HHS Public Access

Author manuscript

Neurosci Biobehav Rev. Author manuscript; available in PMC 2024 February 01.

Published in final edited form as:

Neurosci Biobehav Rev. 2023 February ; 145: 105008. doi:10.1016/j.neubiorev.2022.105008.

Computational models of subjective feelings in psychiatry

Chang-Hao Kao[1], **Gloria W. Feng**[1], **Jihyun K. Hur**[1], **Huw Jarvis**[1,2,3], **Robb B. Rutledge**[1,4]

[1]Department of Psychology, Yale University, New Haven, Connecticut, USA

[2]Turner Institute for Brain and Mental Health, Monash University, Clayton, Victoria, Australia

[3]School of Psychological Sciences, Monash University, Clayton, Victoria, Australia

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

Abstract

Research in computational psychiatry is dominated by models of behavior. Subjective experience during behavioral tasks is not well understood, even though it should be relevant to understanding the symptoms of psychiatric disorders. Here, we bridge this gap and review recent progress in computational models for subjective feelings. For example, happiness reflects not how well people are doing, but whether they are doing better than expected. This dependence on recent reward prediction errors is intact in major depression, although depressive symptoms lower happiness during tasks. Uncertainty predicts subjective feelings of stress in volatile environments. Social prediction errors influence feelings of self-worth more in individuals with low self-esteem despite a reduced willingness to change beliefs due to social feedback. Measuring affective state during behavioral tasks provides a tool for understanding psychiatric symptoms that can be dissociable from behavior. When smartphone tasks are collected longitudinally, subjective feelings provide a potential means to bridge the gap between lab-based behavioral tasks and real-life behavior, emotion, and psychiatric symptoms.

Keywords

computational psychiatry; subjective feelings; depression; happiness; reward prediction errors; computational model; decision making; smartphone

1. Introduction on psychiatric symptoms and subjective feelings

Research on psychiatric disorders is complicated by the complexity and heterogeneity of psychiatric symptoms. For example, following the current symptom-based diagnostic system, researchers observe substantial heterogeneity among individuals diagnosed with

Corresponding author: Chang-Hao Kao: chang-hao.kao@yale.edu, Robb B. Rutledge: robb.rutledge@yale.edu.

Competing interests

The authors declare no competing interests.

mood disorders (Gillan & Rutledge, 2021; Hitchcock et al., 2022; Huys et al., 2016; Yip et al., 2022). One of the most widely used diagnostic tools, the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) (American Psychiatric Association, 2013), allows for at least 256 symptom phenotypes that can be diagnosed as major depressive disorder (MDD). This heterogeneity is thought to be related to variation in symptom trajectory and treatment responses (Buch & Liston, 2021; Drysdale et al., 2017; Fried & Nesse, 2015).

Researchers in the emerging field of computational psychiatry suggest that psychiatric disorders result from aberrant computations that lead to variation in decision making (Huys et al., 2016; Montague et al., 2012). For example, anxiety is associated with increased risk aversion but not loss aversion (Charpentier et al., 2017) and these distinct effects on risky decision making are captured in the Prospect Theory model (Kahneman & Tversky, 1979). Individuals with high compulsion-related symptoms exhibit lower goal-directed control during learning in a two-step task with fixed probabilities of state transitions (Brown et al., 2020; Gillan et al., 2016; Patzelt et al., 2019). Furthermore, chronic worry is associated with greater perseveration on punishment avoidance goals in a learning environment where the probability of state transitions changes (Sharp et al., 2022). Social tasks can also lead to selective effects, with higher depression associated with lower learning rates only for choices from a virtual partner but not from choices made by participants (Safra et al., 2019).

In contrast to a vast literature on cognitive models evaluated in relation to mental health, there have been few studies using computational models to quantify affective dynamics during tasks. Given that psychiatric disorders often feature aberrant mood dynamics, a better understanding of those disorders may arise from a more precise understanding of how affective dynamics change in well-controlled task environments. We can then ask how affective dynamics during tasks relate to and provide insight into the origin of affective dynamics outside of the lab as well as variation in psychiatric symptoms. Furthermore, do affective dynamics explain unique variance in symptoms beyond what is explained by behavior in the same tasks? Addressing these questions requires adopting paradigms that sequentially sample affective states during decision-making tasks.

In this review, we argue that measuring and modeling momentary subjective feelings during decision-making tasks can help to elucidate the affective processes influenced by psychiatric disorders. In this review, we will discuss findings from studies of both emotion and mood to describe the benefits of measuring subjective feelings to address a variety of questions. One way to distinguish affective states and moods from emotions is that affective states can depend on multiple independent events, and any associated stimuli may no longer be present. In contrast, emotions can be thought of as primarily responses to a specific event. Both emotions and affective states lead to subjective feelings and can be studied with questions of a subjective nature. When studying affective states, these questions should not make reference to any specific events or clearly be about a specific event. This distinction is consistent with recent proposed theoretical framework for emotions and moods (Eldar et al., 2021). For obvious reasons, mood dynamics may be particularly relevant to our growing understanding of mood disorders. Computational models that explain behavior in decision-making tasks provides a useful starting point for understanding affective dynamics, because the same variables that influence behavior should also be relevant to affect. For

example, the prediction errors that quantify the difference between received and expected outcomes in reinforcement learning models can also account for momentary mood dynamics in reinforcement learning tasks (Blain & Rutledge, 2020).

Modeling momentary subjective feelings can improve our understanding of affective processes because computational models can quantify the simultaneous influence of multiple factors on affective dynamics. For example, happiness in a social context can depend on outcomes that happen to another person (Rutledge et al., 2016). The extent to which advantageous and disadvantageous inequality impact happiness predicts social preferences. These types of inequality have been linked to the social emotions of guilt and envy, respectively. Participants may be reluctant to honestly report on how much envy they are currently feeling when asked directly. Computational modeling of affective dynamics allows affective impacts of inequality to be quantified in a way that avoids potentially sensitive questions. Some individuals may misrepresent how they emotionally respond to aspects of a task. Tasks with many emotionally relevant events preceding each affective rating may be particularly well suited for obscuring the subject of study. Further studies can delineate under which circumstances and for which populations tasks with affective state ratings have advantages over simpler tasks. In more naturalistic tasks, computational modelling may be particularly useful for quantifying latent affective and cognitive dynamics.

Mood is not only a product of the computations that underlie decision making but can be dissociable from behavior and also have a different relationship to psychiatric symptoms. During reinforcement learning, learning-irrelevant potential rewards can influence participant choices, but not influence momentary happiness (Blain & Rutledge, 2020). In a social context, social prediction errors that are the difference between expected and observed social feedback can lead to dissociable impacts on feelings of self-worth and the predictions that participants make about future social feedback (Will et al., 2020). Distinct cognitive and affective mechanisms can show impairments in psychiatric disorders. People with depression showed lower general mood during a risk-taking task while there was no difference in the influence of reward prediction errors on their mood ratings (Rutledge et al., 2017). In an ultimatum game, two people split a certain amount of money, with one person proposing how to split and the other person accepting or rejecting the offered amount. Both depressed and non-depressed individuals tended to reject offers when the experienced offers were worse than expected. Non-depressed individuals tended to reject offers after emotional prediction errors (e.g., the experienced emotion was more negative than expected) whereas this influence of emotional prediction errors is reduced in depressed individuals (Heffner et al., 2021).

If affect is to play an adaptive role in behavior, affective state should interact with decision making. This also implies that assessing affective dynamics could capture unexplained variance in behavior in any task where affective state varies. In an ultimatum game, emotional prediction errors predicted participants' rejections of unfair offers (Heffner et al., 2021). A recent proposal also argues that mood represents the momentum of rewards in an environment (Bennett et al., 2022; Blain & Rutledge, 2020; Eldar et al., 2016; Eldar & Niv, 2015). The momentum of rewards reflects the moving average of recent prediction errors, a measure that relates to whether an environment is getting better or worse. In the proposal, a

positive mood could then influence behavior by increasing the perceived value of rewards, thereby increasing the rate of value updates after unexpected rewards and exploiting actual momentum in the environment if it exists (Eldar et al., 2016, 2021). However, this could have unintended consequences and increase risk taking in other domains. After surprising positive events such as sport team winning or a sunny day, increased mood is associated with purchase of lottery tickets (Otto & Eichstaedt, 2018). In mood disorders characterized with high mood instability such as bipolar disorder, mood could distort reward perception in a way that leads to extreme value estimates and behavior (Eldar & Niv, 2015; Mason et al., 2017). Thus, measuring mood dynamics during cognitive tasks could clarify the latent mechanisms that underlie aberrant choice behavior.

Sequential sampling of subjective mood ratings can be conveniently implemented both in the lab and in tasks implemented on mobile devices such as smartphones or tablets. Moving toward large-scale and longitudinal data collection through online platforms and smartphone devices can resolve shortcomings from cross-sectional clinical datasets collected at a single time point in a controlled lab environment (Gillan & Rutledge, 2021; Hitchcock et al., 2022). High accessibility facilitates dense sampling of momentary subjective feelings during behavioral tasks, and it allows researchers to conduct longitudinal studies that probe both cognitive and affective processing in psychiatric disorders with low financial and patient burden (Gillan & Rutledge, 2021; Harari et al., 2016). In addition, smartphones are useful for ecological momentary assessment, and can measure affect in the same participant in different real-life contexts (Killingsworth & Gilbert, 2010; MacKerron & Mourato, 2013).

In this review, we discuss how the dynamics of subjective experience during tasks have been assessed through momentary ratings of subjective feelings and utilized in computational psychiatry research. We also show how computational modeling of affective dynamics during tasks has contributed to a better understanding of emotions in psychiatric disorders. We specifically focus on three widely used decision-making contexts: risky decision making (section 2), reinforcement learning (section 3), and decision making in social contexts (section 4). We suggest potential implications of subjective mood modeling for dissecting heterogeneous symptoms of psychiatric disorders like major depression (section 5). We describe how smartphones can be utilized to improve research in computational psychiatry that bridges between behavioral models and subjective experience as it relates to both real-world emotions and psychiatric symptoms (section 6). Lastly, we suggest guidelines for developing computational models of subjective feelings and propose some future directions for this growing field (section 7).

2. Computational models of subjective feelings: risky decision making

2.1 Risky decision making

Decision making under risk or uncertainty was initially thought to be primarily about maximization of expected values (Bernoulli, 1954; Von Neumann & Morgenstern, 1944). More recently, emotion research has inspired more realistic economic theories including in the field of behavioral economics. Prospect theory is a widely used model for economic decision making under risk, inspired by ideas about anticipated emotions and the role they might play in subjective risk preferences (Kahneman & Tversky, 1979). This theory

formally describes risky decision making and focuses in particular on two phenomena: the diminishing subjective utility of increasing outcome magnitudes that explains risk aversion in gains and risk seeking in losses, and the tendency to weigh potential losses more heavily than potential gains, which is referred to as loss aversion (Kahneman & Tversky, 1979; Sokol-Hessner & Rutledge, 2019). By expressing the concepts that comprise decision making under risk in mathematical form, psychologists have been able to rigorously test hypotheses about how these components may differ between individuals and could relate to brain function or mood (De Martino et al., 2010; Tom et al., 2007). Recently, researchers have increasingly considered a potential relationship between feelings and choice, motivated by theories like the "risk-as-feelings" hypothesis (Loewenstein et al., 2001; Loewenstein & Lerner, 2003) and consistent with research linking risky choice to subjective feelings (Charpentier et al., 2016).

2.2 Mood depends on reward prediction errors

While it seems intuitive that happiness is influenced by reward, and increased wealth should lead to improved mood, empirical evidence suggests that reward alone does not capture the full picture and that the relationship between wealth and happiness is not a simple one (Easterlin et al., 2010; Kahneman et al., 2006; Kahneman & Deaton, 2010). The emotional response to a gamble's outcome depends on the value of the obtained outcome and also its likelihood (Mellers et al., 1997). For example, it feels better to win $50 when the odds are 10% compared to when the odds are 90%. Using computational modeling, researchers have recently formalized a model for happiness incorporating a role for expectations (Rutledge et al., 2014). Happiness is suggested to be a recency-weighted average of chosen certain rewards (CR), the expected values of chosen gambles (EV), and reward prediction errors (RPE), the difference between the received reward and the expected value of chosen gambles (Equation 1) (Rutledge et al., 2015, Rutledge et al., 2014). In addition, a baseline mood parameter (w_0) captures overall mood during the task after accounting for the mood fluctuations that can be attributed to task events. Thus, it may reflect overall how a participant experiences a task and could differ between tasks for an individual. Researchers found that, despite no immediate impact on current wealth, expectations about the future influence happiness, but prediction errors have an even stronger impact (Figure 1A). These results were replicated by an independent group in a pre-registered study (Vanhasbroeck et al., 2021).

$$Happiness_t = w_0 + w_1 \sum_{j=1}^{t} \gamma^{t-j} CR_j + w_2 \sum_{j=1}^{t} \gamma^{t-j} EV_j + w_3 \sum_{j=1}^{t} \gamma^{t-j} RPE_j \quad (1)$$

Using computational modeling allows researchers to quantify the different factors that influence happiness and help to bridge the gap between subjective emotional experience and the neurophysiology of affective processing during risky decision making. Using functional magnetic resonance imaging (fMRI), blood oxygen level dependent (BOLD) activity in the ventral striatum preceding happiness ratings was found to correlate with later self-reported happiness ratings (Figure 1B) (Rutledge et al., 2014). Furthermore, activity in the ventral striatum also correlated with the magnitude of certain rewards, expected values, and reward

prediction errors that all influence momentary happiness. Neurons that release dopamine show activity patterns that resemble these reward prediction errors (Schultz et al., 1997) and this is consistent with BOLD responses in the ventral striatum thought to be due to dopaminergic input (Caplin et al., 2010). Moreover, right anterior insula activity at the time when participants were asked to rate their current happiness was positively correlated with happiness ratings (Rutledge et al., 2014), consistent with evidence that this area supports interoceptive awareness (Critchley et al., 2004; Damasio, 1999).

Examining the relationship between happiness and risk taking can inform our understanding of how mood disorders influence affective experiences and behavior in a context that is well understood from a psychological and neurobiological perspective. While depressive symptom severity negatively correlates with overall happiness during risk-taking tasks, the neural and emotional impact of reward prediction errors is intact in major depression (Rutledge et al., 2017). This result suggests that the aberrant processing of reward prediction errors during reinforcement learning tasks in previous studies may reflect more downstream impairments in behavior or cognitive appraisal (Kumar et al., 2018). Clinical anxiety is linked to increased risk aversion (Maner et al., 2007), but not loss aversion (Charpentier et al., 2017), although it is less clear whether risk taking is influenced by depression (Chung et al., 2017). Affective experience has only been evaluated in a small number of risk-taking paradigms but modeling the dynamics of subjective feelings like happiness first in these well-understood decision paradigms will be a key step in understanding how these dynamics become dysfunctional in psychopathology.

2.3 Mood is influenced by counterfactual outcome

Mood can also depend on the unobtained outcome of unchosen options. Several studies have shown how emotions can relate to comparisons between the outcome of the chosen option and the unobtained outcome in the unchosen option (Bennett et al., 2022; Coricelli et al., 2005; Mellers et al., 1997). For example, participants make a choice between option A and option B. Option A can win $50 or lose $50 and option B can win $200 or lose $200. After choosing option A, participants feel better after a $50 win than a $50 loss. Seeing the unobtained outcome from the unchosen option B influences emotions. They feel better if the unobtained outcome is losing $200 than if the unobtained outcome is winning $200. Considering the unobtained outcome introduces counterfactual thinking about what they would have obtained if they made a different choice, and this counterfactual outcome can influence emotions (Coricelli et al., 2007; Coricelli & Rustichini, 2010). Furthermore, participants report disappointment when the obtained outcome is worse than they expected, and regret when the obtained outcome from the chosen option is worse than the unobtained outcome from the unchosen option. Greater regret was associated with elevated activity in orbitofrontal cortex, anterior cingulate cortex, and hippocampus (Coricelli et al., 2005). Regret for counterfactual outcomes has also been shown to be reduced in depressed patients (Chase et al., 2010). Obsessive-compulsive disorder patients showed more extreme emotional responses to counterfactual outcome than healthy controls but no difference for outcomes from chosen options (Gillan et al., 2014). These findings suggest that subjective feelings for counterfactual outcomes may be relevant to a deeper understanding of psychiatric disorders.

2.4 Interactions between mood and risky decision making

Mood could also influence subsequent risky decisions. The “mood maintenance theory” proposed that people in happy moods are actually more reluctant to take risks because they want to avoid undermining their positive emotional state (Isen et al., 1988). This tendency to overweigh the pain of potential losses relative to gains is consistent with heightened loss aversion, attributed to a raised reference point (Mellers et al., 2021). However, some researchers have found the opposite relationship. Elevated mood has been related to increased risk seeking (Forgas, 1995; Stanton et al., 2014). This effect is consistent with analyses of real-world urban populations, which showed that positive incidental outcomes like local sporting events and weather patterns predict greater participation in lottery gambling (Otto et al., 2016). Analyses of day-to-day mood language extracted from Twitter and localized to the same location established that such surprising positive outcomes increase mood, and this increased mood is associated with increased gambling (Otto & Eichstaedt, 2018). Consistent with the theory that mood represents momentum of reward (see section 1 for more information of the theory), a positive mood reflects increasing overall reward availability, and risky options may represent a novel reward source to be approached. Consistent with this possibility, people choose more novel stimuli in a positive mood (Dreisbach & Goschke, 2004).

Affective experience is also amenable to intentional cognitive regulation. Notably, researchers have found that cognitive regulation strategies such as “perspective-taking” reduce physiological arousal to losses, and this has the effect of reducing loss aversion (Sokol-Hessner et al., 2009). These findings demonstrate that emotions influence subjective valuations of risk and behavior just as the positive and negative outcomes of those risks have an influence on emotions. Moreover, subjective feeling associated with potential risky options can be used to predict risky taking better than using the established Prospect Theory model (Charpentier et al., 2016).

3. Computational models of subjective feelings: learning and uncertain environments

3.1 Learning in uncertain environments

In an uncertain environment, values of options are often not observable. Under the framework of reinforcement learning, people can learn these values by trial and error, and make adaptive decisions that maximize cumulative expected reward (Sutton & Barto, 2018). When the received outcome is better than predicted (i.e., a positive prediction error), the value of the chosen option should be increased, leading to an increase in the probability of repeating the same behavior. When the received outcome is lower than expected (i.e., a negative prediction error), the value of the chosen option should be decreased.

3.2 Mood depends on prediction errors during learning

Subjective feelings should play a role in adaptive behavior, but it remains unclear what this role is. Many studies show that physiological arousal changes during learning in response to prediction errors, surprise, and uncertainty. For example, pupil diameter increases as belief

surprise or belief uncertainty increases (Nassar et al., 2012). A recent study measured skin conductance, pupil diameter, and subjective ratings of stress during learning (de Berker et al., 2016). Participants predicted whether pictures of rocks would lead to a snake, which resulted in a mild electrical shock when it was present. The probability of the snake for these stimuli changed occasionally during the task. This study explores three types of uncertainty: irreducible uncertainty, estimation uncertainty, and volatility uncertainty. Irreducible uncertainty emerges from the probabilistic association between action and outcome. It is highest when the probability of shock is 50% and gradually decreases as the probability of shock goes to 0% or 100%. Estimation uncertainty reflects imprecision in estimated shock probability and decreases with learning. Volatility uncertainty captures imprecision in the estimated volatility, which reflects instability in the shock probability. As irreducible uncertainty estimated from participant predictions increased, subjective stress, skin conductance, and pupil diameter also increased. The influence of irreducible uncertainty on subjective stress was associated with the influence of irreducible uncertainty on both skin conductance and pupil diameter, supporting a link between subjective stress and physiological arousal in uncertain environments. Stress has also been shown to affect decision processes including valuation, learning, and risk taking in the lab and real life (Morgado et al., 2015; Porcelli & Delgado, 2017).

In addition to risky decision making (section 2.2), the influence of prediction errors on momentary happiness has also been shown during learning in uncertain environments (Blain & Rutledge, 2020; Eldar & Niv, 2015). In these studies, participants made a choice from multiple options with different reward probabilities. Trial-by-trial expected probabilities and prediction errors were estimated from a reinforcement learning model that best explained choice. During learning, participants were asked periodically about their momentary happiness. Their mood dynamics were driven both by expected probabilities and experienced prediction errors (Blain & Rutledge, 2020). These findings are consistent with a recent theory that mood represents the momentum of reward (see section 1 for more information of the theory).

Building on this work, Bennett and colleagues proposed that mood depends on the integration of advantages from multiple sources (Bennett et al., 2022). Advantage captures the difference between the value of taking a specific action in a specific state and the value of that state. Thus, a positive advantage indicates that this action can increase expected reward. People can adjust behavior based on this advantage to maximize expected future reward. Recent studies show that this advantage can influence mood (Equation 2) (Bennett et al., 2022).

$$Mood_{t+1} = Mood_t + \eta_{mood}(Advantage_t - Mood_t) \quad (2)$$

This model considers multiple sources for advantage. The first advantage is the reward prediction error of the chosen action. The second advantage is the difference between the learned value of the chosen action and the learned value of the unchosen actions. The third advantage is the difference between the learned value of the chosen action and the actual outcomes of the unchosen actions when participants receive this counterfactual information.

In this model, mood can be simultaneously influenced by these three advantages with different weights that capture the influence of multiple factors (e.g., expectation, prediction errors, counterfactual outcomes) on mood identified in past studies.

Mood dynamics during learning may not reflect the momentum of all recent outcomes but instead specifically the prediction errors that are relevant to learning. In a recent study (Blain & Rutledge, 2020), participants chose between two options with different reward probabilities, and each option was randomly assigned a potential reward. Participants must integrate their current beliefs about the probability of reward with the potential rewards on each trial to make decisions. In this task, there are two sources of prediction errors. One is the probability prediction error, which indicates the difference between whether participants receive a reward and the expected probability of the chosen option. This probability prediction error can be used to update beliefs about the reward probability of the chosen option. The reward prediction error is the difference between the magnitude of a received reward and the expected value of the chosen option. This reward prediction error is not informative in learning because the potential rewards are randomly assigned on each trial. Participants were asked to rate their happiness periodically during the task. The model that included probability prediction errors performed better than the model that instead included reward prediction errors in both stable and volatile environments (Figure 2A). Reward magnitude influenced participant choices but did not influence momentary happiness. This was consistent for the stable environment (where the reward probabilities of the two options was stable for the entire task) and the volatile environment (where the reward probabilities of the two options switched periodically during the task).

Modeling subjective feelings during learning could help to understand the aberrant beliefs and decisions present in psychiatric disorders. For example, using computational models to evaluate the baseline mood parameters in stable and volatile environments separately (Blain & Rutledge, 2020) showed differences related to depression. Depressive symptoms were associated with lower baseline mood parameters in volatile but not stable environments (Figure 2B). In volatile environments, anxiety symptoms are associated with irregular learning (Browning et al., 2015), and subjective feelings measured in different types of uncertain environments could help in understanding the experience of people with psychiatric disorders.

3.3 Interactions between mood and learning

Mood is not just a byproduct during learning but can also influence learning. Manipulating mood with task-irrelevant stimuli can influence later preferences (Eldar & Niv, 2015; Michely et al., 2020). In a learning task (Eldar & Niv, 2015), participants learned to choose among three slot machines with reward probabilities of 20%, 40%, or 60%. In the middle of the task, people played a task-irrelevant wheel of fortune. Participants were happier after winning than losing the wheel of fortune. After this wheel of fortune, participants learned to choose among three new slot machines with reward probabilities of 20%, 40%, and 60%. In the test phase, participants chose between pairs of slot machines they had learned about but in the absence of any additional feedback. For people with high trait mood instability, indicating vulnerability for bipolar disorder, preferences were influenced

by the wheel of fortune outcome. For the slot machines with the same reward probabilities, they preferred the one learned after winning the wheel of fortune to the matched ones learned before winning the wheel of fortune (Eldar & Niv, 2015). Conversely, they preferred the slot machines learned before losing the wheel of fortune to matched ones learned after losing the wheel of fortune. Even though the wheel of fortune was not relevant to learning, mood changed as a result might have influenced perceived outcomes and biased subsequent preferences. Additionally, this mood impact also modulated the neural encoding of reward in striatum during learning. People with high trait mood instability showed stronger neural responses to reward after winning compared to before winning the wheel of fortune. Conversely, they showed weaker neural responses to reward after losing compared to before losing the wheel of fortune.

Computational models help in understanding the association between mood dynamics and learning dynamics. A mood bias parameter in this model influenced the perception of a received reward (Eldar & Niv, 2015). As learning is driven by prediction errors between expectations and rewards, this biased perception on received reward can influence learning. The mood bias parameter was associated with trait mood instability. Moreover, this theoretical framework suggested that a high mood bias parameter is a risk factor of bipolar disorder (Mason et al., 2017). For example, a positive prediction error leads to higher mood. The higher mood biases the perception of received outcome to generate larger positive prediction errors, and then this large positive prediction error updates expectation upward more than they would have otherwise. As expectation becomes excessively high, individuals could enter a manic phase where they expect everything to go well and experience any small reward as being large. However, this state would increase the probability of large negative prediction errors and eventually could contribute to a depressive phase. A number of large negative prediction errors together lead to low mood, and this low mood biases the perception of received outcomes downwards and thereby expectations. Stronger mood bias parameters could lead to stronger positive feedback dynamics that encourage manic and depressive phases.

4. Computational models of subjective feelings: social environments

4.1 Decision making in social contexts

Decision making in social contexts is complex because people care not only about their own choices and outcomes but also the choices and outcomes of others. In real life, we make many decisions involving interaction with others, from negotiating salary for a new job, to responding to a tweet, to asking someone out on a date. Such decisions recruit a broad range of cognitive processes, including mental state inference and the evaluation of social norms (Lee, 2008; Lee & Harris, 2013; Rilling & Sanfey, 2011; Xiang et al., 2013). Computational modeling is increasingly being used to understand how these processes contribute to social decision making (Cheong et al., 2017; Cushman & Gershman, 2019; FeldmanHall & Nassar, 2021). Emotional expressions and how they change over time could reflect recent events like self-reports of mood. Individuals can use emotional expressions to infer the likely causes (Ong et al., 2019; Wu et al., 2021). Computational models can test to what extent multiple past events influence current expressions, allowing tests of whether others can infer the

causes of emotional expressions in a way that matches the factors that are most predictive of expressions.

A prominent finding is that social decisions often deviate from normative theories of reward maximization (Bernoulli, 1954; Kahneman & Tversky, 1979; Von Neumann & Morgenstern, 1944). This has been shown empirically using the ultimatum game, a two-player economic choice paradigm in which a proposer decides how to split money with another player and a responder decides whether to accept or reject the offer (Güth et al., 1982; Harsanyi, 1961; van ‘t Wout et al., 2006). Responders reject around half of all offers that fall below 20% of the total, even though the rational (i.e., reward-maximizing) choice strategy in non-repeated interactions is to accept any non-zero offer (Nowak et al., 2000). Rejected offers result in no money for either player, and are believed to reflect negative emotions (i.e., anger) that relate to a desire to punish the proposer (Nelissen & Zeelenberg, 2009; Pillutla & Murnighan, 1996). Neuroimaging using fMRI suggests a link between heightened activity in anterior insula, dorsolateral prefrontal cortex, and anterior cingulate cortex and increased rejection of unfair offers, and this signal, as well as rejection rates, are increased during interaction with a human compared to a computer player (Sanfey et al., 2003).

Computational modeling has led to progress understanding disorders with a social dimension, such as autism, social anxiety, and borderline personality disorder, which have been linked to differences in learning and decision making in interpersonal settings (Fineberg et al., 2018; Forgeot d’Arc et al., 2020; Henco et al., 2020; Hopkins et al., 2021; King-Casas et al., 2008; Siegel et al., 2020). An ongoing challenge is to understand the mechanisms by which affective experience during social decision making is related to psychiatric disorders, something that computational models of subjective feelings have begun to shed light on.

4.2 Mood is influenced by social comparison

Mood dynamics are influenced by comparison between outcomes for the self and for other people, consistent with ongoing affective experience being influenced by other reference points than just expectations. This is also consistent with measures of subjective well-being at a population level, where how satisfied people are with their lives depends partially on how they compare to others in their social environment (Boyce et al., 2010; Luttmer, 2005). In an ultimatum game, people exhibited stronger arousal-related skin conductance responses when rejecting versus accepting an offer proposed by a human partner (van ‘t Wout et al., 2006). In contrast, there was no difference in skin conductance between rejecting and accepting an offer proposed by a computer. Momentary happiness in individuals also reflects social comparison (Rutledge et al., 2016). Participants rated their momentary happiness as they played a risky decision-making task in which they saw the outcomes not only of their own choices, but also those of a social partner. Happiness was predicted best by a model that accounted for subjective feelings elicited by social comparison, revealing that both advantageous inequality (i.e., conditions that could elicit guilt) and disadvantageous inequality (i.e., conditions that could elicit envy) reduced happiness (Figure 3A). Model parameters also predicted how generous participants were in a separate dictator game: greater guilt was associated with more generous decisions, and greater envy with less

generous decisions (Figure 3B). This pattern of results highlights how computational models can be used to distinguish many simultaneous influences on happiness, including some that may be socially undesirable to admit (e.g., envy).

Previous work has shown that generosity itself is associated with greater happiness (Dunn et al., 2008). In one study, participants were instructed that they would receive a monetary endowment to spend over a period of four weeks, allocating one group to pledge to spend the money on other people (experiment group) and a separate group to pledge to spend the money on themselves (control group) (Park et al., 2017). Participants made a series of choices to accept or reject proposals from a social partner while undergoing an fMRI scan. Each proposal consisted of a monetary benefit for their social partner and a monetary cost to themselves. The researchers found that participants who had pledged to spend money on others made more generous decisions, and reported a greater increase in happiness over the course of the experiment (Park et al., 2017). Generous decision making corresponded to increased BOLD activity in the temporoparietal junction, and increased connectivity between that area and ventral striatum. Remarkably, these effects were seen after participants had made a commitment to being generous, but before they had spent any money.

4.3 Mood is modulated by social norms

Mood dynamics can also be influenced by social norms. One approach has been to model social emotions as affective and motivational state changes in response to violations of social norms. This builds on functional theories that compare emotions to homeostatic mechanisms (Chang & Smith, 2015; Damasio, 1999; Seth, 2013), positioning social norms such as “fairness” as learned set points (Montague & Lohrenz, 2007). In a variation of the ultimatum game, participants acted as responders to offers from a computer (Xiang et al., 2013). Offers from the computer were drawn from a Gaussian distribution, the mean and variance of which differed between the first and second half of the experiment. Throughout the task, participants rated their subjective feelings about the current offer using emoticons (Lang, 1980). Using computational models of subjective ratings, the researchers found that momentary happiness depended not only on the fairness of the current offer, but on how much that offer deviated from the fairness norm established in the first half of the experiment (Xiang et al., 2013). The extent to which each offer deviated from this model-based norm covaried with BOLD activity in medial prefrontal cortex, nucleus accumbens, and posterior cingulate cortex.

Prediction errors derived from social feedback can also influence momentary feels of self-worth (Will et al., 2017, 2020). Participants received a series of “likes” and “dislikes” that they were told were from people who had viewed a social media profile they had previously submitted to researchers. On each trial, participants were presented with the name of the rater, and a color cue indicating which of four groups the rater belonged to based on how likely they were to like profiles in general. The researchers found that self-esteem was not only sensitive to social approval or disapproval, but to social approval prediction errors: receiving a like resulted in a bigger increase in self-esteem if it came from a rater who liked few profiles in general, compared to a rater who liked most profiles. Social approval

prediction errors modulated BOLD activity in ventral striatum, but changes in self-esteem modulated ventromedial prefrontal cortex activity.

These findings are consistent with the notion that self-esteem is shaped over time by social evaluation by others (Gruenenfelder-Steiger et al., 2016), suggesting that self-esteem may serve as a type of dynamic learning signal used to update beliefs about changes in one's own social standing (Low et al., 2022). Low self-esteem is an important risk factor across a range of psychiatric disorders (Orth et al., 2012), which raises the possibility that such disorders are driven in part by aberrant cognitive or affective processes during social learning. Indeed, participants with low trait self-esteem exhibited impaired social learning and tended to persist in their expectations of disapproval (Will et al., 2020). At the same time, momentary feelings of self-worth in this group were more volatile and susceptible to change based on social prediction errors, identifying a dissociation between the impact of social prediction errors on learning and feelings that relate to trait self-esteem (Will et al., 2020). Participants reported feelings of self-worth throughout the experiment, and computational modeling was used to explain fluctuations in self-worth in relation to recent task events. The researchers then used canonical correlation analysis to derive a computational phenotype based on both psychiatric symptoms and parameters from the computational model. Participants with high scores on a single dimension of "interpersonal vulnerability" not only had lower trait self-esteem, but also exhibited attenuated BOLD signal in ventromedial prefrontal cortex, which corresponded to lower expectations of positive social approval during the task (Figure 4) (Will et al., 2020).

4.4 Interactions between mood and social decision making

Social behavior varies widely across individuals, but subjective feelings do not always track decision making. In a social context, subjective feelings of self-worth can be highly reactive to social feedback in individuals with low self-esteem who do not update their expectations about future social feedback, despite both expectations and subjective feelings of self-worth being subject to the same social prediction errors (Will et al., 2020). Another study investigated the role of reward prediction errors and also emotion prediction errors in a non-learning social context (Heffner et al., 2021). In an ultimatum game, participants rated their momentary affect along valence and arousal dimensions twice on each trial: once before the offer was made, capturing emotion expectations, and once after the offer was made, capturing emotion experience. Reward prediction errors between the observed offer and expected offer were predictive of rejection. Differences between experienced and expected emotion were computed as valence prediction errors and arousal prediction errors, whose role in decisions was distinct from reward prediction errors. Participants were more likely to reject offers (and thus punish their partner) after experiencing less valence or more arousal than expected. Critically, depressed participants showed diminished use of emotion prediction errors in guiding decisions, but intact use of reward prediction errors (Figure 5). Moreover, depression was associated with a reduced overall range of reported emotional experience (Heffner et al., 2021). Together, these results suggest that emotional responses to social feedback can be blunted in depression even if responses to some types of reward feedback are not (Rutledge et al., 2017).

5. How computational models of subjective feelings could help us understand mood disorders

Despite the fact that mood disorders are diagnosed based on self-reported subjective symptoms, there has been little research on using computational models to understand subjective feelings in controlled task conditions. Experience sampling provides information about emotional variability but does so in an uncontrolled environment with minimal information as to ongoing experience. Measuring affective states with questionnaires depends on how well individuals remember past emotions. Measuring momentary mood changes during different tasks provides a way to measure affective experience in controlled task conditions that is complementary to standard approaches to measuring emotion.

Psychiatric disorders such as major depressive disorder or generalized anxiety disorder have shown a wide range of neurocomputational deficits in behavior (Hitchcock et al., 2022). However, there is considerable heterogeneity and the underlying mechanisms are not well understood. For example, there was mixed evidence about model parameters in reinforcement learning tasks in depression (Chen et al., 2015). Compared with control participants, a meta-analysis showed that depressed patients showed higher learning rates for punishments and slightly lower learning rates for reward (Pike & Robinson, 2022). However, there was considerable variability across studies. Investigating subjective feelings during tasks may help to resolve the inconsistency of these findings. For example, different effects of learning rates for punishment and reward may only be shown on participants who change their mood in response to punishment and reward, but not on the participants who showed no change in affective state.

The link between cognitive and affective mechanisms can be influenced by psychiatric disorders. Psychiatric disorders may lead to different impacts on behavior and subjective feelings. For example, depressive symptoms did not lead to impairments in performance in risk-taking tasks (Rutledge et al., 2017) (see section 2.2 for more information of the study) or reinforcement learning tasks (Blain & Rutledge, 2020) (see section 3.2 for more information of the study). However, higher depressive symptoms were associated with lower baseline mood parameters, suggesting that depression influences the affective experience of individuals completing these tasks. In a learning task, mood instability assessed with a standard clinical questionnaire was associated with a mood bias parameter quantifying the impact of mood on learning (Eldar & Niv, 2015) (see section 3.3 for more information of the study). These findings suggest that measuring affective processes can reveal distinct cognitive and affective mechanisms in psychiatric disorders. Furthermore, modeling subjective feelings has the potential to disentangle the stable and dynamic components of affective processes. For example, bipolar disorder and borderline personality disorder are both characterized by high mood variability. Modeling daily mood ratings collected for a long time period revealed that mood changes in bipolar disorder persist (i.e., mood volatility) longer than mood changes in borderline personality disorder (i.e., mood noise) (Pulcu et al., 2022). Baseline mood parameters could capture relatively more stable components, although drift in mood can also be modeled, providing useful additional

information. For example, a higher decay on mood over time during rest is associated with lower depression risk (Jangraw et al., 2021).

Having these affective measures aid in the identification of subtypes for different disorders. Past studies have used self-report scores to cluster depressed patients along anxiety and anhedonia dimensions, and revealed putative neural subtypes based on brain functional network connectivity (Drysdale et al., 2017). We can apply dimensional approaches to both cognitive and affective measures. These components can be linked to specific neural subtypes or aid in the identification of new subtypes that reflect brain, behavior, and emotion. In addition, subjective feelings can be investigated in relation to brain network measurements. Greater daily variability in physical location was related to increased positive affect and this link was stronger for people who show greater functional connectivity between ventral striatum and hippocampus (Heller et al., 2020).

More dimensional data are required to understand the heterogeneity of mood disorders. Bigger data sets are useful for increasing statistical power, but only if they measure the right things. Especially because subjective feelings should relate to subjective symptoms, adding ratings of subjective feelings to existing tasks is an intuitive and efficient way to collect additional data that should be highly relevant to psychiatric disorders. Given such multi-dimensional data from individuals, we could make better symptom predictions (Rutledge et al., 2019). Furthermore, using parameters estimated from computational models for behavior and subjective feeling can increase power over machine-learning approaches that do not employ computational models to reduce the dimensionality of the data (Rutledge et al., 2019). Computational models can help to better understand and categorize individuals and be useful in designing effective interventions for specific individuals (Nair et al., 2020). For example, the adaptation of learning rate to the volatility of rewards is intact for greater anxiety symptoms but the adaptation of learning rate to punishments was impaired for greater anxiety symptoms (Pulcu & Browning, 2017). Compulsivity is associated with impaired model-based learning (Gillan et al., 2016). In a volatile environment, compulsivity was associated with impaired learning (Sharp et al., 2021; Vaghi et al., 2017) but confidence ratings in response to volatility were unchanged (Vaghi et al., 2017). In addition, adaptation of learning rate to reward or punishment volatility was not associated with depressive symptoms (Blain & Rutledge, 2020; Pulcu & Browning, 2017), but greater depressive symptoms were associated with lower baseline mood parameters reflecting a different affective experience (Blain & Rutledge, 2020). Understand this computational heterogeneity can inspire different intervention for different individuals (Pulcu & Browning, 2017).

6. The value of smartphones for computational models of subjective feelings

Smartphone-based research methods have the potential to dramatically advance our scientific understanding of subjective feelings and mental health (Gillan & Rutledge, 2021). Research in mental health today has focused on making descriptive claims about mental illness and its contributing factors in the population. For the field to provide insights that are clinically useful, a major paradigm shift is needed that can move the field beyond

description and toward prediction (Browning et al., 2020). The highly individualized nature of subjective experiences makes it a good candidate for being relevant to this problem of predicting treatment responses and symptom severity.

Subjective well-being is complex, and it is influenced by a large array of competing and interacting factors such as sleep, stress, early life trauma, social factors, and diet. It is challenging to make advances in understanding mental illnesses because large comprehensive studies that can capture different dimensions of mental illness and lifestyle are difficult to conduct in traditional lab settings. Due to the wide availability of smartphones, which offer capabilities like sleep monitoring, geolocation tracking, accelerometer data, and social media activity logs, smartphones are uniquely positioned to deliver substantially larger and richer multivariate datasets than feasible through traditional single-site studies. As people experience a wide range of emotions in daily life, smartphone can also provide a convenient tool to measure the richness of real-life subjective feelings (Trampe et al., 2015). Smartphone-based experiential sampling can be used in innovative combinations with neuroimaging methods to reveal the links between lifestyle, brain connectivity, and mental health outcomes. Greater diversity of daily-life activities predicts positive affect in humans and increased hippocampalstriatal functional activity (Heller et al., 2020). Daily-life activities can be used to estimate an individual-specific mobility “footprint” and the more consistent and distinctive the footprint, the lower the mood instability. This footprint was also predictive of sleep irregularity and functional brain network connectivity (Xia et al., 2022).

Self-reported symptom inventory scores capture a static snapshot in time that can be related to task data. However, when administered only once, these scores fail to encompass the reality that mood disorders follow dynamic trajectories over time, and that the instruments themselves might not be accurately capturing the latent traits that are most important for predicting future outcomes (Sharp et al., 2020). Formal computational models can be used to relate internal traits to symptom change over time. In addition to the dynamics of mood in tasks, the dynamics of traits or symptoms over time can be important features of psychiatric disorders. Furthermore, understanding the association between mood and symptom dynamics can help to predict future symptoms changes. Smartphone-based methodologies are especially useful because they lower practical barriers to acquire densely sampled datasets, and have advantages over in-lab data collection in allowing ecological experience sampling during daily life on the same platform. Several studies have shown how smartphones can conveniently collect subjective feelings over a long period (e.g., hours or days). In one study, students reported their positive and negative affect periodically over several hours each day on several days, and showed that real-life prediction errors resulting from exam results influenced mood for multiple days (Villano et al., 2020). Another study used smartphone data collection to show that electroencephalographic measurements of neural responses to reward prediction errors during learning tasks predicted mood changes up to 24 hours later (Eldar et al., 2018). These studies illustrate the value of employing longitudinal techniques on smartphones to understand real-world mood and behavior.

7. Future directions

Adding subjective ratings to existing tasks provides additional data that can be used to understand psychiatric disorders. Different subjective feelings can be measured depending on the disorders or processes under study. Different questions can probe specific types of emotion (Heffner et al., 2021; Heffner & FeldmanHall, 2022), which have different relationships with decision making (Heffner & FeldmanHall, 2022). For example, researchers interested in the conditions that make people angry might consider repeatedly asking questions about anger. Researchers interested in social behavior could ask questions about self-esteem. Different tasks should modulate different kinds of subjective feelings. For example, subjective stress during learning was associated with uncertainty (de Berker et al., 2016) (see section 3.2 for more information of the study). Subjective feelings of self-worth were related to trait self-esteem (Will et al., 2017, 2020) (see section 4.3 for more information of the study).. Valence prediction errors were more related to rejecting offers from other people compared with arousal prediction errors (Heffner et al., 2021) (see section 4.4 for more information of the study). Future studies can also investigate decision making in more complicated situations where people may form a cognitive map of the environment (Behrens et al., 2018). For example, how does momentary happiness changes in response to map complexity and deviations from the learned cognitive map?

To model subjective feelings, a good setting for measuring subjective feelings is required. While emotion researchers have often focused on affective responses to specific events, less is understood about affective states like moods which often change more slowly (see section 1 for the distinction between emotion and mood). For any task in which affective state might vary in relation to multiple previous events, we recommend the following guidelines for developing computational models of affective states:

1. Questions that probe affective state can be included in a wide variety of tasks, but the task should be such that affective state varies over time for most participants. For example, probabilistic reward is a reliable way to influence happiness. Even if the major focus of the study is not reward, probabilistic reward can be a way to keep participants engaged and to provide a reference point to compare affect in other domains. For example, the happiness related to a task that only some participants found intrinsically rewarding can be related to the happiness derived from a probabilistic reward task that most people found extrinsically rewarding (Chew et al., 2021).

2. Affective state questions should be related to participant affective state and not mention task events. For example, for happiness, participants can rate between “very unhappy” and “very happy” for the question “How happy are you right now?” (Blain & Rutledge, 2020; Rutledge et al., 2014). For self-esteem, participants can rate between “very bad” and “very good” for the question “How good do you feel about yourself at this moment?”(Will et al., 2017, 2020). A continuous scale without numbers or markings reduces the probability that participants remember previous ratings.

3. Affective state questions should not be asked too frequently. For most paradigms, no more than twice per minute is a good rule of thumb. For a trial-based paradigm, there should always be at least two trials between each rating. Computational modeling can be used to separate out the influences of multiple previous events. Questions asked more frequently risk annoying participants.

4. Repeatedly answering affective state questions is an additional task that participants perform. When asked only a single question repeatedly, participants typically respond quickly and without substantial reflection. Asking multiple types of questions can introduce additional task switching costs that reduce participant engagement and data quality. Thus we recommend sticking to one question per experiment. Rating along valence and arousal dimensions simultaneously does offer one way to get multiple measures without overcomplicating the task (Heffner et al., 2021).

5. Task-relevant information should not be presented on the screen when participants answer affective state questions. Any task cues or information about overall performance (e.g., total score), could lead to an additional impact on affective state different from the subject of study.

These guidelines may be useful for investigating the roles of emotion and mood in decision making. A recent theoretic framework mapped different types of emotion to different computations during decision making (Emanuel & Eldar, 2022). Under this proposed framework, decision making is decomposed into multiple processes: outcome evaluation, value learning, policy learning, and planning. Pleasure and pain relate to outcome evaluation; happiness and sadness relate to value learning; frustration and content are related to outcomes due to our actions; anger and gratitude relate to outcomes due to others' actions; desire and hope relate to plans to realize uncertain outcomes; fear and anxiety relate to avoiding uncertain outcomes. Future studies following our guidelines can test specific predictions of this framework by measuring feelings associated with specific emotions in environments where behavior can be explained by the emotion-relevant computations.

One important question is whether the subjective feelings we measure reflect the latent state we wish to study. First, we can evaluate the model fit on momentary mood. If model performance and reliability are high when considering data collected at different times, this suggests that task events can be related to self-reports in a consistent way. Second, we can evaluate how ratings respond to specific task events. For example, people should be happier after wins compared to losses. Third, we can evaluate baseline mood parameters across tasks. If correlations are high, that suggests this measure is coherent across tasks. Fourth, we can directly manipulate mood with standard manipulations. For example, people felt happier for multiple ratings after winning a wheel of fortune (Eldar & Niv, 2015). Fifth, we can evaluate whether self-report ratings are associated with subsequent behavior. For example, low mood in the current situation (e.g., mood in the current job) can be associated with change from the current situation (e.g., switching jobs) (Kaiser & Oswald, 2022). Participants in a high mood act as if rewards are perceived as better than they are, choosing rewarded options more frequently (Eldar & Niv, 2015). Last, we can collect questionnaires

specifically about emotion awareness. High emotion awareness is associated with insula activity (Sharp et al., 2018), and insula activity is associated with self-report happiness (Rutledge et al., 2014). Here, there is overlap with the issues concerning other subjective reports including confidence (Vaghi et al., 2017) and metacognitive awareness (Fleming & Lau, 2014). A detailed understanding of affective states will benefit from use of the tools developed to study other kinds of subjective self-reports.

Researchers should characterize affective processing across a wide variety of tasks in relation to psychiatric disorders. Research Domain Criteria (RDoC) provides a research framework to study psychiatric disorders (Cuthbert & Insel, 2013; Insel et al., 2010). For example, Positive Valence Systems delineate several constructs relevant for decision making: reward responsiveness (including reward anticipation, initial response to reward, reward satiation), reward learning (including probabilistic and reinforcement learning, reward prediction error, habit), and reward valuation (including reward, probability, delay, effort). These constructs are shared between multiple decision-making tasks. Past studies based on this framework have focused on behaviors and not subjective feelings. In the same way that behavioral processes should be shared across tasks, affective processes (e.g., affective responses to reward prediction error) should be shared between tasks (e.g., between risk-taking and reinforcement learning tasks). Studies of subjective feelings also provide insight into neurobiological processes that contribute to subjective symptoms but are difficult to evaluate in animal models where selfreports of affective states are unavailable. A focus in pharmaceutical research on animal models means that drug development has focused primarily on behavioral differences and largely ignored subjective aspects of psychiatric disorders, despite subjective aspects being a major source of patient distress (LeDoux & Pine, 2016). The neural circuits that generate aberrant behaviors should overlap with but not be identical to those that generate aberrant feelings. Any difference in symptom-behavior and symptom-feeling associations are an indication of how much the processes are dissociable. A goal of psychiatry is to treat subjective symptoms, and thus adding measurements of subjective feelings to established tasks can enrich the RDoC framework and improve our understanding of psychiatric disorders and design of effective treatments.

With the collection of larger datasets, we can enrich transdiagnostic dimensional approaches to understand symptoms instead of focusing on specific disorders. For example, past studies used factor analysis on multiple self-report questionnaires to extract three factors (e.g., compulsive behavior and intrusive thought, anxious depression, and social withdrawal), and then evaluated the association between these factors and task performance (Gillan et al., 2016; Gillan & Seow, 2020). Given anxiety as an example, past studies discussed cognitive and neural difference between two anxiety subtypes: anxious apprehension and anxious arousal (Sharp et al., 2015). Anxious apprehension is related to worry while anxious arousal is related to fear and panic. This distinction can be computationally linked to different decision-making processes. High chronic worry was associated with difficulty in disengaging from the goal of punishment avoidance when the current goal has changed to seek reward (Sharp et al., 2022). In an aversive environment, high physiological symptoms of anxiety were associated with enhanced learning from safety (Wise & Dolan, 2020). These approaches can help to understand individual processes and symptoms. As we

collect affective measures in addition to cognitive measures, we can expand transdiagnostic dimensions and better stratify patients, aiding in the design of effective treatment to target different individuals. Additionally, with increasing data from cognitive and affective perspectives, we will need tools for integrating diverse information to account for current symptoms and to predict future symptoms. Theory-driven approaches allow task data to be compactly summarized with a small number of parameters. Data-driven machine learning approaches are complimentary in allowing parameters estimated from multiple tasks to be combined with other data sources to make predictions (Rutledge et al., 2019). In large data sets, machine learning approaches can also account for behavioral and affective variability in complicated tasks in ways that lie outside of existing computational models (Dezfouli et al., 2019). Using these approaches on data sets including both task and non-task data will improve predictions and aid in the design effective treatments for psychiatric disorders.

Subjective feelings are not just a byproduct of behavior but may be our best way of understanding emotional process that play an important role in behavior. Task-irrelevant information to manipulate mood is one way to test for an influence of mood on behavior. After a positive mood manipulation, people more quickly update the value of option upwards (Eldar & Niv, 2015; Vinckier et al., 2018). Subjective feelings can also better predict participant risky choices compared with the Prospect Theory model (Charpentier et al., 2016). Approaches that change task conditions as a result of past affective reports (e.g., to increase or decrease happiness) are one way to understand how mood and behavior interact (Keren et al., 2021). Another view of the relationship between mood and behavior is the idea of mood homeostasis which suggests that people stabilize their mood by engaging in mood-modifying activities such that they are more likely to engage in activities that should increase mood while in a low mood state and more likely to engage in activities that lower mood (i.e., washing dishes) while in a high mood state (Quoidbach et al., 2019; Taquet et al., 2016). This mood regulation appeared to be impaired in people with depression (Taquet et al., 2021). These findings together illustrate the importance of using measurements of subjective feelings to understand the relationship between emotion and behavior. Furthermore, the association between emotion and behavior could change in psychiatric disorders. This has been found to be the case for other types of subjective reports. In a perceptual task, through transdiagnostic dimensional approaches, high anxious depression was related to low confidence level and high metacognitive efficiency whereas high compulsive behavior and intrusive thought was related to high confidence level and low metacognitive efficiency (Rouault et al., 2018). However, none of these transdiagnostic symptoms was correlated with task accuracy. Additionally, patients with high compulsivity showed weaker associations between changes in behavior in a volatile environment and changes in confidence (Seow & Gillan, 2020; Vaghi et al., 2017).

Computational models of subjective feelings in tasks can also be applied to real life. Using smartphone-based measurement of emotions over multiple days, emotional responses of students receiving exam results was found to depend strongly on expectations and resulting prediction errors (Villano et al., 2020) (see section 6 for more information about smartphone-based research methods). This measure of real-life emotional response could be considered in relation to psychiatric disorders. Prediction errors can also derive from successfully performing a learned skill (e.g., cooking, playing piano, riding a bicycle) in a

manner that might be thought of as more intrinsically than extrinsically rewarding (Chew et al., 2021). Greater modulation of BOLD activity in ventromedial prefrontal cortex by a motor performance aspect of a task was associated with a greater influence of motor performance on mood. Individuals with a more consistent and distinct mobility "footprint" had lower mood instability (Xia et al., 2022). Furthermore, subjective feelings can also help to predict city-level behaviors. Real-world unexpected positive outcomes (e.g., sport results or weather) can increase mood states and risk-taking behaviors in a city (Otto et al., 2016; Otto & Eichstaedt, 2018). Better study of these relationships could help in understanding societal mental health and guiding policymakers. Many studies focus on the association between tasks and individuals, but within-subject variance is just as important. Little is known about the temporal dynamics of symptoms within subjects (Sharp et al., 2020). With smartphones, it is easier to measure dense longitudinal data from individuals. If we repeatedly collect data in a wide variety of tasks, we can better understand how cognitive and affective processes relate to changes in symptoms. In addition to mood in the tasks, it is also important to understand mood dynamics outside of tasks. The pattern of real-life mood dynamics may be a marker of psychiatric disorders. For example, bipolar disorder may lead to mood fluctuations that do not affect subjective feelings during tasks (Pulcu et al., 2022). We can collect mood at different times (e.g., morning, evening), days (e.g., weekday, weekend), seasons (e.g., summer, winter), and in relation to major societal or personal events. Computational models of subjective feelings have the possibility to bridge the gap between behavior and symptoms, offering a new way to understand the heterogeneity of psychiatric disorders and to better predict treatment outcomes.

Acknowledgements

The authors were supported by the National Institute of Mental Health (1R01MH124110).

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th edition). 10.1176/appi.books.9780890425596

Behrens TEJ, Muller TH, Whittington JCR, Mark S, Baram AB, Stachenfeld KL, & Kurth-Nelson Z. (2018). What is a cognitive map? Organizing knowledge for flexible behavior. Neuron, 100(2), 490–509. 10.1016/j.neuron.2018.10.002 [PubMed: 30359611]

Bennett D, Davidson G, & Niv Y. (2022). A model of mood as integrated advantage. Psychological Review, 129(3), 513–541. 10.1037/rev0000294 [PubMed: 34516150]

Bernoulli D. (1954). Exposition of a new theory on the measurement of risk. Econometrica, 22(1), 23–36. 10.2307/1909829

Blain B, & Rutledge RB (2020). Momentary subjective well-being depends on learning and not reward. ELife, 9, e57977. 10.7554/eLife.57977

Boyce CJ, Brown GD, & Moore SC (2010). Money and happiness: Rank of income, not income, affects life satisfaction. Psychological Science, 21(4), 471–475. 10.1177/0956797610362671 [PubMed: 20424085]

Brown VM, Chen J, Gillan CM, & Price RB (2020). Improving the reliability of computational analyses: Model-based planning and its relationship with compulsivity. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 5(6), 601–609. 10.1016/j.bpsc.2019.12.019 [PubMed: 32249207]

Browning M, Behrens TE, Jocham G, O'Reilly JX, & Bishop SJ (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. Nature Neuroscience, 18(4), Article 4. 10.1038/nn.3961

Browning M, Carter CS, Chatham C, Ouden HD, Gillan CM, Baker JT, Chekroud AM, Cools R, Dayan P, Gold J, Goldstein RZ, Hartley CA, Kepecs A, Lawson RP, Mourao-Miranda J, Phillips ML, Pizzagalli DA, Powers A, Rindskopf D, … Paulus M. (2020). Realizing the clinical potential of computational psychiatry: Report from the Banbury Center Meeting, February 2019. Biological Psychiatry, 88(2), e5–e10. 10.1016/j.biopsych.2019.12.026 [PubMed: 32113656]

Buch AM, & Liston C. (2021). Dissecting diagnostic heterogeneity in depression by integrating neuroimaging and genetics. Neuropsychopharmacology, 46(1), Article 1. 10.1038/s41386-020-00789-3

Caplin A, Dean M, Glimcher PW, & Rutledge RB (2010). Measuring beliefs and rewards: A neuroeconomic approach. The Quarterly Journal of Economics, 125(3), 923–960. 10.1162/qjec.2010.125.3.923 [PubMed: 25018564]

Chang LJ, & Smith A. (2015). Social emotions and psychological games. Current Opinion in Behavioral Sciences, 5, 133–140. 10.1016/j.cobeha.2015.09.010

Charpentier CJ, Aylward J, Roiser JP, & Robinson OJ (2017). Enhanced risk aversion, but not loss aversion, in unmedicated pathological anxiety. Biological Psychiatry, 81(12), 1014–1022. 10.1016/j.biopsych.2016.12.010 [PubMed: 28126210]

Charpentier CJ, De Neve J-E, Li X, Roiser JP, & Sharot T. (2016). Models of affective decision making: How do feelings predict choice? Psychological Science, 27(6), 763–775. 10.1177/0956797616634654 [PubMed: 27071751]

Chase HW, Camille N, Michael A, Bullmore ET, Robbins TW, & Sahakian BJ (2010). Regret and the negative evaluation of decision outcomes in major depression. Cognitive, Affective, & Behavioral Neuroscience, 10(3), 406–413. 10.3758/CABN.10.3.406

Chen C, Takahashi T, Nakagawa S, Inoue T, & Kusumi I. (2015). Reinforcement learning in depression: A review of computational research. Neuroscience & Biobehavioral Reviews, 55, 247–267. 10.1016/j.neubiorev.2015.05.005 [PubMed: 25979140]

Cheong JH, Jolly E, Sul S, & Chang LJ (2017). Computational models in social neuroscience. In Computational models of brain and behavior (pp. 229–244). John Wiley & Sons, Ltd. 10.1002/9781119159193.ch17

Chew B, Blain B, Dolan RJ, & Rutledge RB (2021). A neurocomputational model for intrinsic reward. Journal of Neuroscience, 41(43), 8963–8971. 10.1523/JNEUROSCI.0858-20.2021 [PubMed: 34544831]

Chung D, Kadlec K, Aimone JA, McCurry K, King-Casas B, & Chiu PH (2017). Valuation in major depression is intact and stable in a non-learning environment. Scientific Reports, 7(1), 44374. 10.1038/srep44374

Coricelli G, Critchley HD, Joffily M, O'Doherty JP, Sirigu A, & Dolan RJ (2005). Regret and its avoidance: A neuroimaging study of choice behavior. Nature Neuroscience, 8(9), Article 9. 10.1038/nn1514

Coricelli G, Dolan RJ, & Sirigu A. (2007). Brain, emotion and decision making: The paradigmatic example of regret. Trends in Cognitive Sciences, 11(6), 258–265. 10.1016/j.tics.2007.04.003 [PubMed: 17475537]

Coricelli G, & Rustichini A. (2010). Counterfactual thinking and emotions: Regret and envy learning. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1538), 241–247. 10.1098/rstb.2009.0159

Critchley HD, Wiens S, Rotshtein P, Ohman A, & Dolan RJ (2004). Neural systems supporting interoceptive awareness. Nature Neuroscience, 7(2), 189–195. 10.1038/nn1176 [PubMed: 14730305]

Cushman F, & Gershman S. (2019). Editors' introduction: Computational approaches to social cognition. Topics in Cognitive Science, 11(2), 281–298. 10.1111/tops.12424 [PubMed: 31025547]

Cuthbert BN, & Insel TR (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. BMC Medicine, 11(1), 126. 10.1186/1741-7015-11-126 [PubMed: 23672542]

Damasio AR (1999). The feeling of what happens: Body and emotion in the making of consciousness. Houghton Mifflin Harcourt.

de Berker AO, Rutledge RB, Mathys C, Marshall L, Cross GF, Dolan RJ, & Bestmann S. (2016). Computations of uncertainty mediate acute stress responses in humans. Nature Communications, 7(1), Article 1. 10.1038/ncomms10996

De Martino B, Camerer CF, & Adolphs R. (2010). Amygdala damage eliminates monetary loss aversion. Proceedings of the National Academy of Sciences, 107(8), 3788–3792. 10.1073/pnas.0910230107

Dezfouli A, Griffiths K, Ramos F, Dayan P, & Balleine BW (2019). Models that learn how humans learn: The case of decision-making and its disorders. PLOS Computational Biology, 15(6), e1006903. 10.1371/journal.pcbi.1006903

Dreisbach G, & Goschke T. (2004). How positive affect modulates cognitive control: Reduced perseveration at the cost of increased distractibility. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 343–353. 10.1037/02787393.30.2.343 [PubMed: 14979809]

Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, Fetcho RN, Zebley B, Oathes DJ, Etkin A, Schatzberg AF, Sudheimer K, Keller J, Mayberg HS, Gunning FM, Alexopoulos GS, Fox MD, Pascual-Leone A, Voss HU, … Liston C. (2017). Resting-state connectivity biomarkers define neurophysiological subtypes of depression. Nature Medicine, 23(1), Article 1. 10.1038/nm.4246

Dunn EW, Aknin LB, & Norton MI (2008). Spending money on others promotes happiness. Science, 319(5870), 1687–1688. 10.1126/science.1150952 [PubMed: 18356530]

Easterlin RA, McVey LA, Switek M, Sawangfa O, & Zweig JS (2010). The happiness–income paradox revisited. Proceedings of the National Academy of Sciences, 107(52), 22463–22468. 10.1073/pnas.1015962107

Eldar E, & Niv Y. (2015). Interaction between emotional state and learning underlies mood instability. Nature Communications, 6(1), Article 1. 10.1038/ncomms7149

Eldar E, Pessiglione M, & van Dillen L. (2021). Positive affect as a computational mechanism. Current Opinion in Behavioral Sciences, 39, 52–57. 10.1016/j.cobeha.2021.01.007

Eldar E, Roth C, Dayan P, & Dolan RJ (2018). Decodability of reward learning signals predicts mood fluctuations. Current Biology, 28(9), 1433–1439.e7. 10.1016/j.cub.2018.03.038 [PubMed: 29706512]

Eldar E, Rutledge RB, Dolan RJ, & Niv Y. (2016). Mood as representation of momentum. Trends in Cognitive Sciences, 20(1), 15–24. 10.1016/j.tics.2015.07.010 [PubMed: 26545853]

Emanuel A, & Eldar E. (2023). Emotions as computations. Neuroscience & Biobehavioral Reviews, 144, 104977. 10.1016/j.neubiorev.2022.104977 [PubMed: 36435390]

FeldmanHall O, & Nassar MR (2021). The computational challenge of social learning. Trends in Cognitive Sciences, 25(12), 1045–1057. 10.1016/j.tics.2021.09.002 [PubMed: 34583876]

Fineberg SK, Leavitt J, Stahl DS, Kronemer S, Landry CD, Alexander-Bloch A, Hunt LT, & Corlett PR (2018). Differential valuation and learning from social and nonsocial cues in borderline personality disorder. Biological Psychiatry, 84(11), 838–845. 10.1016/j.biopsych.2018.05.020 [PubMed: 30041970]

Fleming SM, & Lau HC (2014). How to measure metacognition. Frontiers in Human Neuroscience, 8. https://www.frontiersin.org/articles/10.3389/fnhum.2014.00443

Forgas JP (1995). Mood and judgment: The affect infusion model (AIM). Psychological Bulletin, 117(1), 39–66. 10.1037/0033-2909.117.1.39 [PubMed: 7870863]

Forgeot d'Arc B, Devaine M, & Daunizeau J. (2020). Social behavioural adaptation in Autism. PLOS Computational Biology, 16(3), e1007700. 10.1371/journal.pcbi.1007700

Fried EI, & Nesse RM (2015). Depression is not a consistent syndrome: An investigation of unique symptom patterns in the STAR*D study. Journal of Affective Disorders, 172, 96–102. 10.1016/j.jad.2014.10.010 [PubMed: 25451401]

Gillan CM, Kosinski M, Whelan R, Phelps EA, & Daw ND (2016). Characterizing a psychiatric symptom dimension related to deficits in goal-directed control. ELife, 5, e11305. 10.7554/eLife.11305

Gillan CM, Morein-Zamir S, Kaser M, Fineberg NA, Sule A, Sahakian BJ, Cardinal RN, & Robbins TW (2014). Counterfactual processing of economic action-outcome alternatives in obsessive-compulsive disorder: Further evidence of impaired goal-directed behavior. Biological Psychiatry, 75(8), 639–646. 10.1016/j.biopsych.2013.01.018 [PubMed: 23452663]

Gillan CM, & Rutledge RB (2021). Smartphones and the neuroscience of mental health. Annual Review of Neuroscience, 44(1), 129–151. 10.1146/annurev-neuro101220-014053

Gillan CM, & Seow TXF (2020). Carving out new transdiagnostic dimensions for research in mental health. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 5(10), 932–934. 10.1016/j.bpsc.2020.04.013 [PubMed: 32532686]

Gruenenfelder-Steiger AE, Harris MA, & Fend HA (2016). Subjective and objective peer approval evaluations and self-esteem development: A test of reciprocal, prospective, and long-term effects. Developmental Psychology, 52(10), 1563–1577. 10.1037/dev0000147 [PubMed: 27690495]

Güth W, Schmittberger R, & Schwarze B. (1982). An experimental analysis of ultimatum bargaining. Journal of Economic Behavior & Organization, 3(4), 367–388. 10.1016/0167-2681(82)90011-7

Harari GM, Lane ND, Wang R, Crosier BS, Campbell AT, & Gosling SD (2016). Using smartphones to collect behavioral data in psychological science: Opportunities, practical considerations, and challenges. Perspectives on Psychological Science, 11(6), 838–854. 10.1177/1745691616650285 [PubMed: 27899727]

Harsanyi JC (1961). On the rationality postulates underlying the theory of cooperative games. Journal of Conflict Resolution, 5(2), 179–196. 10.1177/002200276100500205

Heffner J, & FeldmanHall O. (2022). A probabilistic map of emotional experiences during competitive social interactions. Nature Communications, 13(1), Article 1. 10.1038/s41467-022-29372-8

Heffner J, Son J-Y, & FeldmanHall O. (2021). Emotion prediction errors guide socially adaptive behaviour. Nature Human Behaviour, 5(10), 1391–1401. 10.1038/s41562-021-01213-6

Heller AS, Shi TC, Ezie CEC, Reneau TR, Baez LM, Gibbons CJ, & Hartley CA (2020). Association between real-world experiential diversity and positive affect relates to hippocampal–striatal functional connectivity. Nature Neuroscience, 23(7), Article 7. 10.1038/s41593-020-0636-4

Henco L, Diaconescu AO, Lahnakoski JM, Brandi M-L, Hörmann S, Hennings J, Hasan A, Papazova I, Strube W, Bolis D, Schilbach L, & Mathys C. (2020). Aberrant computational mechanisms of social learning and decision-making in schizophrenia and borderline personality disorder. PLOS Computational Biology, 16(9), e1008162. 10.1371/journal.pcbi.1008162

Hitchcock PF, Fried EI, & Frank MJ (2022). Computational psychiatry needs time and context. Annual Review of Psychology, 73(1), 243–270. 10.1146/annurevpsych-021621-124910

Hopkins AK, Dolan R, Button KS, & Moutoussis M. (2021). A reduced self-positive belief underpins greater sensitivity to negative evaluation in socially anxious individuals. Computational Psychiatry, 5(1), 21. 10.5334/cpsy.57 [PubMed: 34212077]

Huys QJM, Maia TV, & Frank MJ (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. Nature Neuroscience, 19(3), 404–413. 10.1038/nn.4238 [PubMed: 26906507]

Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, Sanislow C, & Wang P. (2010). Research domain criteria (RDoC): Toward a new classification framework for research on mental disorders. The American Journal of Psychiatry, 167(7), 748–751. 10.1176/appi.ajp.2010.09091379 [PubMed: 20595427]

Isen AM, Nygren TE, & Ashby FG (1988). Influence of positive affect on the subjective utility of gains and losses: It is just not worth the risk. Journal of Personality and Social Psychology, 55(5), 710–717. 10.1037/0022-3514.55.5.710 [PubMed: 3210141]

Jangraw D, Keren H, Sun H, Bedder R, Rutledge R, Pereira F, Thomas AG, Pine D, Zheng C, Nielson D, & Stringaris A. (2021). Passage-of-time dysphoria: A highly replicable decline in mood during rest and simple tasks that is moderated by depression. PsyArXiv. 10.31234/osf.io/bwv58

Kahneman D, & Deaton A. (2010). High income improves evaluation of life but not emotional well-being. Proceedings of the National Academy of Sciences, 107(38), 16489–16493. 10.1073/pnas.1011492107

Kahneman D, Krueger AB, Schkade D, Schwarz N, & Stone AA (2006). Would you be happier if you were richer? A focusing illusion. Science, 312(5782), 1908–1910. 10.1126/science.1129688 [PubMed: 16809528]

Kahneman D, & Tversky A. (1979). Prospect Theory: An analysis of decision under risk. Econometrica, 47(2), 263–291. 10.2307/1914185

Kaiser C, & Oswald AJ (2022). The scientific value of numerical measures of human feelings. Proceedings of the National Academy of Sciences, 119(42), e2210412119. 10.1073/pnas.2210412119

Keren H, Zheng C, Jangraw DC, Chang K, Vitale A, Rutledge RB, Pereira F, Nielson DM, & Stringaris A. (2021). The temporal representation of experience in subjective mood. ELife, 10, e62051. 10.7554/eLife.62051

Killingsworth MA, & Gilbert DT (2010). A wandering mind is an unhappy mind. Science, 330(6006), 932–932. 10.1126/science.1192439 [PubMed: 21071660]

King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, & Montague PR (2008). The rupture and repair of cooperation in borderline personality disorder. Science, 321(5890), 806–810. 10.1126/science.1156902 [PubMed: 18687957]

Kumar P, Goer F, Murray L, Dillon DG, Beltzer ML, Cohen AL, Brooks NH, & Pizzagalli DA (2018). Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. Neuropsychopharmacology, 43(7), Article 7. 10.1038/s41386-018-0032-x

Lang P. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. Technology in Mental Health Care Delivery Systems, 119–137.

LeDoux JE, & Pine DS (2016). Using neuroscience to help understand fear and anxiety: A two-system framework. American Journal of Psychiatry, 173(11), 1083–1093. 10.1176/appi.ajp.2016.16030353 [PubMed: 27609244]

Lee D. (2008). Game theory and neural basis of social decision making. Nature Neuroscience, 11(4), Article 4. 10.1038/nn2065

Lee VK, & Harris LT (2013). How social cognition can inform social decision making. Frontiers in Neuroscience, 7. https://www.frontiersin.org/articles/10.3389/fnins.2013.00259

Loewenstein GF, & Lerner JS (2003). The role of affect in decision making. In Handbook of affective sciences (pp. 619–642). Oxford University Press.

Loewenstein GF, Weber EU, Hsee CK, & Welch N. (2001). Risk as feelings. Psychological Bulletin, 127(2), 267–286. 10.1037/0033-2909.127.2.267 [PubMed: 11316014]

Low AAY, Hopper WJT, Angelescu I, Mason L, Will G-J, & Moutoussis M. (2022). Self-esteem depends on beliefs about the rate of change of social approval. Scientific Reports, 12(1), 6643. 10.1038/s41598-022-10260-6 [PubMed: 35459920]

Luttmer EFP (2005). Neighbors as negatives: Relative earnings and well-being*. The Quarterly Journal of Economics, 120(3), 963–1002. 10.1093/qje/120.3.963

MacKerron G, & Mourato S. (2013). Happiness is greater in natural environments. Global Environmental Change, 23(5), 992–1000. 10.1016/j.gloenvcha.2013.03.010

Maner JK, Richey JA, Cromer K, Mallott M, Lejuez CW, Joiner TE, & Schmidt NB (2007). Dispositional anxiety and risk-avoidant decision-making. Personality and Individual Differences, 42(4), 665–675. 10.1016/j.paid.2006.08.016

Mason L, Eldar E, & Rutledge RB (2017). Mood instability and reward dysregulation—A neurocomputational model of bipolar disorder. JAMA Psychiatry, 74(12), 1275–1276. 10.1001/jamapsychiatry.2017.3163 [PubMed: 29049438]

Mellers BA, Schwartz A, Ho K, & Ritov I. (1997). Decision affect theory: Emotional reactions to the outcomes of risky options. Psychological Science, 8(6), 423–429. 10.1111/j.1467-9280.1997.tb00455.x

Mellers BA, Yin S, & Berman JZ (2021). Reconciling loss aversion and gain seeking in judged emotions. Current Directions in Psychological Science, 30(2), 95–102. 10.1177/0963721421992043

Michely J, Eldar E, Martin IM, & Dolan RJ (2020). A mechanistic account of serotonin's impact on mood. Nature Communications, 11(1), Article 1. 10.1038/s41467-020-16090-2

Montague PR, Dolan RJ, Friston KJ, & Dayan P. (2012). Computational psychiatry. Trends in Cognitive Sciences, 16(1), 72–80. 10.1016/j.tics.2011.11.018 [PubMed: 22177032]

Montague PR, & Lohrenz T. (2007). To detect and correct: Norm violations and their enforcement. Neuron, 56(1), 14–18. 10.1016/j.neuron.2007.09.020 [PubMed: 17920011]

Morgado P, Sousa N, & Cerqueira J. j. (2015). The impact of stress in decision making in the context of uncertainty. Journal of Neuroscience Research, 93(6), 839–847. 10.1002/jnr.23521 [PubMed: 25483118]

Nair A, Rutledge RB, & Mason L. (2020). Under the hood: Using computational psychiatry to make psychological therapies more mechanism-focused. Frontiers in Psychiatry, 11. https://www.frontiersin.org/article/10.3389/fpsyt.2020.00140

Nassar MR, Rumsey KM, Wilson RC, Parikh K, Heasly B, & Gold JI (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. Nature Neuroscience, 15(7), Article 7. 10.1038/nn.3130

Nelissen RMA, & Zeelenberg M. (2009). Moral emotions as determinants of third-party punishment: Anger, guilt, and the functions of altruistic sanctions. Judgment and Decision Making, 4(7), 543–553.

Nowak MA, Page KM, & Sigmund K. (2000). Fairness versus reason in the ultimatum game. Science, 289(5485), 1773–1775. 10.1126/science.289.5485.1773 [PubMed: 10976075]

Ong DC, Zaki J, & Goodman ND (2019). Computational models of emotion inference in theory of mind: A review and roadmap. Topics in Cognitive Science, 11(2), 338–357. 10.1111/tops.12371 [PubMed: 30066475]

Orth U, Robins RW, & Widaman KF (2012). Life-span development of self-esteem and its effects on important life outcomes. Journal of Personality and Social Psychology, 102(6), 1271–1288. 10.1037/a0025558 [PubMed: 21942279]

Otto AR, & Eichstaedt JC (2018). Real-world unexpected outcomes predict city-level mood states and risk-taking behavior. PLOS ONE, 13(11), e0206923. 10.1371/journal.pone.0206923

Otto AR, Fleming SM, & Glimcher PW (2016). Unexpected but incidental positive outcomes predict real-world gambling. Psychological Science, 27(3), 299–311. 10.1177/0956797615618366 [PubMed: 26796614]

Park SQ, Kahnt T, Dogan A, Strang S, Fehr E, & Tobler PN (2017). A neural link between generosity and happiness. Nature Communications, 8(1), Article 1. 10.1038/ncomms15964

Patzelt EH, Kool W, Millner AJ, & Gershman SJ (2019). Incentives boost model-based control across a range of severity on several psychiatric constructs. Biological Psychiatry, 85(5), 425–433. 10.1016/j.biopsych.2018.06.018 [PubMed: 30077331]

Pike AC, & Robinson OJ (2022). Reinforcement learning in patients with mood and anxiety disorders vs control individuals: A systematic review and meta-analysis. JAMA Psychiatry, 79(4), 313–322. 10.1001/jamapsychiatry.2022.0051 [PubMed: 35234834]

Pillutla MM, & Murnighan JK (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. Organizational Behavior and Human Decision Processes, 68(3), 208–224. 10.1006/obhd.1996.0100

Porcelli AJ, & Delgado MR (2017). Stress and decision making: Effects on valuation, learning, and risk-taking. Current Opinion in Behavioral Sciences, 14, 33–39. 10.1016/j.cobeha.2016.11.015 [PubMed: 28044144]

Pulcu E, & Browning M. (2017). Affective bias as a rational response to the statistics of rewards and punishments. ELife, 6, e27879. 10.7554/eLife.27879

Pulcu E, Saunders KEA, Harmer CJ, Harrison PJ, Goodwin GM, Geddes JR, & Browning M. (2022). Using a generative model of affect to characterize affective variability and its response to treatment in bipolar disorder. Proceedings of the National Academy of Sciences, 119(28), e2202983119. 10.1073/pnas.2202983119

Quoidbach J, Taquet M, Desseilles M, de Montjoye Y-A, & Gross JJ (2019). Happiness and social behavior. Psychological Science, 30(8), 1111–1122. 10.1177/0956797619849666 [PubMed: 31268832]

Rilling JK, & Sanfey AG (2011). The neuroscience of social decision-making. Annual Review of Psychology, 62(1), 23–48. 10.1146/annurev.psych.121208.131647

Rouault M, Seow T, Gillan CM, & Fleming SM (2018). Psychiatric symptom dimensions are associated with dissociable shifts in metacognition but not task performance. Biological Psychiatry, 84(6), 443–451. 10.1016/j.biopsych.2017.12.017 [PubMed: 29458997]

Rutledge RB, Chekroud AM, & Huys QJ (2019). Machine learning and big data in psychiatry: Toward clinical applications. Current Opinion in Neurobiology, 55, 152–159. 10.1016/j.conb.2019.02.006 [PubMed: 30999271]

Rutledge RB, de Berker AO, Espenhahn S, Dayan P, & Dolan RJ (2016). The social contingency of momentary subjective well-being. Nature Communications, 7(1), Article 1. 10.1038/ncomms11825

Rutledge RB, Moutoussis M, Smittenaar P, Zeidman P, Taylor T, Hrynkiewicz L, Lam J, Skandali N, Siegel JZ, Ousdal OT, Prabhu G, Dayan P, Fonagy P, & Dolan RJ (2017). Association of neural and emotional impacts of reward prediction errors with major depression. JAMA Psychiatry, 74(8), 790–797. 10.1001/jamapsychiatry.2017.1713 [PubMed: 28678984]

Rutledge RB, Skandali N, Dayan P, & Dolan RJ (2014). A computational and neural model of momentary subjective well-being. Proceedings of the National Academy of Sciences, 111(33), 12252–12257. 10.1073/pnas.1407535111

Rutledge RB, Skandali N, Dayan P, & Dolan RJ (2015). Dopaminergic modulation of decision making and subjective well-being. Journal of Neuroscience, 35(27), 9811–9822. 10.1523/JNEUROSCI.0702-15.2015 [PubMed: 26156984]

Safra L, Chevallier C, & Palminteri S. (2019). Depressive symptoms are associated with blunted reward learning in social contexts. PLOS Computational Biology, 15(7), e1007224. 10.1371/journal.pcbi.1007224

Sanfey AG, Rilling JK, Aronson JA, Nystrom LE, & Cohen JD (2003). The neural basis of economic decision-making in the ultimatum game. Science, 300(5626), 1755–1758. 10.1126/science.1082976 [PubMed: 12805551]

Schultz W, Dayan P, & Montague PR (1997). A neural substrate of prediction and reward. Science, 275(5306), 1593–1599. 10.1126/science.275.5306.1593 [PubMed: 9054347]

Seow TXF, & Gillan CM (2020). Transdiagnostic phenotyping reveals a host of metacognitive deficits implicated in compulsivity. Scientific Reports, 10(1), Article 1. 10.1038/s41598-020-59646-4

Seth AK (2013). Interoceptive inference, emotion, and the embodied self. Trends in Cognitive Sciences, 17(11), 565–573. 10.1016/j.tics.2013.09.007 [PubMed: 24126130]

Sharp PB, Dolan RJ, & Eldar E. (2021). Disrupted state transition learning as a computational marker of compulsivity. Psychological Medicine, 1–11. 10.1017/S0033291721003846

Sharp PB, Miller GA, Dolan RJ, & Eldar E. (2020). Towards formal models of psychopathological traits that explain symptom trajectories. BMC Medicine, 18(1), 264. 10.1186/s12916-020-01725-4 [PubMed: 32981516]

Sharp PB, Miller GA, & Heller W. (2015). Transdiagnostic dimensions of anxiety: Neural mechanisms, executive functions, and new directions. International Journal of Psychophysiology, 98(2, Part 2), 365–377. 10.1016/j.ijpsycho.2015.07.001 [PubMed: 26156938]

Sharp PB, Russek EM, Huys QJ, Dolan RJ, & Eldar E. (2022). Humans perseverate on punishment avoidance goals in multigoal reinforcement learning. ELife, 11, e74402. 10.7554/eLife.74402

Sharp PB, Sutton BP, Paul EJ, Sherepa N, Hillman CH, Cohen NJ, Kramer AF, Prakash RS, Heller W, Telzer EH, & Barbey AK (2018). Mindfulness training induces structural connectome changes in insula networks. Scientific Reports, 8(1), Article 1. 10.1038/s41598-018-26268-w

Siegel JZ, Curwell-Parry O, Pearce S, Saunders KEA, & Crockett MJ (2020). A computational phenotype of disrupted moral inference in borderline personality disorder. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 5(12), 1134–1141. 10.1016/j.bpsc.2020.07.013 [PubMed: 33012682]

Sokol-Hessner P, Hsu M, Curley NG, Delgado MR, Camerer CF, & Phelps EA (2009). Thinking like a trader selectively reduces individuals' loss aversion. Proceedings of the National Academy of Sciences, 106(13), 5035–5040. 10.1073/pnas.0806761106

Sokol-Hessner P, & Rutledge RB (2019). The psychological and neural basis of loss aversion. Current Directions in Psychological Science, 28(1), 20–27. 10.1177/0963721418806510

Stanton SJ, Reeck C, Huettel SA, & LaBar KS (2014). Effects of induced moods on economic choices. Judgment and Decision Making, 9(2), 9.

Sutton RS, & Barto AG (2018). Reinforcement learning: An introduction. MIT Press.

Taquet M, Quoidbach J, de Montjoye Y-A, Desseilles M, & Gross JJ (2016). Hedonism and the choice of everyday activities. Proceedings of the National Academy of Sciences, 113(35), 9769–9773. 10.1073/pnas.1519998113

Taquet M, Quoidbach J, Fried EI, & Goodwin GM (2021). Mood homeostasis before and during the coronavirus disease 2019 (COVID-19) lockdown among students in the Netherlands. JAMA Psychiatry, 78(1), 110–112. 10.1001/jamapsychiatry.2020.2389 [PubMed: 32725176]

Tom SM, Fox CR, Trepel C, & Poldrack RA (2007). The neural basis of loss aversion in decision-making under risk. Science, 315(5811), 515–518. 10.1126/science.1134239 [PubMed: 17255512]

Trampe D, Quoidbach J, & Taquet M. (2015). Emotions in everyday life. PLOS ONE, 10(12), e0145450. 10.1371/journal.pone.0145450

Vaghi MM, Luyckx F, Sule A, Fineberg NA, Robbins TW, & De Martino B. (2017). Compulsivity reveals a novel dissociation between action and confidence. Neuron, 96(2), 348–354.e4. 10.1016/j.neuron.2017.09.006 [PubMed: 28965997]

van ‘t Wout M, Kahn RS, Sanfey AG, & Aleman A. (2006). Affective state and decision-making in the Ultimatum Game. Experimental Brain Research, 169(4), 564–568. 10.1007/s00221-006-0346-5 [PubMed: 16489438]

Vanhasbroeck N, Devos L, Pessers S, Kuppens P, Vanpaemel W, Moors A, & Tuerlinckx F. (2021). Testing a computational model of subjective well-being: A preregistered replication of Rutledge et al. (2014). Cognition and Emotion, 35(4), 822–835. 10.1080/02699931.2021.1891863 [PubMed: 33632071]

Villano WJ, Otto AR, Ezie CEC, Gillis R, & Heller AS (2020). Temporal dynamics of real-world emotion are more strongly linked to prediction error than outcome. Journal of Experimental Psychology: General, 149(9), 1755–1766. 10.1037/xge0000740 [PubMed: 32039625]

Vinckier F, Rigoux L, Oudiette D, & Pessiglione M. (2018). Neuro-computational account of how mood fluctuations arise and affect decision making. Nature Communications, 9(1), Article 1. 10.1038/s41467-018-03774-z

Von Neumann J, & Morgenstern O. (1944). Theory of games and economic behavior (pp. xviii, 625). Princeton University Press.

Will G-J, Moutoussis M, Womack PM, Bullmore ET, Goodyer IM, Fonagy P, Jones PB, Rutledge RB, & Dolan RJ (2020). Neurocomputational mechanisms underpinning aberrant social learning in young adults with low self-esteem. Translational Psychiatry, 10(1), Article 1. 10.1038/s41398-020-0702-4

Will G-J, Rutledge RB, Moutoussis M, & Dolan RJ (2017). Neural and computational processes underlying dynamic changes in self-esteem. ELife, 6, e28098. 10.7554/eLife.28098

Wise T, & Dolan RJ (2020). Associations between aversive learning processes and transdiagnostic psychiatric symptoms in a general population sample. Nature Communications, 11(1), 4179. 10.1038/s41467-020-17977-w

Wu Y, Schulz LE, Frank MC, & Gweon H. (2021). Emotion as information in early social learning. Current Directions in Psychological Science, 30(6), 468–475. 10.1177/09637214211040779

Xia CH, Barnett I, Tapera TM, Adebimpe A, Baker JT, Bassett DS, Brotman MA, Calkins ME, Cui Z, Leibenluft E, Linguiti S, Lydon-Staley DM, Martin ML, Moore TM, Murtha K, Piiwaa K, Pines A, Roalf DR, Rush-Goebel S, … Satterthwaite TD (2022). Mobile footprinting: Linking individual distinctiveness in mobility patterns to mood, sleep, and brain functional connectivity. Neuropsychopharmacology, 1–10. 10.1038/s41386-022-01351-z

Xiang T, Lohrenz T, & Montague PR (2013). Computational substrates of norms and their violations during social exchange. Journal of Neuroscience, 33(3), 1099–1108. 10.1523/JNEUROSCI.1642-12.2013 [PubMed: 23325247]

Yip SW, Barch DM, Chase HW, Flagel S, Huys QJM, Konova AB, Montague R, & Paulus M. (2022). From computation to clinic. Biological Psychiatry Global Open Science, S2667174322000507. 10.1016/j.bpsgos.2022.03.011

Highlights:

- Models of subjective feelings can quantify the influence of many factors.
- Subjective feelings can be dissociable from behavior in psychiatric disorders.
- Smartphones can be used to measure subjective feelings in tasks and real life.

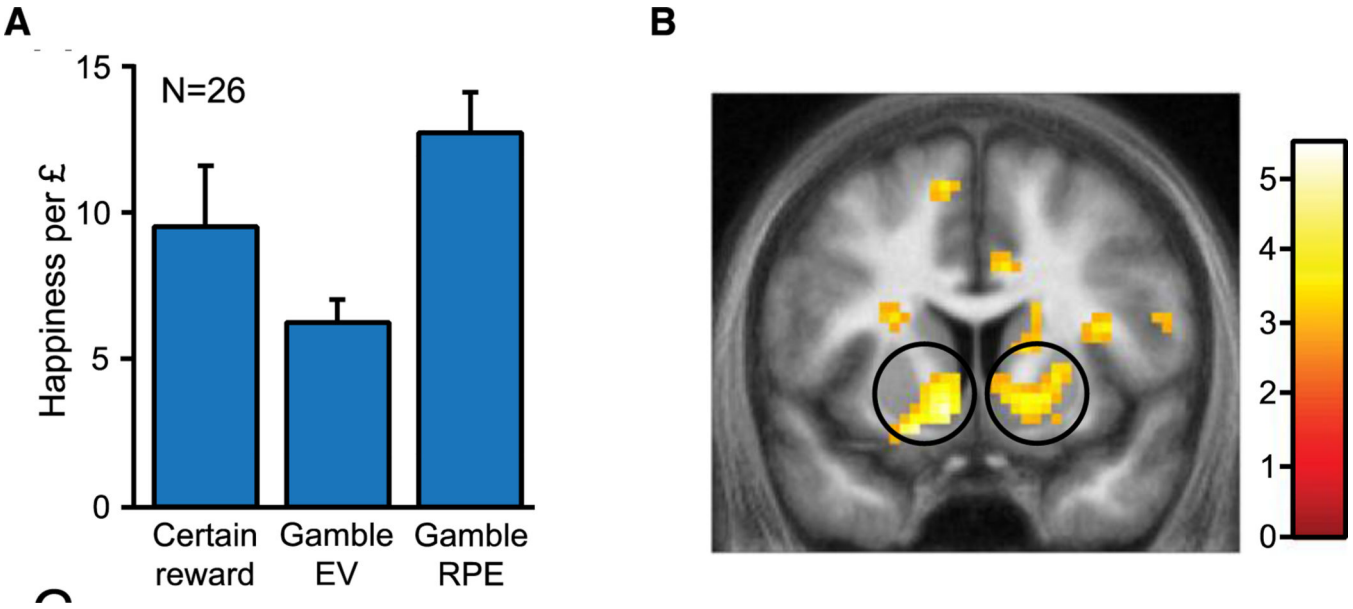

Figure 1. Computational modeling of subjective well-being and fMRI analysis of striatal activity during risky decision making.

(A) The computational model that explained happiness had positive weights for certain reward, gamble expected value, and gamble reward prediction errors. **(B)** Neural responses in the ventral striatum preceding happiness ratings correlated with later self-reported happiness. These neural responses were explained by the same task variables used to explain happiness in the computational model (Rutledge et al., 2014).

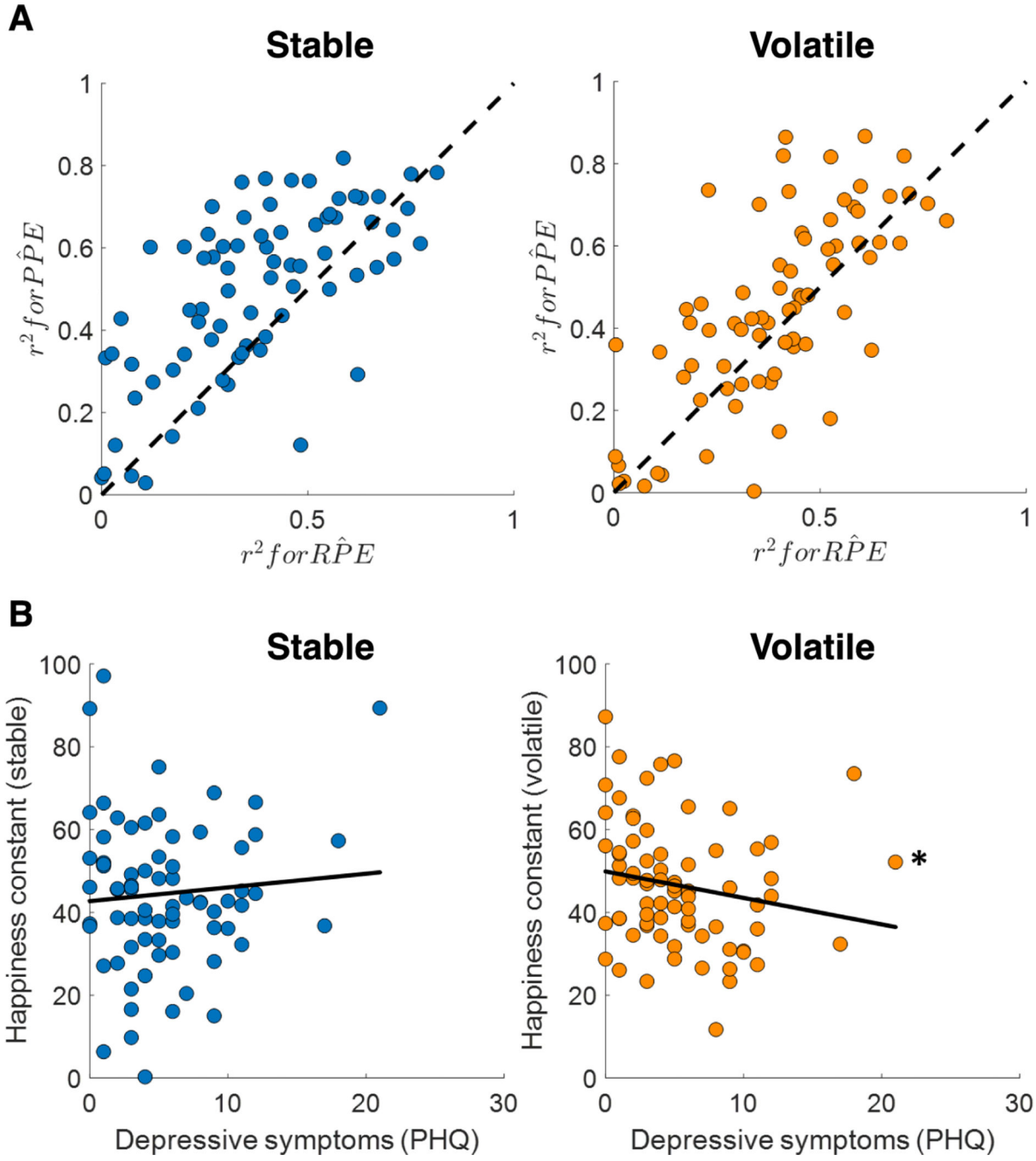

Figure 2. Momentary happiness in a learning task

(A) The model including probability prediction errors (PPE) performed better than the model including reward prediction errors (RPE) in both stable and volatile environments. Each data point indicates a participant. **(B)** The happiness constant or baseline mood parameter was correlated with depressive symptoms in volatile but not stable environments. This parameter was estimated from the happiness model that simultaneously quantifies the influence of expected probabilities and probability prediction errors on happiness (Blain & Rutledge, 2020). * $p < 0.05$.

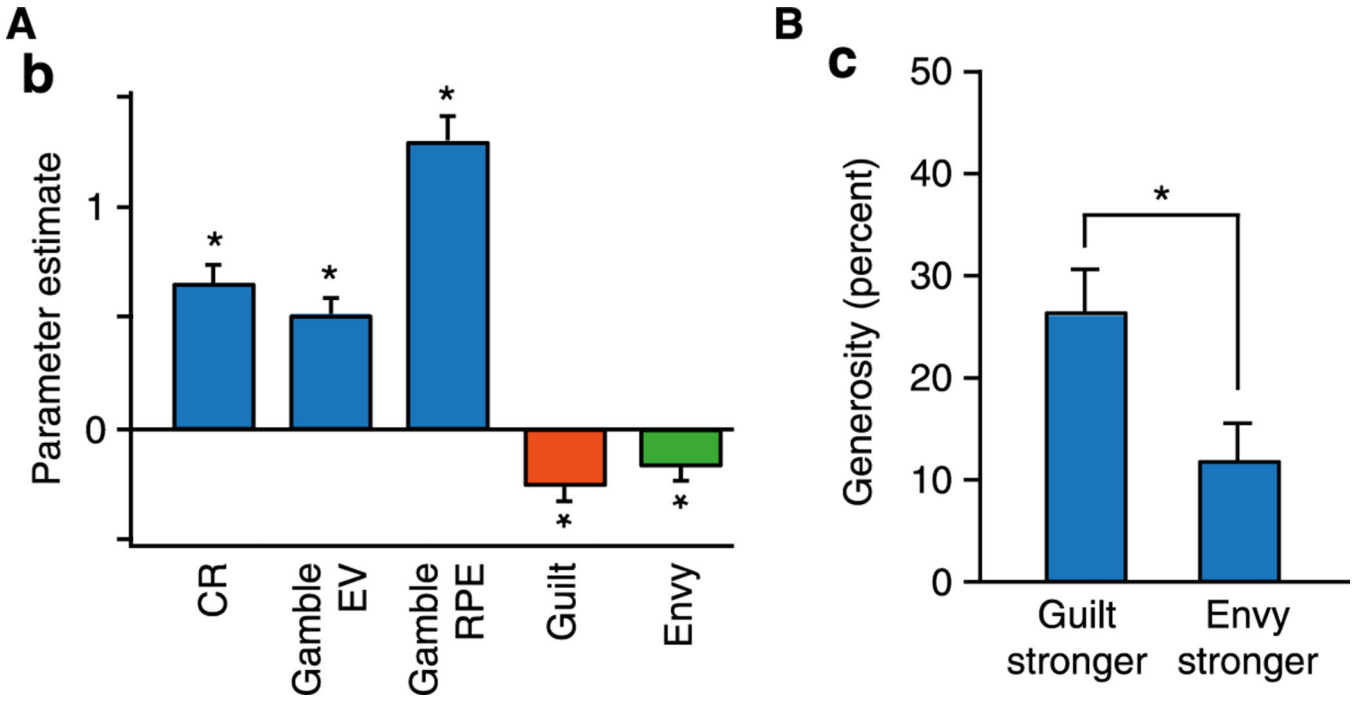

Figure 3. Computational modeling of affective responses to inequality during social decision making.

(A) Consistent with previous work modeling subjective feelings in a risky decision task, happiness depended on the value of recent chosen rewards (CR), the expected value of recent gambles (Gamble EV), and reward prediction errors resulting from gambles (Gamble RPE). Critically, in addition to these reward values, more negative parameter estimates of guilt (orange) and envy (green) predicted lower happiness. **(B)** Guilt and envy parameter estimates predicted generosity in a separate dictator game, as measured by the percentage of a monetary sum that participants allocated to their social partner. Participants whose happiness was reduced more by guilt than envy gave more on average compared to participants whose happiness was reduced more by envy (Rutledge et al., 2016). * $p<0.05$.

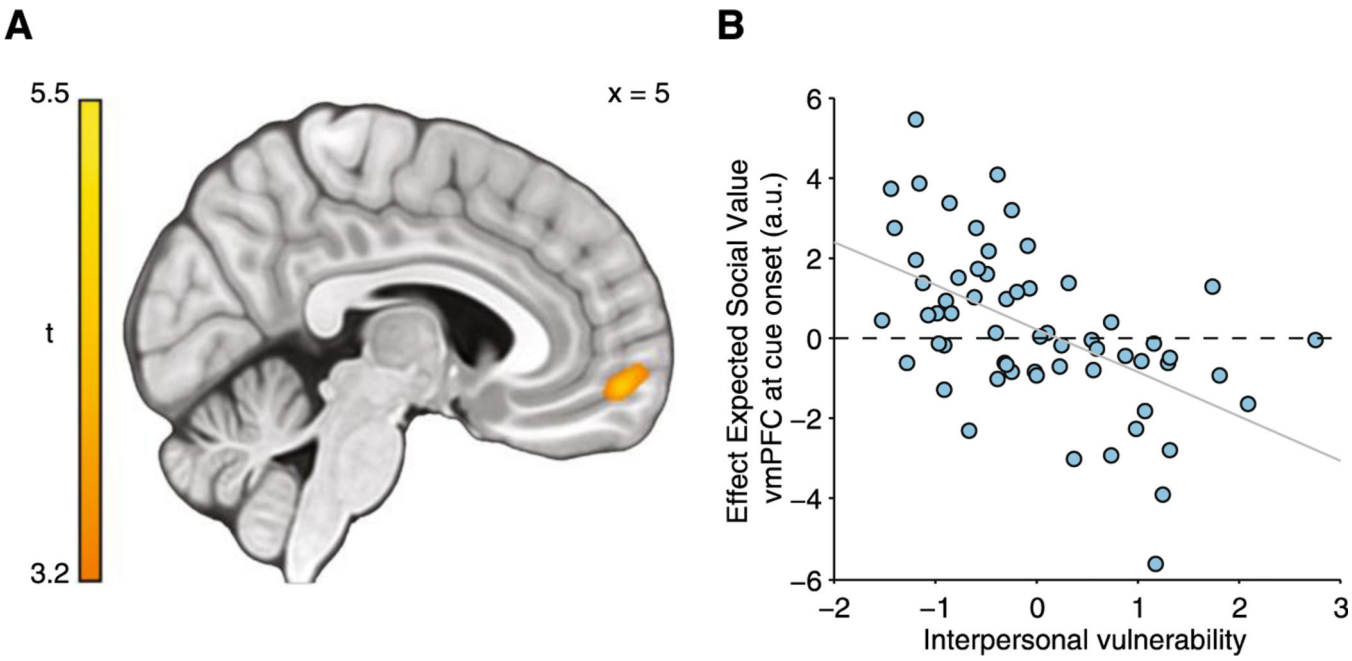

Figure 4. Self-esteem was associated with an "interpersonal vulnerability" dimension reflecting both model parameters and symptoms.

Will et al. (2020) derived a computational phenotype with a single dimension of "interpersonal vulnerability". **(A)** BOLD activity in ventromedial prefrontal cortex correlated with the extent to which participants expected approval on the current trial. **(B)** Higher scores on this dimension predicted attenuated expected approval signal in ventromedial prefrontal cortex (Will et al., 2020).

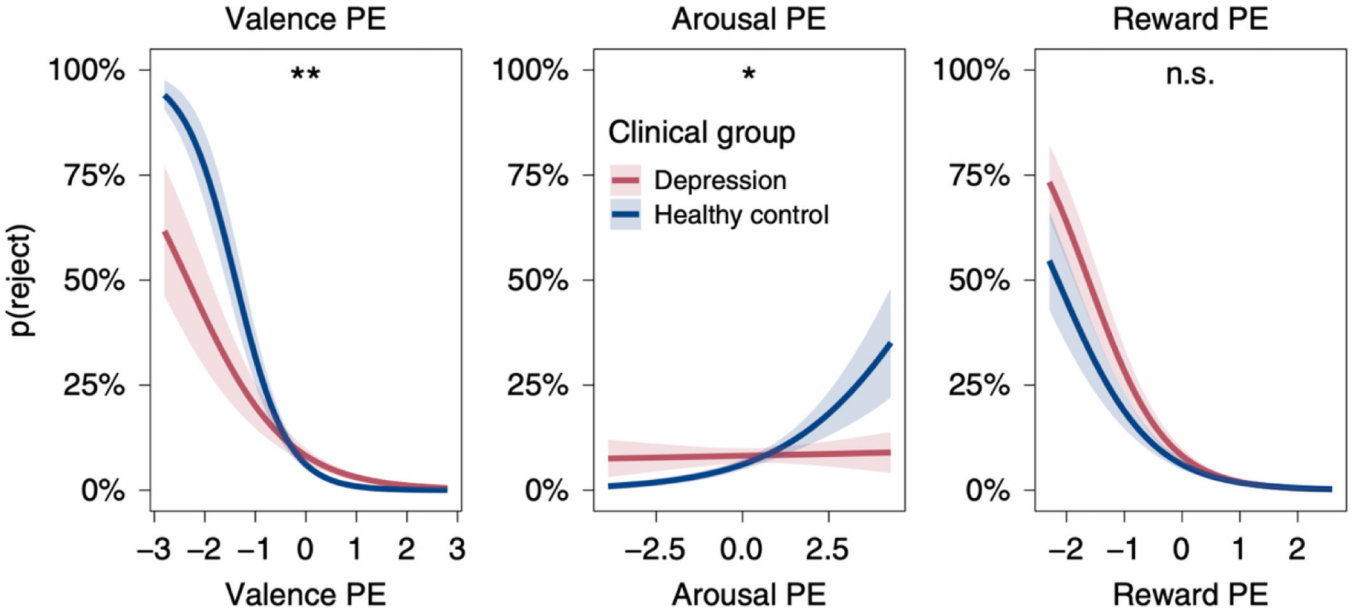

Figure 5. Emotion prediction errors in the ultimatum game.

Emotion prediction errors were computed as the difference between expected and experienced emotion. People were more likely to reject unfair offers when experiencing less valence or more arousal than expected. This influence of emotion prediction errors was significantly reduced in depression. (Heffner et al., 2021). * $p<0.05$, ** $p<0.01$.