild' using real-world data.

unique real-world setting: students learning what exam grades to expect in successive exams in an introductory class (Figure 1A–C) [43] (see also [44]). Over the course of a semester, students were prompted to predict their grades after taking an exam. When grades were revealed, the authors computed prediction errors as the difference between the actual and predicted grades and quantified the amount of learning from these prediction errors through differences in successive grade expectations. They found that neuroticism (a personality trait associated with risk for mental health conditions [45]) moderated expectation updating, with more neurotic individuals showing a pessimistic bias and updating their predictions downward, even in the absence of a prediction error. Future work could use computational modeling to estimate individual learning rates and test whether neuroticism is associated with lower learning rates for positive prediction errors and higher learning rates for negative prediction errors: a pattern hypothesized for depression and anxiety but never found in experimental settings [46]. However, sampling of prediction errors at more timepoints might be needed for psychometrically valid estimates.

The more traditional use of experience sampling to track affective states is also valuable in computational psychiatry. Laboratory studies have found that the dynamics of mood are affected by reward prediction errors [47,48], leading to the theory that mood tracks the overall perceived change in the availability of reward [49,50]. These findings were replicated in the same exam-grade setting, where students' grade prediction errors explained their positive and negative affect more strongly than the grades themselves [51]. Changes or dysregulation in these mood computations have been hypothesized to underlie mood disorders [50,52]. Indeed, although previous laboratory-based work found no association between depression and altered behavioral or neural impacts of prediction errors [53], the exam-grade paradigm revealed a blunted effect of real-world positive prediction errors on mood in depression [54]. This result highlights the value of

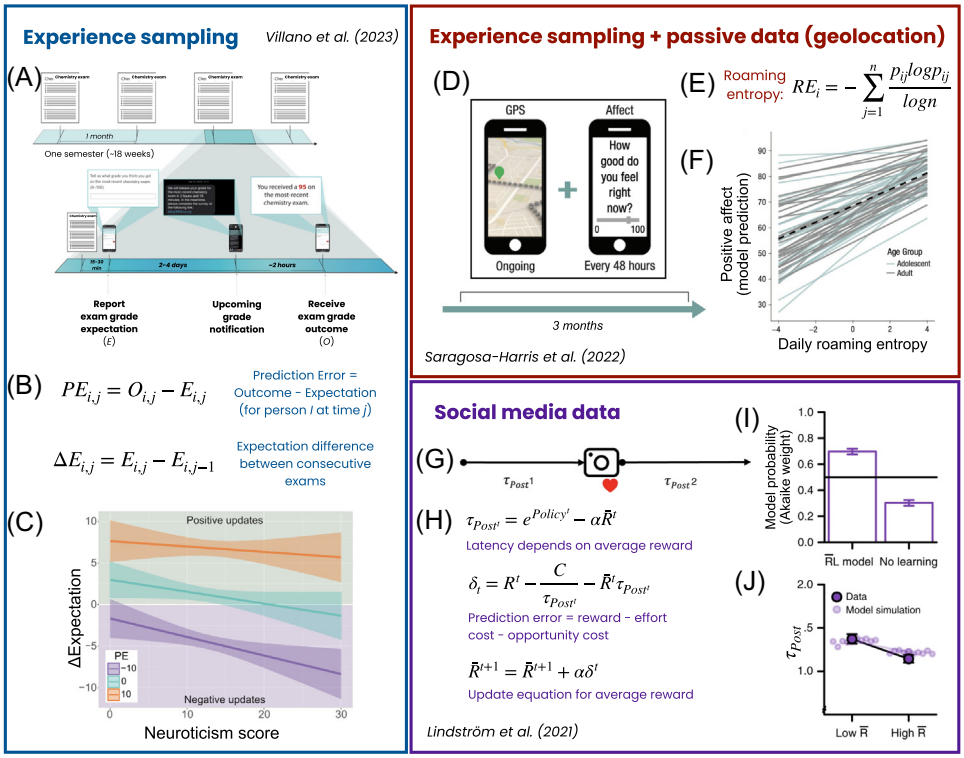

Figure 1. Examples of using real-world data in theory-driven computational psychiatry. Left panel: experience sampling of prediction errors in predicting examination grades, adapted from [43] under CC BY-NC license. (A) Schematic of the data collection set-up. (B) Equations used to compute prediction errors and expectation updates. (C) Main result: the higher the neuroticism score, the less participants increased their expectations when receiving positive prediction errors, and the more they decreased their expectations when encountering negative prediction errors. Top right panel: measuring exploration using geolocation (GPS) data (adapted from [57]). (D) Schematic of the data collection set-up. (E) Equation used to quantify exploration. (F) Main result: higher daily roaming entropy is associated with higher positive affect. Bottom right panel: modeling social media posting behavior using reinforcement-learning models (adapted from [70]). (G) Schematic highlighting the type of model used: modeling the effect of social rewards (likes) on inter-post latencies. (H) Model equations. (I) Main result 1: a reinforcement-learning model fit to the posting data explains behavior better (higher Akaike weights) than a baseline model. (J) Main result 2: the model is able to capture the key behavioral pattern seen in the human data: lower latencies following higher rates of average reward ($\bar{R}$).

studying cognition 'in the wild' for uncovering previously unobserved mechanisms underlying mental health.

The studies presented so far took advantage of a natural reinforcement-learning setting where rewards were already precisely quantified in the form of grades, but they did not measure students' decision-making behavior. However, experience sampling can also quantify reward-driven behaviors and how they are affected by real-world rewards through self-report. For example, one study showed that experiencing social interactions and physical activity as rewarding increased later engagement with those activities, whereas experiencing these activities as punishment (quantified as negative affect) reduced participation in those same activities [55] (although only learning from punishment was found in a later replication [56]). Overall, these studies offer a proof-of-principle use case of experience sampling within the framework of reinforcement learning, paving the way for more computational psychiatry work.

Passive data to complement experience sampling

Although experience sampling is a powerful tool for capturing cognitive and affective processes as they unfold in daily life, there is a limit to how frequently data can be sampled. This is where passive data can help, adding context and ecological validity beyond self-report. In one study, researchers collected GPS data tracking adolescent and adult participants' location over several months and coupled it with sampling their affect every 48 h (Figure 1D–F) [57]. For each participant, they computed daily 'roaming entropy' – a measure of how varied the locations they visited were – and found that this correlated with self-reported positive affect. In a related study, the link between roaming entropy and positive affect was stronger in individuals with stronger connectivity between the hippocampus and the ventral striatum [58]. Although these studies did not use computational models, balancing exploration of new actions and exploitation of known actions is a classic reinforcement learning problem [59], and novelty has long been thought to be intrinsically reinforcing [60]. A similar design could directly probe individual differences in the tendency to explore versus exploit by sampling rewards throughout the day and linking them to different locations, coupled with reinforcement learning models capturing decisions to return to previously rewarding locations as opposed to exploring new ones. These patterns could then be related to psychiatric symptoms. For example, atypical explore–exploit trade-offs have been described in the laboratory in neurodivergent individuals (e.g., people with ADHD tend to over-explore [61]).

Passive collection of physiological measures (e.g., heart rate, skin conductance) could also be used to estimate subjective or latent quantities relevant to cognition (e.g., rewards), allowing for denser sampling than the more burdensome experience sampling and boosting ecological validity. Using machine-learning methods to estimate affective states from physiological signals is an area of active research, with recent big-team efforts finding limited success [62]. While most studies to date are laboratory-based, using explicit emotional stimuli [63], recent work is pushing the field towards the real world [64]. One study managed to separately decode reward prediction errors, outcomes, and expectations from heart rate and portable electroencephalography (EEG) data while participants performed a probabilistic reward task at various time points during a week [65]. Interestingly, the extent to which prediction errors were decodable was predictive of future mood, suggesting that higher versus lower physiological reactivity to reward is linked to higher versus lower mood. This suggests that latent, computational quantities could be inferred from physiological data and related to fluctuations in affect. Future research has the opportunity to use these approaches to answer psychiatric questions.

Digital-behavior data

Another way to enrich self-report data obtained through experience sampling is using what we call 'digital-behavior data', reflecting behavior on one's smartphone/computer. This includes screen time, usage of different apps, conversation data from texting and other types of messaging apps (including therapy apps), and social media data. These data are not only promising for computational psychiatry but interesting in their own right, as there is a growing concern about the impact that such technology, especially social media, might have on mental health [66–68]. Directly analyzing these data is important, as there is a known discrepancy between self-reports of social media use and actual use [69].

Social media is, in many ways, an ideal real-world environment for studying (social) reinforcement-learning behaviors and their link to mental health. This is because the 'likes', followers, and comments that users receive for their posts can be thought of as social rewards that, in turn, may shape future posting behavior. One study showed that reinforcement learning models capture decisions of when to post on social media [70] (Figure 1G–J). Older theoretical work [71] and more recent work in humans [72] suggest that the optimal behavioral latency (i.e., time between

two actions) depends on a tradeoff between potential rewards and two cost terms: the effort cost and the opportunity cost of time. A higher average reward rate means that the opportunity cost of delaying the next action is higher, prescribing shorter latencies. Similarly, receiving more rewards for an action should lead to shorter latencies. This was shown empirically to apply to posting behavior on Instagram, with a recent study extending this model to include a habit component [73]. Another study using the same modeling approach found that adolescents have higher learning rates than adults, which the authors conceptualized as an increased sensitivity to social media rewards [74]. Recent work investigating the link between depression and reinforcement learning on social media found an association between depression and a heightened sensitivity to social media rewards [75]. Interestingly, this result is in line with other studies finding elevated emotional reactivity to daily events and stressors in depression [76,77], but it contrasts with laboratory-based experiments that suggest reduced sensitivity to rewards in depression [78] or an association between reward sensitivity and better mental health [79]. This finding is also in line with interpersonal theories of depression, which posit increased valuation and seeking of social feedback in depression [25].

Other types of digital-behavior data might also be useful in computational psychiatry. The study of screen time has been termed 'screenomics' and is the focus of the Human Screenome Project [80]. Snapshots of participants' phone screens are sampled throughout the day over extended periods. This type of data lends itself well to decision-making models, such as examining how people make decisions about whether to stay on an app or switch to a different one, and how this is influenced by the reward structures (gamification) of apps. Conversation data are also relevant here [81,82]. In particular, psychotherapy is a social setting itself, and the social dynamics between the client and their therapist are crucial to therapeutic success. Data from conversations between clients and therapists on therapy apps, which have been shown to track therapy outcomes [83], could offer unique insights into such social dynamics if examined using a computational lens. However, more work is needed in extending traditional reinforcement-learning frameworks to social interactions.

Other modeling frameworks and cognitive processes

In this section we have reviewed several studies that have used real-world data within a computational cognitive framework, primarily reinforcement learning. These studies illustrate the feasibility of applying theories and modeling approaches from computational psychiatry to understand daily-life behavior and shed further light on the cognitive underpinnings of mental health. Although we are not aware of any existing examples, other classes of models that have been used in computational psychiatry, such as Bayesian models (Box 2), could also be applicable to real-world data. Moreover, many cognitive processes beyond reward processing are potentially amenable to a real-world modeling approach (we provide some examples in Table 1). In the next sections we discuss challenges and opportunities of using real-world data in computational psychiatry.

Opportunities

Testing the generalizability of in-laboratory findings

At the beginning of this article we discussed some issues with cognitive tasks, including poor convergent validity and unclear generalizability [14]. Real-world data have the potential to improve the generalizability of findings in computational psychiatry due to their intrinsic ecological validity. At the very least, real-world data can be used to test the generalizability of existing findings from laboratory-based cognitive tasks. While some of the studies presented earlier have replicated in-laboratory findings (e.g., pessimistic expectations in more neurotic individuals [43]), others have uncovered novel links to psychopathology that had not been observed, or even opposite results from laboratory settings (e.g., blunted impact of positive prediction errors on mood [54] or

Box 2. Bayesian cognitive models in computational psychiatry

Although the field of computational psychiatry has focused primarily on reinforcement learning as a modeling framework, other models exist to capture cognitive processes beyond reward learning, such as belief updating, decision-making, and categorization. The major alternative class of models is Bayesian inference models, which are based on a reformulation of Bayes' theorem of conditional probabilities:

$$p(belief|observation) \propto p(observation|belief) \times p(belief) \quad \text{I}$$

Here, to update *p(belief|observation)* – the posterior probability of a belief (hypothesis) after observing some data – the model combines the prior probability of the belief *p(belief)* with the likelihood of the observation if the belief were true *p(observation|belief)*. Although Bayesian models are used in many fields for statistical inference, their conceptualization as mental belief updating lends itself well to modeling a variety of cognitive processes, from perception (where a percept is determined by integrating prior beliefs with sensory evidence) to learning (where models of Bayesian inference of the best course of action are alternative formulations to reinforcement learning). For example, Bayesian updating models have been used to model disruptions in perceptual inference in psychosis [100], impaired social learning in anxiety [101], and differences in learning about safety versus danger in anxiety and compulsivity [102].

A subclass of Bayesian inference models that have been prominent in computational psychiatry are latent-cause inference models. In these, the agent infers a latent causal structure from observations like outcomes (similar to inferring latent variables from observable ones in latent variable modeling approaches) that is presumed to have generated these observations, effectively clustering experiences [103]. Such a framework has been used to formalize several cognitive processes, including categorization, generalization, and representation learning. Individual differences in latent-cause inference have been linked to mental health. For example, latent-cause models explain Pavlovian learning better than gradual learning models in more anxious individuals [104]. Additionally, patients with post-traumatic stress disorder (PTSD) were more likely to infer a single underlying cause for both dangerous and safe contexts in fear conditioning [105].

Finally, many studies in computational psychiatry have employed the hierarchical Gaussian filter [106]. This model captures hierarchical belief structure, allowing the estimation of higher-order beliefs such as beliefs about environment volatility (i.e., how much the environment is changing). Perceiving heightened volatility in the environment has been associated with symptoms of paranoia [107], whereas a reduced capacity to adapt learning depending on volatility has been linked to symptoms of anxiety and depression [80].

To our knowledge, none of these modeling approaches has been translated to real-world settings to assess processes related to mental health. However, we believe such a future direction would align with the areas of opportunity we have identified and discussed.

heightened sensitivity to social rewards in depression [75]). Beyond testing generalization from the laboratory to the real world, a future goal of computational psychiatry 'in the wild' should be to measure the convergent validity between tasks and real-world environments, and between analogous real-world environments. This could be done through fitting similar models to task and real-world data, and correlating parameter estimates from the two (e.g., reward sensitivity on social media versus during real-world social interactions).

Incorporating language into cognitive models

Language is a window into subjective experience and therefore holds great promise in building better explanatory and predictive models of mental health [84]. However, language is rarely measured in cognitive tasks or included in cognitive models, in part because it has traditionally been difficult to quantify language with the same precision as behavior. The recent advent of large-language models (LLMs) has the potential to change this. Pre-trained LLMs such as GPT have been shown to outperform word-count-based methods and non-transformer-based classifiers in a variety of text analysis tasks, including coding of sentiment and emotional language [85] and mental health symptom prediction [86]. Because they work through flexible prompts, LLMs can be interrogated about various psychological constructs that are traditionally human-coded, although more validation research is needed to establish their accuracy. Moreover, LLM-based embeddings, which can project text into a low-dimensional vector space, allow researchers to compute the semantic similarity between chunks of text, such as social media posts. These approaches can be incorporated into cognitive models. For example, semantic

Table 1. Examples of cognitive processes that could be studied using real-world data[a]

Cognitive process	Traditional cognitive tasks	Cognitive modeling framework	Example real-world data
Reward processing and reward learning: updating expectations and behavior in response to rewards and punishments	Bandit tasks Conditioning tasks Foraging tasks	Reinforcement learning Rescorla–Wagner Bayesian inference	Exam grades [43,51,54] Social media posting [70,73,74,75] Video games [112] Gamified app behavior (e.g., Duolingo)
Decision-making: evaluating and selecting between alternatives, often under uncertainty or risk	Risky decision-making tasks Economic choice tasks Intertemporal choice/delay discounting tasks Information-seeking tasks	Drift–diffusion models Reinforcement learning Hyperbolic discounting	Purchasing decisions [113,114] Gambling behavior [115] Web-browsing behavior [116,117] Self-report of information-seeking strategies [118]
Planning: mentally organizing or sequencing actions to achieve future goals	Tower of Hanoi Navigation tasks	Tree search Model-based reinforcement learning	Chess gameplay [119] Planning apps Navigation (+ sequencing, e.g., while shopping) [57,58]
Cognitive control: maintaining goal-directed behavior in the face of distraction or competition	Stroop task Stop-signal task Oddball go/no-go tasks	Neural network models Drift–diffusion models	Compulsive phone/social media use (e.g., checking) [75,120] Dieting behavior
Attention: prioritizing certain stimuli or tasks over distractions	Posner cueing tasks	Neural network models Drift–diffusion models	Web-browsing or video gameplay (with eye-tracking) [116]
Memory: encoding, storing and retrieving information over time	Recall/recognition/source memory tests	Connectionist/neural network models Temporal context model	Web searches Spontaneous reminiscing in conversations Note-taking app behavior
Perceptual decision-making: accumulating sensory evidence to make judgements about stimuli	Random dot motion tasks	Drift–diffusion models Neural network models	Captcha task solving Voice assistant logs
Social cognition: reasoning about the cognitive processes and behaviors of other people and behaving according to these inferences	False belief tasks	Bayesian inference	Conversation data Multiplayer game behavior [121,122]

[a]Where possible, we cite recent studies that have used each type of real-world data with a cognitive lens. In most cases, there has not been any application to mental health research.

similarity derived from LLM embeddings could be used to compute the extent to which a person continues the same topic or changes to a new one in their social media posts or text messages, a kind of linguistic explore–exploit measure (Figure 2A). Computational models could then combine this behavioral metric with measures of reward, either explicit (e.g., social media likes) or language-derived (from comments and direct messages). Language is also amenable to Bayesian modeling approaches: for example, capturing how someone updates their beliefs in response to evidence from another person in a conversation (Figure 2B), or identifying latent semantic clusters and how they are started or traversed as a person generates language, which possibly reflects their underlying mental representations (Figure 2C). Overall, there is an opportunity for researchers to incorporate language into their cognitive models, though further research should address the feasibility of this approach.

Challenges

Unexplained variance

Real-world behavior inherently has many more degrees of freedom than behavior in constrained cognitive tasks. This means that there will be more unexplained variance (noise) when modeling real-world data. For example, social media posting may be influenced by rewards such as 'likes', but also by other potentially idiosyncratic factors such as the time of day and constraints

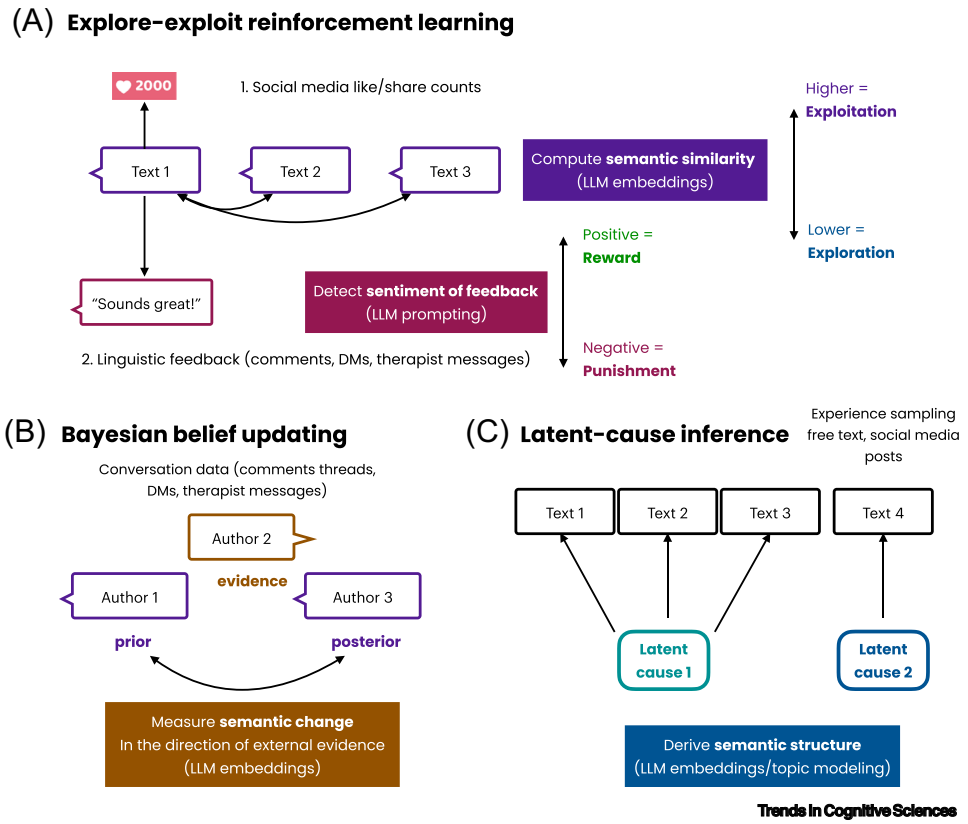

Figure 2. Incorporating language into cognitive models using large language models (LLMs). (A) Semantic similarity between successive chunks of text (e.g., successive social media posts or messages sent in a texting app) can be computed using LLM embeddings. Social rewards received for each message or post can be either retrieved from data (e.g., objective like/share counts on social media) or inferred from linguistic measures (e.g., agreement or disagreement in subsequent conversation). This allows studying the modulation of language by social rewards within a reinforcement-learning framework, in the same way as studying the modulation of other behaviors by reward. (B) LLM embeddings can also be used with Bayesian belief updating, by computing semantic change in the direction of external evidence (e.g., before and after receiving a reply in a comment thread, direct conversation, or therapeutic input). The setup mimics a Bayesian updating process, with the first message representing the prior belief, evidence is provided in the form of a message from another author, and the posterior belief is evidenced by the reply from the original author. (C) Language data are also amenable to a latent-cause inference framework, which is similar to some natural language processing models like topic models: both approaches estimate the extent to which people 'cluster' experiences and how they move through clusters or topics. Thus, these models can be fit to text from social media or experience sampling to derive a latent semantic structure and examine individual differences in how people start new semantic clusters or switch between semantic clusters, which could be altered in psychopathology. Abbreviation: DM, direct message.

on posting (e.g., due to being at school or at work). This is both a curse and a potential blessing. On the one hand, models explaining only a small fraction of variance exacerbate challenges in statistical inference and hypothesis testing (e.g., larger sample sizes are required to detect a signal, and/or noise reduction techniques should be used) and diminish the practical significance of findings due to small effect sizes (that said, scholars argue that it is important that we take small effect sizes seriously, both for replicability and for understanding their cumulative impact over time [87,88]). On the other hand, unexplained variance offers opportunities for (i) testing contextual moderators, (ii) modeling additional systematic sources of variance that are harder for laboratory studies to capture (e.g., circadian rhythms), and (iii) uncovering other patterns of within-subject variation in continuously collected passive data. Thus, additional variance might be scientifically meaningful, potentially leading to an expansion of our theories of human behavior.

Analytical challenges

Less constrained data also pose specific challenges for analysis and modeling. For example, real-world data can be on very different scales for different people (e.g., the distribution of likes for social media influencers is orders of magnitude greater than for casual social media users). This underscores the importance of separating between-subject and within-subject variance [89]. Moreover, the many types of real-world data will require extensions of traditional modeling approaches [22,90], which have been developed mainly for task data constrained to choices and reaction times. A particular challenge is performing causal inference with purely observational data. In laboratory experiments, causal inference about the drivers of behavior can be achieved through careful manipulation of task conditions while controlling for nuisance variables. This is not possible for real-world data, where the relationship between behavior and stimuli/rewards (e.g., posting and likes on social media) is often bidirectional, autocorrelated, and confounded by third variables. Incorporating causal inference methods for observational data from other social sciences, such as difference-in-difference event studies [91], could be beneficial as the field moves forward.

Data collection burden

Finally, although passive data allow the continuous assessment of behavior and socio-cognitive processes, data collection is not without burden. For instance, monitoring or sharing data can use up battery, requiring participants to charge their device more often, use multiple devices, or even replace batteries more frequently [92]. Data donation is burdensome, time-consuming, and ethically aversive for participants, making it difficult to collect large samples [93]. Relatedly, we highlight important ethical considerations when working with digital and passive data in Box 3.

Concluding remarks

We have explored the synergy between real-world data and cognitive theories and models in computational psychiatry. This synergy could improve the ecological validity and generalizability

Outstanding questions

What types of real-world data are amenable to cognitive modeling, and conversely, what classes of cognitive models are applicable to real-world behavior?

What adjustments to current models are needed to accommodate the additional variance of real-world behavior?

How correlated are computational measures derived from cognitive tasks and from real-world behavior?

To what extent can modeling real-world data reveal new associations with psychiatric symptoms or provide new predictors of treatment response that are not already known from laboratory-based studies?

How can cognitive tasks, real-world data, computational modeling, and text analysis using LLMs be optimally integrated to advance computational psychiatry?

Box 3. The ethical challenges of using digital behavior and passive data

Working with real-world data that are acquired incidentally (e.g., from publicly available social media posts or other passive means) poses unique ethical challenges, especially when these data are related to mental health outcomes. Several issues have been described in the literature, such as informed consent (is there consent and, if so, do participants fully understand the implications of giving this consent?) and data protection and privacy (how are the data processed and stored and how is identification prevented?) [108]. The use of LLMs brings its own share of issues, from concerns around consent and privacy in training data to algorithmic biases [109].

One grave concern is that the products of research done with informed consent (e.g., an algorithm that predicts mental health conditions based on linguistic data) will later be applied to other people without their consent, potentially by ill-intending actors (e.g., by social media companies to target ads, or worse, to prevent employment or other such discriminations based on mental health). This is particularly concerning for predictive modeling; however, any explanatory model could potentially be used for predictive purposes as well. The risk is compounded by the fact that passive/digital data are often difficult to de-identify, and this concern is especially pronounced with highly sensitive, health-related data (e.g., physiological data such as heart rate, or activity/GPS data). Another ethically challenging case is social media, as at least some of the data are publicly available, which makes it both a gray area in terms of regulation by ethics committees and an easy point of exploitation by bad actors [110].

With the recent closing down of public access to data from many social media platforms through application programming interfaces (APIs), researchers are orienting themselves more and more towards data donation strategies [111]. These more closely resemble the typical process of data acquisition for other types of real-world data and can be more tightly regulated, providing a slight ethical benefit against misuse by bad actors. However, digital data (even beyond social media) continue to be used and likely misused internally by the corporations that own these data. At a large scale, regulatory powers should mandate access to digital data for researchers under strict ethical guidelines that address the concerns listed earlier. On a small scale, individual researchers should think critically about and consult or create ethical guidelines for their particular data type and use case, perhaps even going beyond what their local ethics committee recommends.

of computational psychiatry findings and extend the use of real-world data from purely predictive to explanatory. Given advances in data availability and modeling approaches, including models of natural language, computational psychiatry 'in the wild' might be the future of mental health research, although analytical and ethical challenges should be carefully considered. Immediate next steps for the field include: (i) identifying amenable types of real-world data, (ii) integrating real-world data into data collection pipelines (ideally alongside laboratory-based tasks), (iii) assessing links to mental health, (iv) assessing laboratory-based versus real-world convergent validity of parameters, and (v) expanding existing models (see Outstanding questions). Whether this endeavor will indeed provide meaningful solutions to the puzzle of mental health mechanisms remains to be seen.

Acknowledgments

D-M.M. is supported by a Princeton Precision Health Grant. E.C.N. is supported by a NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation and a Princeton Precision Health Grant. Y.N. is supported by NIMH Conte Center grant P50MH136296 and NIMH grant R01MH119511.

Declaration of interests

The authors declare no competing interests.

References

1. Morris, S.E. and Cuthbert, B.N. (2012) Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. *Dialogues Clin. Neurosci.* 14, 29–37
2. Huys, Q.J.M. *et al.* (2016) Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* 19, 404–413
3. Bennett, D. *et al.* (2019) The two cultures of computational psychiatry. *JAMA Psychiatry* 76, 563–564
4. Maia, T.V. *et al.* (2017) Theory-based computational psychiatry. *Biol. Psychiatry* 82, 382–384
5. Galatzer-Levy, I.R. and Onnela, J.-P. (2023) Machine learning and the digital measurement of psychological health. *Annu. Rev. Clin. Psychol.* 19, 133–154
6. Torous, J. *et al.* (2021) The growing field of digital psychiatry: current evidence and the future of apps, social media, chatbots, and virtual reality. *World Psychiatry* 20, 318–335
7. Hauser, T.U. *et al.* (2022) The promise of a model-based psychiatry: building computational models of mental ill health. *Lancet Digit. Health* 4, e816–e828
8. Griffiths, T.L. *et al.* (2008) Bayesian models of cognition. In *The Cambridge Handbook of Computational Psychology* (Sun, R., ed.), pp. 59–100, Cambridge University Press
9. Sutton, R.S. and Barto, A.G. (2018) *Reinforcement Learning, An Introduction* (2nd edn), 2018. MIT Press
10. Huys, Q.J.M. *et al.* (2021) Advances in the computational understanding of mental illness. *Neuropsychopharmacology* 46, 3–19
11. Berwian, I.M. *et al.* (2020) Computational mechanisms of effort and reward decisions in patients with depression and their association with relapse after antidepressant discontinuation. *JAMA Psychiatry* 77, 513–522
12. Reiter, A.M.F. *et al.* (2021) Neuro-cognitive processes as mediators of psychological treatment effects. *Curr. Opin. Behav. Sci.* 38, 103–109
13. Norbury, A. *et al.* (2024) Different components of cognitive-behavioral therapy affect specific cognitive mechanisms. *Sci. Adv.* 10, eadk3222
14. Karvelis, P. *et al.* (2023) Individual differences in computational psychiatry: a review of current challenges. *Neurosci. Biobehav. Rev.* 148, 105137
15. Schurr, R. *et al.* (2024) Dynamic computational phenotyping of human cognition. *Nat. Hum. Behav.* 8, 917–931
16. Hitchcock, P.F. (2022) Computational psychiatry needs time and context. *Annu. Rev. Psychol.* 73, 243–270
17. Gillan, C.M. and Rutledge, R.B. (2021) Smartphones and the neuroscience of mental health. *Annu. Rev. Neurosci.* 44, 129–151
18. Bedder, R.L. *et al.* (2023) Risk taking for potential losses but not gains increases with time of day. *Sci. Rep.* 13, 5534
19. Hewitt, S.R.C. *et al.* (2025) Day-to-day fluctuations in motivation drive effort-based decision-making. *Proc. Natl. Acad. Sci.* 122, e2417964122
20. Fox, C.A. *et al.* (2024) Reliable, rapid, and remote measurement of metacognitive bias. *Sci. Rep.* 14, 14941
21. Dawson, D.R. and Marcotte, T.D. (2017) Special issue on ecological validity and cognitive assessment. *Neuropsychol. Rehabil.* 27, 599–602
22. Maselli, A. *et al.* (2023) Beyond simple laboratory studies: developing sophisticated models to study rich behavior. *Phys. Life Rev.* 46, 220–244
23. Gross, J.J. and Jazaieri, H. (2014) Emotion, emotion regulation, and psychopathology: an affective science perspective. *Clin. Psychol. Sci.* 2, 387–401
24. Busfield, J. (2000) Introduction: rethinking the sociology of mental health. *Sociol. Health Illn.* 22, 543–558
25. Hames, J.L. *et al.* (2013) Interpersonal processes in depression. *Annu. Rev. Clin. Psychol.* 9, 355–377
26. Slavich, G.M. and Sacher, J. (2019) Stress, sex hormones, inflammation, and major depressive disorder: extending social signal transduction theory of depression to account for sex differences in mood disorders. *Psychopharmacology* 236, 3063–3079
27. Pickard, H. (2021) Addiction and the self. *Noûs* 55, 737–761
28. Wise, T.K. *et al.* (2024) Naturalistic reinforcement learning. *Trends Cogn. Sci.* 28, 144–158
29. Allen, K. *et al.* (2024) Using games to understand the mind. *Nat. Hum. Behav.* 8, 1035–1043
30. Hakim, A. and Hammad, S. (2022) Use of virtual reality in psychology. In *Digital Interaction and Machine Intelligence* (Biele, C. *et al.*, eds), pp. 208–217, Springer
31. Mestdagh, M. and Dejonckheere, E. (2021) Ambulatory assessment in psychopathology research: current achievements and future ambitions. *Curr. Opin. Psychol.* 41, 1–8
32. Myin-Germeys, I. *et al.* (2018) Experience sampling methodology in mental health research: new insights and technical developments. *World Psychiatry* 17, 123–132
33. Vijayan, V.J. *et al.* (2021) Review of wearable devices and data collection considerations for connected health. *Sensors (Basel)* 21, 5589

34. Sheikh, M. *et al.* (2023) Wearable, environmental, and smartphone-based passive sensing for mental health monitoring. *Front. Digit. Health* 3, 662811
35. De Angel, V. *et al.* (2022) Digital health tools for the passive monitoring of depression: a systematic review of methods. *npj Digit. Med.* 5, 3
36. Chikersal, P. *et al.* (2021) Detecting depression and predicting its onset using longitudinal symptoms captured by passive sensing: a machine learning approach with robust feature selection. *ACM Trans. Comput. Hum. Interact.* 28, 1–41
37. Eichstaedt, J.C. *et al.* (2018) Facebook language predicts depression in medical records. *Proc. Natl. Acad. Sci.* 115, 11203–11208
38. Bathina, K.C. *et al.* (2021) Individuals with depression express more distorted thinking on social media. *Nat. Hum. Behav.* 5, 458–466
39. Kelley, S.W. and Gillan, C.M. (2022) Using language in social media posts to study the network dynamics of depression longitudinally. *Nat. Commun.* 13, 870
40. Guntuku, S.C. *et al.* (2019) Language of ADHD in adults on social media. *J. Atten. Disord.* 23, 1475–1485
41. Parry, D.A. *et al.* (2022) Social media and well-being: a methodological perspective. *Curr. Opin. Psychol.* 45, 101285
42. Orben, A. *et al.* (2024) Mechanisms linking social media use to adolescent mental health vulnerability. *Nat. Rev. Psychol.* 3, 407–423
43. Villano, W.J. *et al.* (2023) Individual differences in naturalistic learning link negative emotionality to the development of anxiety. *Sci. Adv.* 9, eadd2976
44. Vandendriessche, H. and Palminteri, S. (2023) Neurocognitive biases from the lab to real life. *Commun. Biol.* 6, 158
45. Ormel, J. *et al.* (2013) Neuroticism and common mental disorders: meaning and utility of a complex relationship. *Clin. Psychol. Rev.* 33, 686–697
46. Bishop, S.J. and Gagne, C. (2018) Anxiety, depression, and decision making: a computational perspective. *Annu. Rev. Neurosci.* 41, 371–388
47. Rutledge, R.B. *et al.* (2014) A computational and neural model of momentary subjective well-being. *Proc. Natl. Acad. Sci.* 111, 12252–12257
48. Eldar, E. and Niv, Y. (2015) Interaction between emotional state and learning underlies mood instability. *Nat. Commun.* 6, 6149
49. Eldar, E. *et al.* (2016) Mood as representation of momentum. *Trends Cogn. Sci.* 20, 15–24
50. Eldar, E. *et al.* (2021) Positive affect as a computational mechanism. *Curr. Opin. Behav. Sci.* 39, 52–57
51. Villano, W.J. *et al.* (2020) Temporal dynamics of real-world emotion are more strongly linked to prediction error than outcome. *J. Exp. Psychol. Gen.* 149, 1755–1766
52. Mason, L. *et al.* (2017) Mood instability and reward dysregulation – a neurocomputational model of bipolar disorder. *JAMA Psychiatry* 74, 1275–1276
53. Rutledge, R.B. *et al.* (2017) Association of neural and emotional impacts of reward prediction errors with major depression. *JAMA Psychiatry* 74, 790–797
54. Villano, W.J. and Heller, A.S. (2024) Depression is associated with blunted affective responses to naturalistic reward prediction errors. *Psychol. Med.* 54, 1956–1964
55. Wichers, M. *et al.* (2015) From affective experience to motivated action: tracking reward-seeking and punishment-avoidant behaviour in real-life. *PLoS One* 10, e0129722
56. Heininga, V.E. *et al.* (2017) Reward and punishment learning in daily life: a replication study. *PLoS One* 12, e0180753
57. Saragosa-Harris, N.M. *et al.* (2022) Real-world exploration increases across adolescence and relates to affect, risk taking, and social connectivity. *Psychol. Sci.* 33, 1664–1679
58. Heller, A.S. *et al.* (2020) Association between real-world experiential diversity and positive affect relates to hippocampal–striatal functional connectivity. *Nat. Neurosci.* 23, 800–804
59. Dayan, P. and Daw, N.D. (2008) Decision theory, reinforcement learning, and the brain. *Cogn. Affect. Behav. Neurosci.* 8, 429–453
60. Berlyne, D.E. (1970) Novelty, complexity, and hedonic value. *Percept. Psychophys.* 8, 279–286
61. Addicott, M.A. *et al.* (2021) Attention-deficit/hyperactivity disorder and the explore/exploit trade-off. *Neuropsychopharmacology* 46, 614–621
62. Coles, N. *et al.* (2025) Big team science reveals promises and limitations of machine learning efforts to model the physiological basis of affective experience. *R. Soc. Open Sci.* Published June 25, 2025. https://doi.org/10.1098/rsos.241778
63. Khalid, M. and Willis, E. (2022) A brief survey of machine learning methods for emotion prediction using physiological data. *arXiv* Published online January 17, 2022. http://doi.org/10.48550/arXiv.2201.06610
64. Hoemann, K. *et al.* (2021) Investigating the relationship between emotional granularity and cardiorespiratory physiological activity in daily life. *Psychophysiology* 58, e13818
65. Eldar, E. *et al.* (2018) Decodability of reward learning signals predicts mood fluctuations. *Curr. Biol.* 28, 1433–1439.e7
66. Scott, D.A. *et al.* (2017) Mental health concerns in the digital age. *Int. J. Ment. Heal. Addict.* 15, 604–613
67. Abi-Jaoude, E. *et al.* (2020) Smartphones, social media use and youth mental health. *CMAJ* 192, E136–E141
68. Corpuz, J.C.G. (2023) Metaverse: a public health concern? *J. Public Health* 45, e591
69. Parry, D.A. *et al.* (2021) A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nat. Hum. Behav.* 5, 1535–1547
70. Lindström, B. *et al.* (2021) A computational reward learning account of social media engagement. *Nat. Commun.* 12, 1311
71. Niv, Y. *et al.* (2007) Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology* 191, 507–520
72. Constantino, S.M. and Daw, N.D. (2015) Learning the opportunity cost of time in a patch-foraging task. *Cogn. Affect. Behav. Neurosci.* 15, 837–853
73. Turner, G. *et al.* (2024) A computational model of reward learning and habits on social media. *OSF* Published online November 8, 2024. http://doi.org/10.31234/osf.io/xe25k
74. da Silva Pinho, A. *et al.* (2024) Youths' sensitivity to social media feedback: a computational account. *Sci. Adv.* 10, eadp8775
75. Mirea, D.-M. *et al.* (2024) Depression is associated with higher sensitivity to social media rewards. *OSF* Published online June 21, 2024. http://doi.org/10.31234/osf.io/4ynbc
76. Myin-Germeys, I. *et al.* (2003) Emotional reactivity to daily life stress in psychosis and affective disorder: an experience sampling study. *Acta Psychiatr. Scand.* 107, 124–131
77. Bylsma, L.M. *et al.* (2011) Emotional reactivity to daily events in major and minor depression. *J. Abnorm. Psychol.* 120, 155–167
78. Huys, Q.J. *et al.* (2013) Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis. *Biol. Mood Anxiety Disord.* 3, 12
79. Blain, B. *et al.* (2023) Sensitivity to intrinsic rewards is domain general and related to mental health. *Nat. Mental Health* 1, 679–691
80. Reeves, B. *et al.* (2021) Screenomics: a framework to capture and analyze personal life experiences and the ways that technology shapes them. *Hum. Comput. Interact.* 36, 150–201
81. Stamatis, C.A. *et al.* (2022) Prospective associations of text-message-based sentiment with symptoms of depression, generalized anxiety, and social anxiety. *Depress. Anxiety* 39, 794–804
82. Meyerhoff, J. *et al.* (2023) Analyzing text message linguistic features: do people with depression communicate differently with their close and non-close contacts? *Behav. Res. Ther.* 166, 104342
83. Nook, E.C. *et al.* (2022) Linguistic measures of psychological distance track symptom levels and treatment outcomes in a large set of psychotherapy transcripts. *Proc. Natl. Acad. Sci.* 119, e2114737119
84. Nook, E.C. (2023) The promise of affective language for identifying and intervening on psychopathology. *Affect. Sci.* 4, 517–521
85. Rathje, S. *et al.* (2024) GPT is an effective tool for multilingual psychological text analysis. *Proc. Natl. Acad. Sci.* 121, e2308950121
86. Hur, J.K. *et al.* (2024) Language sentiment predicts changes in depressive symptoms. *Proc. Natl. Acad. Sci.* 121, e2321321121

87. Funder, D.C. and Ozer, D.J. (2019) Evaluating effect size in psychological research: sense and nonsense. *Adv. Methods Pract. Psychol. Sci.* 2, 156–168
88. Götz, F.M. *et al.* (2022) Small effects: the indispensable foundation for a cumulative psychological science. *Perspect. Psychol. Sci.* 17, 205–215
89. Bolger, N. and Laurenceau, J.-P. (2013) *Intensive Longitudinal Methods: An Introduction to Diary and Experience Sampling Research*, Guilford Press, p. xv, 256
90. Collins, A.G.E. and Shenhav, A. (2022) Advances in modeling learning and decision-making in neuroscience. *Neuropsychopharmacology* 47, 104–118
91. Li, Z. and Strezhnev, A. (2024) A guide to dynamic difference-in-differences regressions for political scientists. *OSF* Published online June 25, 2024. http://doi.org/10.31235/osf.io/kxw92
92. Boonstra, T.W. *et al.* (2018) Using Mobile Phone Sensor Technology for Mental Health Research: Integrated Analysis to Identify Hidden Challenges and Potential Solutions. *J. Med. Internet Res.* 20, e10131
93. van Driel, I.I. *et al.* (2022) Promises and Pitfalls of Social Media Data Donations. *Commun. Methods Meas.* 16, 266–282
94. Niv, Y. (2009) Reinforcement learning in the brain. *J. Math. Psychol.* 53, 139–154
95. Montague, P.R. *et al.* (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J. Neurosci.* 16, 1936–1947
96. Schultz, W. *et al.* (1997) A neural substrate of prediction and reward. *Science* 275, 1593–1599
97. Niv, Y. *et al.* (2006) A normative perspective on motivation. *Trends Cogn. Sci.* 10, 375–381
98. Chen, C. *et al.* (2015) Reinforcement learning in depression: a review of computational research. *Neurosci. Biobehav. Rev.* 55, 247–267
99. Gillan, C.M. *et al.* (2016) Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. *eLife* 5, e11305
100. Horga, G. and Abi-Dargham, A. (2019) An integrative framework for perceptual disturbances in psychosis. *Nat. Rev. Neurosci.* 20, 763–778
101. Lamba, A. *et al.* (2020) Anxiety impedes adaptive social learning under uncertainty. *Psychol. Sci.* 31, 592–603
102. Wise, T. and Dolan, R.J. (2020) Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. *Nat. Commun.* 11, 4179
103. Mirea, D.-M. *et al.* (2024) The ubiquity of time in latent-cause inference. *J. Cogn. Neurosci.* 36, 2442–2454
104. Zika, O. *et al.* (2023) Trait anxiety is associated with hidden state inference during aversive reversal learning. *Nat. Commun.* 14, 4203
105. Norbury, A. *et al.* (2021) Latent cause inference during extinction learning in trauma-exposed individuals with and without PTSD. *Psychol. Med.* 2021, 1–12
106. Mathys, C.D. *et al.* (2014) Uncertainty in perception and the Hierarchical Gaussian Filter. *Front. Hum. Neurosci.* 8, 825
107. Reed, E.J. *et al.* (2020) Paranoia as a deficit in non-social belief updating. *eLife* 9, e56345
108. Shen, F.X. *et al.* (2022) An ethics checklist for digital health research in psychiatry: viewpoint. *J. Med. Internet Res.* 24, e31146
109. Thirunavukarasu, A.J. *et al.* (2023) Large language models in medicine. *Nat. Med.* 29, 1930–1940
110. Chancellor, S. *et al.* (2019) A taxonomy of ethical tensions in inferring mental health states from social media. In *Proceedings of the Conference on Fairness, Accountability, and Transparency, in FAT* '19*, pp. 79–88, Association for Computing Machinery, New York, NY, USA
111. Ohme, J. *et al.* (2023) Digital trace data collection for social media effects research: APIs, data donation, and (screen) tracking. *Commun. Methods Meas.* 18, 1–18
112. Brändle, F. *et al.* (2023) Empowerment contributes to exploration behaviour in a creative video game. *Nat. Hum. Behav.* 7, 1481–1489
113. Cotet, M. *et al.* (2025) Deliberation during online bargaining reveals strategic information. *Proc. Natl. Acad. Sci.* 122, e2410956122
114. Schulz, E. *et al.* (2019) Structured, uncertainty-driven exploration in real-world consumer choice. *Proc. Natl. Acad. Sci.* 116, 13903–13908
115. Otto, A.R. *et al.* (2016) Unexpected but incidental positive outcomes predict real-world gambling. *Psychol. Sci.* 27, 299–311
116. Kelly, C.A. and Sharot, T. (2025) Web-browsing patterns reflect and shape mood and mental health. *Nat. Hum. Behav.* 9, 133–146
117. Zhou, D. *et al.* (2024) Architectural styles of curiosity in global Wikipedia mobile app readership. *Sci. Adv.* 10, eadn3268
118. Loosen, A.M. *et al.* (2021) Obsessive–compulsive symptoms and information seeking during the Covid-19 pandemic. *Transl. Psychiatry* 11, 309
119. Russek, E. *et al.* (2022) Time spent thinking in online chess reflects the value of computation. *OSF* Published online October 12, 2022. http://doi.org/10.31234/osf.io/8j9zx
120. Anderson, I.A. and Wood, W. (2023) Social motivations' limited influence on habitual behavior: tests from social media engagement. *Motiv. Sci.* 9, 107
121. Vélez, N. *et al.* (2024) The rise and fall of technological development in virtual communities. *OSF* Published online September 13, 2024. http://doi.org/10.31234/osf.io/tz4dn
122. Sharma, S. *et al.* (2024) Toward human–AI alignment in large-scale multi-player games. *arXiv* Published online June 18, 2024. http://doi.org/10.48550/arXiv.2402.03575