

Understanding the Neurobiology and Computational Mechanisms of Social Conformity: Implications for Psychiatric Disorders

Yutong Li, Yafeng Pan, and Di Zhao

ABSTRACT

Social conformity and psychiatric disorders share overlapping brain regions and neural pathways, arousing our interest in uncovering their potentially shared underlying neural and computational mechanisms. Critically, the dynamics of group behavior may either mitigate or exacerbate mental health conditions, highlighting the need to bridge social neuroscience and psychiatry. Our work examines how aberrant neurobiological circuits and computations influence social conformity. We propose a hierarchical computational framework, based on dynamic systems and active inference, to facilitate the interpretation of the multilayered interplay among processes that drive social conformity. We underscore the significant implications of this hierarchical computational framework for guiding future research on psychiatry, particularly with respect to the clinical translation of interventions such as targeted pharmacotherapy and neurostimulation techniques. Interdisciplinary efforts hold the potential to propel the fields of social and clinical neuroscience forward, fostering the emergence of more efficacious and individualized therapeutic approaches tailored to psychiatric disorders characterized by aberrant social behaviors.

<https://doi.org/10.1016/j.biopsych.2025.05.011>

Social conformity refers to the act of changing one's attitudes, beliefs, and behaviors to align with others (1). It is a critical aspect of human social behavior, facilitating collective decision making through the optimization of decisions based on information acquisition (2). As social beings, humans demonstrate conforming behaviors from birth, from infants being influenced by others' crying to adolescents following trends (1). Conformity fulfills the need for belonging, but maladaptive conformity can lead to loneliness and depression (2,3). This complex social phenomenon has been the subject of clinical research that takes culture, cognition, and individual differences into account, and this research has yielded insight into the underlying computational and neurobiological processes (4). Recent research has established a compelling connection between aberrations in social conformity and the core symptoms of psychiatric disorders across various animal models (5). This observation underscores the significance of exploring the underlying neural and computational mechanisms that govern social conformity and their potential implications for mental health. Computational approaches provide explanations based on data from neuroimaging and psychiatric disorders with complex interactions ranging from the subcellular to societal levels (6). We start by surveying previous research to uncover the neural and computational mechanisms of social conformity, as well as the overlap seen between the neural circuits responsible for social conformity and those found to be aberrant in psychiatric patients with various diagnoses. This leads us to propose a computational framework that leverages a hierarchical model to quantify social conformity in psychiatric

patients. Next, we highlight how the neurobiological network and hierarchical computational model help to assess the dynamic interaction in social context, particularly for social conformity. Finally, we discuss challenges and opportunities for this framework that may lead to new strategies for helping psychiatric patients to recover social functions.

THE NEUROBIOLOGICAL AND COMPUTATIONAL MECHANISMS OF SOCIAL CONFORMITY

Previous research has extensively investigated the brain regions involved in social conformity, linking this process to mechanisms such as reinforcement learning, cognitive decision making, and cognitive dissonance (7–9). Building on this foundation, in the current review, we aim to explore the intricate neurobiological and computational underpinnings of social conformity, elucidating the complex interplay between these neural systems and an internal model of decision making. The brain regions implicated in these processes are primarily associated with the reward circuit, the punishment-related circuit, and the affective circuit (Figure 1). Social conformity engages core neural circuits involved in social decision making, including reward, punishment, and affective networks. Notably, neural activities are not exclusive to social interactions and are verified to have significant correlations with the basis functions. The projection from neurobiology to basis functions abstracts complex social relationships into a low-dimensional space, enabling a computational understanding of social conformity.

SEE COMMENTARY ON PAGE 350

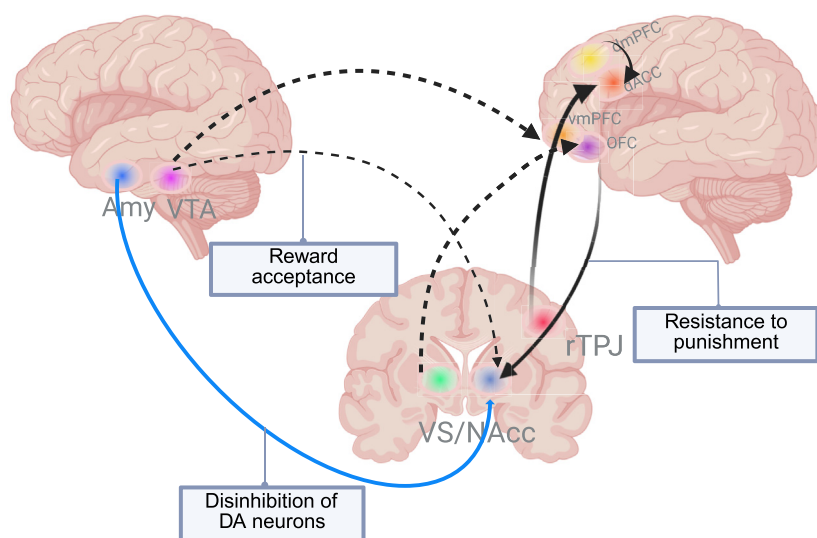


Figure 1. The underlying neurobiological circuits in social conformity. Arrows show 3 neural pathways related to social conformity. Solid black arrows associate major brain regions involved in the punishment-related brain circuit, while dashed black arrows show a loop of activity in the reward circuit. The blue line shows the crucial region of the affective circuit. Amy, amygdala; DA, dopamine; dACC, dorsal anterior cingulate cortex; dmPFC, dorsomedial prefrontal cortex; NAcc, nucleus accumbens; OFC, orbitofrontal cortex; rTPJ, right temporoparietal junction; vmPFC, ventromedial prefrontal cortex; VS, ventral striatum; VTA, ventral tegmental area.

Reward Circuitry

The intricate interplay of the reward circuitry, comprising the ventral striatum (VS), orbitofrontal cortex (OFC), and medial prefrontal cortex (mPFC), fundamentally motivates individuals to seek social recognition and group belongingness (7,10). This circuitry fosters a behavioral tendency wherein individuals align their attitudes to conform with the majority, anticipating positive social reinforcements such as encouraging gestures and verbal affirmations. Classic paradigms for studying social conformity originate from the work of Muzafer Sherif and Solomon Asch. In Asch's line judgment task, individuals accurately identified line lengths when alone but conformed to incorrect group responses under social pressure. These findings laid the groundwork for modern paradigms, such as facial attractiveness evaluation tasks, which engage reward-related neural circuits implicated in social conformity (11). Notably, in facial attractiveness evaluations, heightened activation within the OFC and VS regions was evident when participants demonstrated a preference for faces highly rated by their peers (12). Furthermore, the ventromedial PFC, renowned for encoding subjective values, plays a pivotal role in weighing social values during decision-making processes (13).

According to the framework of adaptive social learning and reinforcement learning theory, the potential value of attitude shifts is dynamically modulated by the choices and behaviors of one's companions, leading to conformity behavior (14,15). Computationally, neural substrates including the dorsal striatum exhibit functions of prediction error detecting, which adjust the model-free system through social learning (16). The emotional bonus derived from valuing social group memberships prompts individuals to align their perception and behaviors with group norms (17), thereby reinforcing behaviors that lead to rewarding outcomes and enhancing their future recurrence. Dopamine (DA), the neurotransmitter that bridges the reward system and reinforcement learning (18), plays a central role in this mechanism. The amygdala, via projections to GABAergic (gamma-aminobutyric acid) interneurons in

the ventral tegmental area (VTA), disinhibits DA neurons, enhancing DA release in the nucleus accumbens (NAcc) during social reward processing (18,19). This circuitry plays a role in assessing stimuli as rewarding or punishing. In autism spectrum disorder (ASD), dysregulated DA release may reduce sensitivity to social rewards (20,21). The pattern of DA bursts is intimately tied to reward prediction errors, which are the primary drivers of reinforcement learning. The action-outcome relationship motivates reinforcement learning computation to guide behaviors (22), suggesting that social conformity emerges as a consequence of adaptive social learning and reinforcement learning processes.

While empirical investigations into the nexus between social conformity and mental disorders remain scarce, some studies have illuminated how aberrant reward circuitry contributes to core symptomatology in psychiatric disorders. For example, individuals with ASD exhibit a notably reduced propensity for conformity, resulting in a deficit in conformity consciousness (21) (Table 1). While consensus on a single aberrant brain structure has not yet been reached, evidence points to abnormalities in the NAcc region as potentially contributing to the difficulties of patients with ASD in encoding group norms (22,23). Loneliness may also result in a weak motivation to comply with social norms (24). The aberrant neural mechanism is impaired at detecting discrepancies between prediction and reality, thereby weakening patients' sensitivity to rewards, according to a computational learning model (25).

Punishment Circuitry

Punishment circuitry, comprising the posterior medial frontal cortex (pmPFC) and anterior insula (AI), plays a pivotal role in avoiding social pain and exclusion (26). This system serves as a safeguard against social isolation because individuals inherently seek alignment with group members to maintain social bonds (27). When exclusion occurs, the resultant distressful social interactions evoke feelings of social pain, triggering the activation of punishment circuits to avert such negative outcomes (7).

Table 1. Aberrant Neural Circuits and Social Conformity in Psychiatric Disorders

Reference	Psychiatric Disorders	Aberrant Neural Circuits	Social Conformity
Shamay-Tsoory <i>et al.</i> (22)	ASD	Reward circuit (NAcc)	Lack of social conformity
Kanai <i>et al.</i> (24)	Loneliness	Punishment circuit (STS and dmPFC)	Lack of social conformity
Volkow and Morales (37)	SUD	Reward and punishment circuits (VS and mPFC)	Conform to risky choices
Chechko <i>et al.</i> (39)	BPD	Punishment circuit (dmPFC and insula)	Maladaptive social conformity
Liu <i>et al.</i> (41)	OCD	Punishment circuit (hyperactive dACC)	Excessive social conformity
Dyck <i>et al.</i> (50)	ADHD	Affective circuit (amygdala)	Lack of social conformity
Wolf <i>et al.</i> (54)	PT-DD	Reward and affective circuits (amygdala and NAcc)	Lack of social conformity
Ganella <i>et al.</i> (55)	SAD	Affective circuit (amygdala)	Excessive social conformity

This table links psychiatric conditions to dysregulated brain circuits and atypical social conformity behaviors. Abnormalities in the NAcc region contributed to the difficulties of patients with ASD in encoding group norms. Individuals experiencing loneliness exhibited reduced gray matter density in the STS and dmPFC, indicating impaired reward circuits. In SUDs, aberrant circuitry suppressed punishment signals, while reduced dopamine D₂ receptor signaling in the VS impaired self-regulation and increased susceptibility to punishment. Individuals with ADHD had a significant size discrepancy between the left and right amygdala surfaces, with the left amygdala implicated in modulating emotional experiences and organizing language related to affective events.

ADHD, attention-deficit/hyperactivity disorder; ASD, autism spectrum disorder; BPD, borderline personality disorder; dACC, dorsal anterior cingulate cortex; dmPFC, dorsomedial prefrontal cortex; mPFC, medial prefrontal cortex; NAcc, nucleus accumbens; OCD, obsessive-compulsive disorder; PT-DD, persistent depressive disorder; SAD, social anxiety disorder; STS, superior temporal sulcus; SUD, substance use disorder; VS, ventral striatum.

However, psychiatric patients often exhibit difficulties in learning from punishment. Illness arises when the computational capacity of the brain fails to match the environmental demand (28). Erroneous predictions and expectancy computations can result from abnormal pmPFC and periaqueductal gray, which are also related to subjective value projection (29,30).

Recent studies have underscored the importance of the dorsomedial PFC (dmPFC) and dorsal anterior cingulate cortex (dACC) within the broader pmPFC network. These regions are implicated in modulating cognitive dissonance and facilitating social conformity (31,32). For example, the dACC is activated during adaptation to group norms, helping to reconcile discrepancies, maintain cognitive harmony, and adjust the inner model (33). Moreover, dmPFC activity is intimately tied to stereotypes and nonconformity behaviors (34), as evidenced by heightened dmPFC activation during interactions with foreigners (35,36). The notion of active inference attributes behavioral choices in social context to the adaptive process of inferring the most probable cause of observed phenomena based on prior knowledge and new evidence.

Substance use disorders, characterized as chronic neurological disorders, disrupt both the reward and the punishment circuits (37,38). Similarly, individuals with borderline personality disorder exhibit impaired conflict-control activity in regions like the dmPFC and insula, indicative of deficits in regulating social negative emotions (39). This dysfunction may contribute to deviant behaviors and disregard for social norms (40). In contrast, patients with obsessive-compulsive disorder (OCD) display hypervigilance in the dACC and dmPFC, leading to exaggerated responses to errors and conflicts (41). This heightened attention to social congruence, coupled with anxiety, underlies the pathological nature of OCD-related social conformity (42). The intricate interplay between punishment circuits, neurochemical modulators, and neuropsychiatric conditions underscores the complexity of social conformity and deviance (43–45).

Affective Circuitry

The affective system, which is tied to fear and other negative emotions, plays a pivotal role in guiding individuals' decisions

to conform or dissent in social settings (46). Social anxiety can arise when individuals experience disagreements with close friends, triggering the activation of the central amygdala (47). Aberrant social conformity can be quantified from the perspective of motivation: Certain parameters influence individuals to either obsess about or avoid emotional experiences in social situations (48). This phenomenon underscores the amygdala's involvement in regulating emotional responses to social challenges. Recent research has further illuminated a robust interconnection between the amygdala and reward circuits, manifesting across both negative and positive social scenarios (49). Positive events such as gaining social recognition promote concurrent activation of the amygdala and the right temporoparietal junction, highlighting the dynamic interplay between emotion and social reinforcement (47).

The amygdala, a pivotal hub for emotion processing, has garnered extensive scientific interest due to its association with various mental disorders. Notably, amygdala structural abnormalities are crucial factors in the development of ASDs and attention-deficit/hyperactivity disorder (50,51). Animal models have also demonstrated the amygdala's involvement in drug addiction. During opiate abstinence, the amygdala receives robust projections from the DA system originating in the VTA, suggesting that the amygdala, together with DA, mediates negative emotions associated with drug treatment cessation (52). Moreover, paranoid-type delusional disorder, a prevalent subtype of schizophrenia, is primarily attributed to functional impairments in the amygdala and OFC (53–55). Computational modeling of the mechanism of anxiety disorders has been attempted, but progress has been hampered by a lack of illuminating neuroimaging data (56).

Neurobiological circuits that underlie social conformity, including reward, punishment, and affective networks, interact dynamically rather than function in isolation. The VS and AI coactivate during group decision making, balancing social acceptance and exclusion risks, while the amygdala enhances the emotional salience of feedback, modulating striatal reward signals (26,30,47). However, neurobiological findings alone cannot fully delineate these interactions or establish causal links to conformity-driven behavior. Computational models

provide a framework to quantify these processes, integrating neural mechanisms and behavioral outcomes to refine our understanding of social conformity in mental disorders. To fully elucidate the intricate mechanisms and accurately quantify the parameters that underlie abnormal social conformity behaviors, the integration of computational models into research is imperative. This approach will facilitate a deeper understanding of the relationships between neural circuits, genetic variation, and behavioral outcomes, ultimately informing the development of targeted interventions for mental health disorders.

COMPUTATIONAL PSYCHIATRY APPROACHES TO STUDYING SOCIAL CONFORMITY

Computational psychiatry provides powerful tools for unraveling the intricacies of social conformity, particularly considering the limitations inherent in traditional decision-making frameworks (56). For example, Von Neumann and Morgenstern's seminal expected utility theory focuses on rational actions and maximizing benefits amidst uncertainty; however, human behavior frequently appears to lack rationality, particularly within social contexts. Notably, it is increasingly understood that there is a pervasive tendency among individuals, especially those with psychiatric conditions, to forgo self-benefit maximization in real-world social decision-making scenarios (19,57). This is readily apparent across various social performance tasks, including two-player economic paradigms, where the complexities of human behavior defy simplistic rational explanations (19). Computational psychiatry offers fresh insights into social conformity, addressing hidden states of social interaction (47).

Moreover, computational techniques have been undervalued in the realm of psychiatric disorders (47). The brain is responsible for an enormous amount of data computation due to the never-ending barrage of environmental and internal

stimuli and can sometimes be overloaded (28). Such computational errors in the brain, which are separate from neurological lesions, can lead to mental illnesses (Figure 2). Therefore, the advancement of computational neuroscience and machine learning, which can shed light on these processes, will be instrumental for future improvements in clinical treatment. Another important issue in clinical neuroscience is understanding transdiagnostic processes. Despite the apparent heterogeneity, similar biology and neural substrates related to social conformity may be shared across a range of different mental disorders. Validated computational approaches show promise for uncovering the shared psychological patterns across social contexts and in different psychological diseases (58). Computational psychiatry utilizes mathematical models and transdiagnostic methods to explore mechanisms of abnormal social behaviors. However, to date there has been only unidirectional translation from clinical data to computational psychiatry. To deepen our understanding of social conformity in psychiatry and to further promote the translation of active guidance from computational models into effective interventions, here we introduce a novel framework based on the integration of hierarchical computation and active inference with the aim of revealing previously undiscovered mechanisms that can lead to better treatments.

Hierarchical Computational Framework

Active inference, based on Bayesian models, has emerged as a promising avenue for understanding, predicting, and monitoring social conformity among psychiatric patients. This computational approach has been successfully applied to human decision making and social coordination in schizophrenia (59,60). According to active inference, social conformity is considered an adaptive behavior approach to achieving interaction between the agent and the environment. Due to the dynamic interactions between individuals, expectations tend

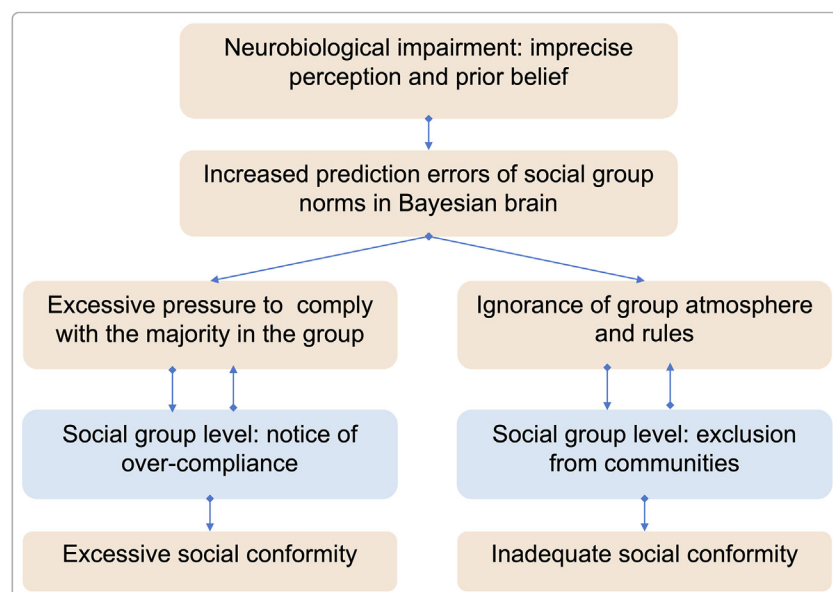


Figure 2. Biased cognitive processes lead to negative conformity outcomes. The relationship between hidden states and aberrant social behaviors in psychiatric patients. The blue panels emphasize group beliefs' influence on individuals during social activities. The orange panels highlight hidden states on the individual's level.

to converge when engaging in social interaction, as in the way that two birds coordinate their singing in a duet (61). To understand the motivation for behavioral choices, active inference conceptualizes the decision-making process as one of inferring the most probable cause of an observed phenomenon based on prior knowledge and new evidence. Under this premise, positive symptoms of schizophrenia such as hallucinations and delusions can be explained by a failed attempt to reduce the uncertainty of interactions between the agent and environment (57); this may stem from a lack of prior knowledge, including forgetting previously effective actions, in the process of optimizing outcomes. Impaired active inference is responsible for aberrant social conformity, while different models rely on multiple parameters (e.g., reward sensitivity parameters, inverse decision temperature parameters) (58). This requires a multilayer model that encompasses these parameters and allows them to vary and interact in sophisticated social situations.

In light of this, a hierarchical computational framework is particularly suitable for modeling the complex structure of interaction between individuals and social groups, which transcends traditional monistic perspectives by recognizing social conformity as an intricate interplay between individuals and communities (62). Social conformity is in essence the outcome of estimation of multilevel interplay, so it benefits from computational understanding of this process (6). The core of the proposed framework relies on the dynamic predictive coding that bridges the individual and group levels, capturing the interactions between agents and their ever-evolving social environments. Through social practice, individuals update their beliefs and decide whether to comply according to prior knowledge (Figure 3).

In the first layer, social groups establish stylistic rules and norms to build bonds between individuals, which may influence prior beliefs through mutual interaction. On the other hand, in the second layer, individuals' interactions exploit social norms, and they gradually explore new subgroups to create distinctive connections. To minimize prediction errors, individuals try to precisely attribute consequences to specific social decisions, a process that often goes wrong in psychiatric patients. Coupled predictions and outcomes separately cultivate bonds between individuals and groups, providing improper guidelines of social conformity for people with psychological pathologies.

To demonstrate the effectiveness and reliability of the proposed framework, we performed an analysis of simulated data. Our hierarchical computational model, which integrates the Kuramoto model and active inference, was implemented in Python 3.12. The model simulates 2 distinct groups: neurotypical individuals (NTs) and patients with psychiatric disorders (PTs), with the latter group assumed to exhibit a reduced interaction level (63,64). Each group comprises 10 individuals, resulting in 45 possible pairs per group, and the simulation spans 200 timesteps. The Kuramoto model describes the phase evolution of coupled oscillators (65,66):

$$\frac{d\theta_i(t)}{dt} = \omega_i(t) + \sum_{j=1}^N K_{ij} * \sin(\theta_j(t) - \theta_i(t)) + \xi(t) \quad (1)$$

Here, ω_i represents the natural frequency of the oscillator, K_{ij} denotes the coupling strength between oscillators i and j , and $\theta_n(t)$ represents the phase of oscillator n at time t . The term ξ introduces Gaussian noise, adding stochasticity to the system

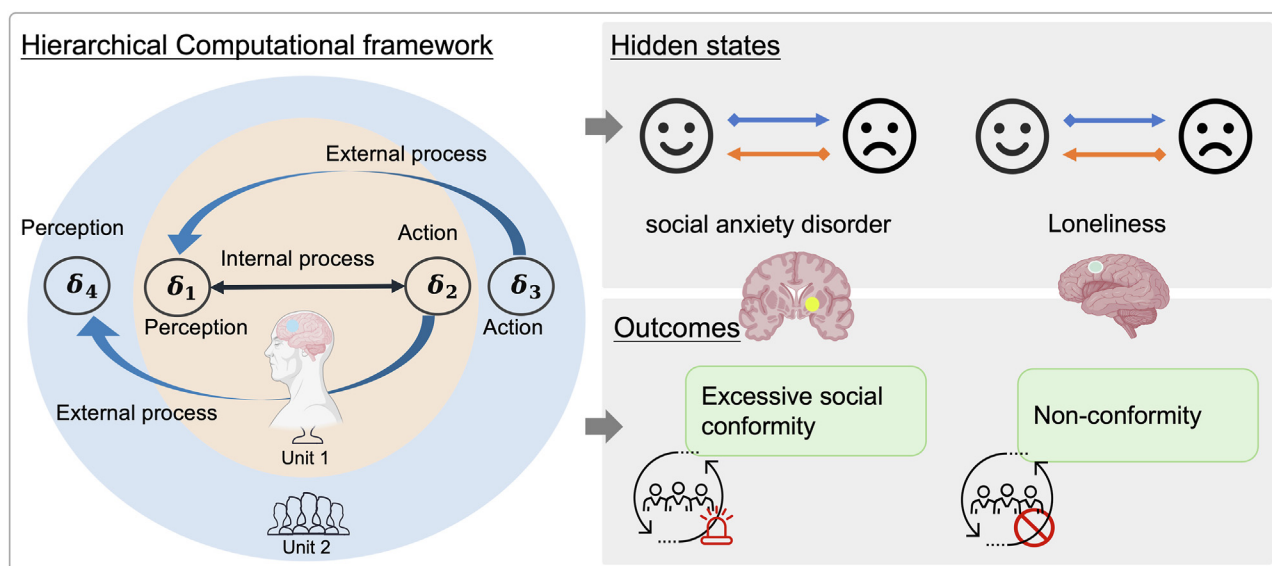


Figure 3. A hierarchical framework with social interaction, computational processes, and social conformity. A computational model of individual-group integration in social interaction can be represented as an integrated Kuramoto model with 4 oscillators. In this model, each person/group in a dyad is abstracted as a unit with 2 interconnected oscillators, reflecting their internal processes of perception and action. δ_1 = oscillator 1 (perception) in unit 1, δ_2 = oscillator 2 (action) in unit 1, δ_3 = oscillator 1 (action) in unit 2, and δ_4 = oscillator 2 (perception) in unit 2 (top). Social interaction-related information extracted from the Kuramoto model can serve as contextual information for further active inference. Active inference offers distinct interpretations of the cognitive processes that contribute to abnormal social conformity.

dynamics. The coupling strength was set higher for NTs (0.3) than for PTs (0.1), reflecting reduced interpersonal interactions in the latter group (59). The coupling strength was chosen arbitrarily, but it is consistent with previous studies on coupling during naturalistic interactions (65). The phase difference and coupling strength derived from the Kuramoto model were then used to inform the statistical model in active inference, including prediction error and prior constraint (59). In active inference, the variational free energy is often decomposed as

$$F = D_{KL}(Q(s)||P(s|o)) - E_{Q(s)}[\ln P(o|s)] \quad (2)$$

where D_{KL} is the Kullback-Leibler divergence, capturing the difference between the posterior $Q(s)$ and prior beliefs $P(s)$. The second term represents surprisal, measuring the likelihood of observations given the internal state [for more details, see (67)]. For social and interpersonal inference (integrated with the Kuramoto model), we need a simplified free energy formulation that captures 1) prediction error—the discrepancy between expected and observed synchronization, and 2) prior constraint on belief, ensuring that beliefs remain within a reasonable range based on interaction strength (i.e., K_{ij}). In this setting, a possible form of free energy F could be

$$F = \frac{1}{2} \sum_i \epsilon_{ij}^2 + \lambda (\theta_i - e^{-K_{ij}})^2 \quad (3)$$

where λ controls the prior strength (68,69). Within the free energy minimization framework (67), the belief μ_i of individual i was updated based on the prediction error ϵ_{ij} , which quantifies the deviation of the current state (an individual's phase) from the expected state (synchronized phase with another individual). λ was set to 0.5 to ensure adaptive belief updates without excessive rigidity or instability. This simplified free energy formula, which captures prediction error and prior constraint on

belief, follows active inference principles while remaining computationally efficient for simulations. Belief updates followed

$$\Delta u_i = -\alpha \epsilon_{ij} \cos(\theta_j - \theta_i) \quad (4)$$

where α is the learning rate, set higher for NTs (0.2) than for PTs (0.1, in arbitrary) to reflect their presumed differences in adaptation speeds (70,71). To quantify synchronization, we computed the phase locking value (PLV) (72):

$$PLV_{ij} = \frac{1}{N} \left| \sum_{t=1}^N \exp(i(\theta_j(t) - \theta_i(t))) \right| \quad (5)$$

PLV values close to 1 indicate strong synchronization, while values near 0 suggest weak synchronization. Belief convergence was measured as the mean absolute difference between an individual's current belief and its initial state, assessing the extent to which beliefs evolve over time. In addition, we tracked approximated free energy over time, with higher values indicating greater uncertainty and instability in social encounters. The simulations revealed that NTs achieved higher synchronization levels compared to PTs. Mean belief convergence was lower for NTs than for PTs, suggesting stronger alignment of internal beliefs in the NT group. Furthermore, free energy values remained stable over time for NTs but increased progressively for PTs, reflecting greater uncertainty in interaction for the latter. Overall, this simulation successfully illustrates that reduced interaction levels in PTs lead to weaker synchronization, maladaptive belief updating, and higher uncertainty during social encounters. These factors may contribute to a more chaotic pattern (vs. adaptive pattern) of social conformity in individuals with psychiatric disorders (Figure 4).

The integration of dynamic systems and active inference illuminates the mathematical underpinnings of cognitive and

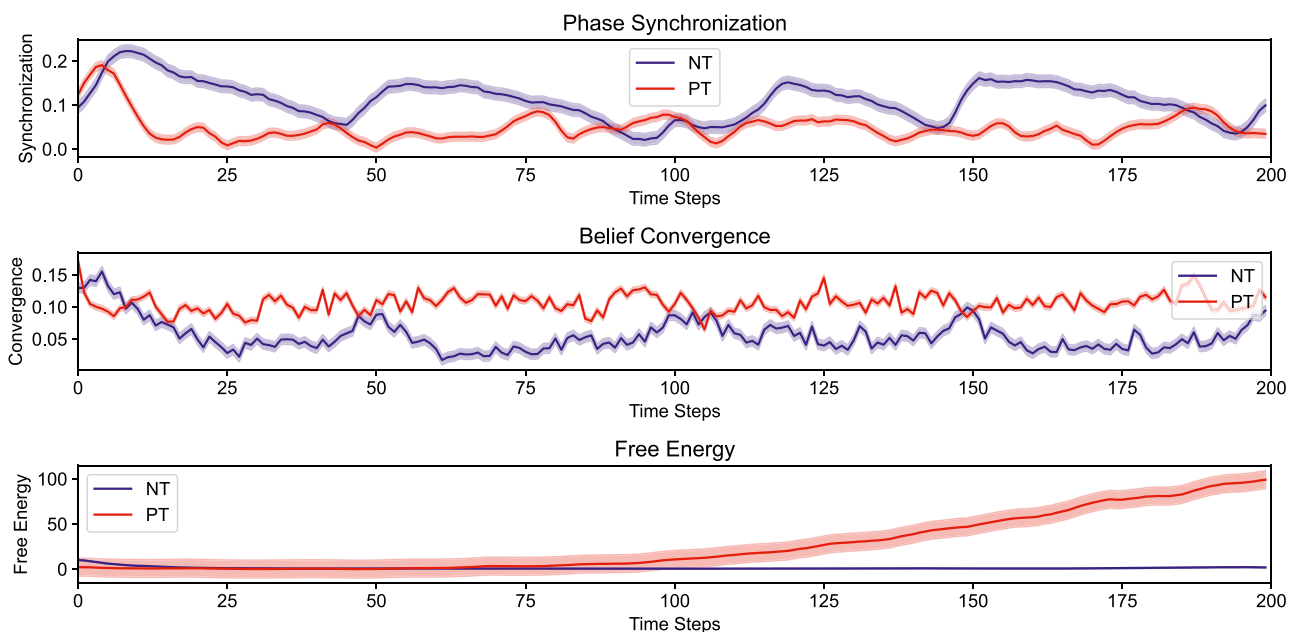


Figure 4. Overview of the simulation results. For both neurotypical (NT) and patient (PT) groups, phase synchronization, belief convergence, and free energy were estimated over time based on the hierarchical computational models. The patterns are plotted as the mean value, with shadows indicating SEMs.

behavioral conformity processes. This framework provides a valuable lens for assessing abnormal mechanisms in psychiatric populations confronted with conformity decisions, inspiring future research endeavors with a holistic and nuanced perspective.

IMPLICATIONS AND FUTURE DIRECTIONS FOR PSYCHIATRY

Advancements in understanding the neurobiological and computational mechanisms that underlie social conformity offer promising avenues for the development of precision treatments in psychiatry (Figure 5). The ability of this novel framework to tease apart the complexities of these mechanisms allows better understanding of psychiatric disorders, with utility for both scientists and clinicians. Advancing treatment will also likely necessitate a multifaceted approach that integrates targeted pharmacotherapy, personalized therapies, and neurostimulation techniques. We look forward to the future application of the proposed computational framework and its clinical translation into more effective therapeutic interventions, but there are still challenges to overcome.

A pivotal opportunity lies in the refinement of pharmacotherapies that target the reward circuitry. Medications designed to modulate dopaminergic activity in the VS have shown the potential to enhance reward sensitivity and social motivation, particularly in depression (73). Beyond traditional

psychostimulants like amphetamines, novel agents such as *TAAR1* agonists, which regulate DA and glutamate signaling, are emerging as promising antipsychotics in clinical trials (74). Their effectiveness in treating depression, OCD, and addiction underscores their potential to revolutionize psychiatric care. From another perspective, pharmaceuticals that reduce anxiety are also likely to find broad applicability in various neuropsychological disorders. Anxiolytic treatments aimed at reducing hyperactivity in the AI and posterior mPFC have the potential to alleviate excessive fear of social judgment in anxiety disorders, as exemplified by the successful combination of anakinra and lamotrigine in epilepsy patients with comorbid anxiety (75).

Personalized behavioral therapy, informed by neuroimaging techniques such as functional magnetic resonance imaging (fMRI), allows for tailored interventions based on individual neural activation patterns. This approach enhances treatment precision by addressing specific deficits underlying abnormal social conformity (11,14). Additionally, integrating cognitive behavioral therapy with virtual reality provides an immersive environment for social skills training, particularly for psychiatric populations in which traditional neurostimulation may have limitations. Other psychotherapeutic approaches including behavioral activation and motivational interviewing also support individuals to form self-affirmation and enhance social involvement (4). The proposed hierarchical computational

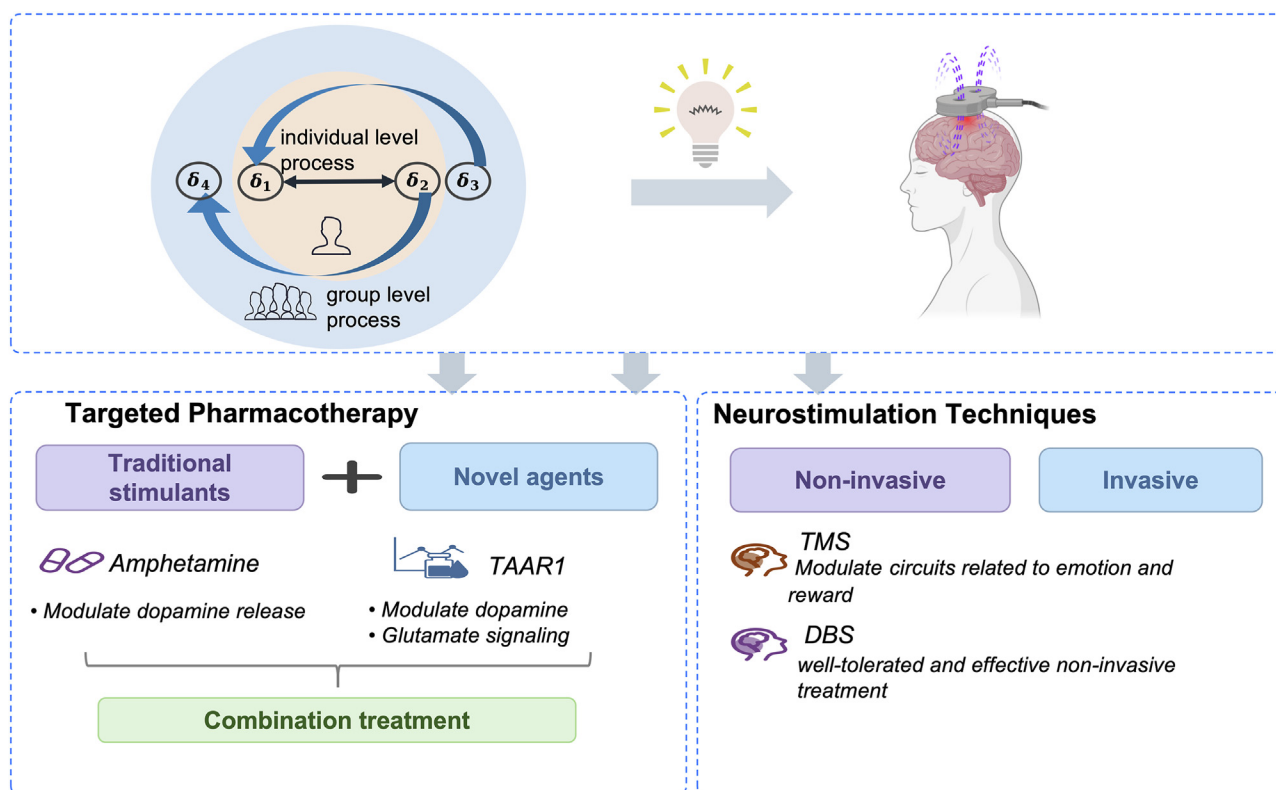


Figure 5. Future inspiration for psychiatric treatments. Top: The hierarchical computational model has implications for future psychiatric interventions. Bottom left: Effective targeted pharmacotherapy necessitates a combination of both traditional stimulants and innovative agents. Bottom right: Neurostimulation techniques, ranging from noninvasive to invasive interventions, should be tailored to the specific condition of each patient. DBS, deep brain stimulation; TMS, transcranial magnetic stimulation.

framework further supports these interventions by generating individualized predictions, which may inform targeted treatment strategies.

However, current research on social conformity predominantly relies on neuroimaging tools. The laboratory-based designs supported by neuroimaging techniques face several methodological limitations. First, fMRI and electroencephalography have limited temporal and spatial resolution, respectively, which hinder the capture of rapid neural dynamics during real-time social interactions (36). Second, laboratory tasks often lack ecological validity, failing to replicate the complexity of real-world social environments and creating gaps in the understanding of the dynamic physiological responses that underlie social conformity (3). Third, neuroimaging data alone cannot establish causal relationships between neural circuits and conformity behaviors. Even well-designed experiments struggle with clinical translation and generalization (7,28). To address these limitations, future research should integrate multimodal data (e.g., genetics, pharmacology, and behavioral metrics) with advanced computational frameworks (6).

From the perspective of modeling social conformity, active inference based on hierarchical computation promotes the stratification of psychiatric patients for diagnostic and prognostic purposes, because social conformity is considered to be an important indicator for prognosis (59). Nonetheless, little work has been dedicated to translation of the findings from computational approaches to clinical trials, which indicates a knowledge gap: Prior computations may be an untapped source of information to guide suitable treatments that could lead to positive outcomes in social functions. This theory-based model can facilitate treatment and intervention for psychiatric disorders. Under the nested structure to calculate interplay of the individual and group levels, empirical data can be put into the framework to investigate the motivation of social conformity. This enables us to uncover the hidden states of social decisions in individuals with psychiatric disorders on a computational level.

Combined with our understanding of neural circuits that are active in social conformity, the proposed framework encourages additional practice to apply neural data and foster stimulus-specific clinical models to analyze psychiatric syndromes in terms of social interaction. Moreover, the mechanistic parameters in the computational model may have implications for the stimulation parameters selected for neurostimulation techniques such as transcranial magnetic stimulation (TMS) and deep brain stimulation (DBS). Algorithmic description of the mental process of social conformity makes it possible to improve clinical conditions of psychiatric patients by selectively adjusting several parameters in interventions (58). Techniques such as TMS have demonstrated efficacy in modulating circuits related to emotion and reward in OCD and major depressive disorder by targeting the mPFC and ACC (76). Furthermore, DBS, as a well-tolerated and effective therapeutic modality, holds promise for treatment-resistant depression and other disorders characterized by reward circuitry dysfunction via

stimulating the NAcc (77,78). The algorithm parameters in our hierarchical computational network model have the potential to make targeted stimulus and outcome interpretation more explicit and reliable.

The limitations of the proposed framework mainly lie in the omission of other potential factors in terms of clinical translation. First, explicit consideration of risk, recurrence, and recovery may be incorporated into the computational framework as supportive parameters. Second, proper training should be provided for biological researchers, clinicians, psychologists, and computer scientists. Thus, we can usher in a new era of precision treatments that effectively address the complexities of psychiatric disorders, thereby fostering individual well-being and societal harmony.

CONCLUSIONS

The reward, punishment, and affective circuits elucidate the neurobiological basis of social conformity. Advanced computational psychiatry, leveraging mathematical models, reveals deviant decision making. Integration with neuroimaging enables a therapeutic continuum from targeted pharmacotherapy to neurostimulation, offering personalized approaches. Critically, the parameters in the hierarchical computational framework may serve as quantifiable biomarkers for predicting maladaptive conformity, thereby guiding neuromodulation interventions. While applications are limited, this approach holds promise for advancing treatments across mental disorders. Future studies should validate these predictions through multimodal clinical trials, combining neuroimaging, genetics, and behavioral monitoring.

ACKNOWLEDGMENTS AND DISCLOSURES

This work was supported by the National Science and Technology Innovation 2030 Major Project of China (Grant No. 2021ZD0203900 [to DZ]), National Natural Science Foundation of China (Grant No. 82422029 [to DZ]), the Fundamental Research Funds for the Central Universities (Grant No. YG2025ZD07 [to DZ]), the Science and Technology Commission of Shanghai Municipality (Grant Nos. 224Y22800200 and 22QA1407900 [to DZ]), National Natural Science Foundation of China (Grant No. 82271530 [to DZ]), innovation teams of high-level universities in Shanghai, and the Scientific Research and Innovation Team of Liaoning Normal University (Grant No. 24TD004 [to DZ]), and the Zhejiang Provincial Natural Science Foundation of China (Grant No. LMS25C090002 [to YP]). The funders had no role in the performance of the study or the decision to pursue publication.

DZ and YP conceived and designed the study. YL drafted the initial article. DZ and YP revised and discussed the final version. All authors read and approved the final article.

We thank Ruinan Zhao and Qianhao Yu for their help with the initial manuscript.

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

ARTICLE INFORMATION

From the Shanghai Key Laboratory of Psychotic Disorders, Brain Health Institute, National Center for Mental Disorders, Shanghai Mental Health Center, Shanghai Jiao Tong University School of Medicine and School of Psychology, Shanghai, China (YL, DZ); Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China (YP); and State Key Laboratory of Brain-Machine Intelligence, Zhejiang University, Hangzhou, China (YP).

Address correspondence to Di Zhao, Ph.D., at dizhao_sunset@sjtu.edu.cn, or Yafeng Pan, Ph.D., at yafeng.pan@zju.edu.cn.

Received Dec 9, 2024; revised Apr 17, 2025; accepted May 14, 2025.

REFERENCES

- Cialdini RB, Goldstein NJ (2004): Social influence: Compliance and conformity. *Annu Rev Psychol* 55:591–621.
- Sherif M (1936): *The Psychology of Social Norms*. New York, NY: Harper & Brothers.
- Asch SE (1951): Effects of group pressure upon the modification and distortion of judgements. In: Guetzkow HS, editor. *Groups, Leadership and Men*. Pittsburgh, PA: Carnegie Press, 177–190.
- Hall AF, Browning M, Huys QJM (2024): The computational structure of consummatory anhedonia. *Trends Cogn Sci* 28:541–553.
- Eslinger PJ, Anders S, Ballarín T, Boutros S, Krach S, Mayer AV, et al. (2021): The neuroscience of social feelings: Mechanisms of adaptive social functioning. *Neurosci Biobehav Rev* 128:592–620.
- Maia TV, Huys QJM, Frank MJ (2017): Theory-based computational psychiatry. *Biol Psychiatry* 82:382–384.
- Zhao C, Liu J, Gong P, Hu J, Zhou X (2016): Investigating the genetic basis of social conformity: The role of the dopamine receptor 3 (DRD3) gene. *Neuropsychobiology* 74:32–40.
- Dominguez DJF, Taing SA, Molenberghs P (2016): Why do some find it hard to disagree? An fMRI study. *Front Hum Neurosci* 9:718.
- Wu H, Luo Y, Feng C (2016): Neural signatures of social conformity: A coordinate-based activation likelihood estimation meta-analysis of functional brain imaging studies. *Neurosci Biobehav Rev* 71:101–111.
- Martins D, Rademacher L, Gabay AS, Taylor R, Richey JA, Smith DV, et al. (2021): Mapping social reward and punishment processing in the human brain: A voxel-based meta-analysis of neuroimaging findings using the social incentive delay task. *Neurosci Biobehav Rev* 122:1–17.
- Campbell-Meiklejohn DK, Simonsen A, Jensen M, Wohler V, Gjerloff T, Scheel-Kruger J, et al. (2012): Modulation of social influence by methylphenidate. *Neuropsychopharmacology* 37:1517–1525.
- Zaki J, Schirmer J, Mitchell JP (2011): Social influence modulates the neural computation of value. *Psychol Sci* 22:894–900.
- Chung D, Christopoulos GI, King-Casas B, Ball SB, Chiu PH (2015): Social signals of safety and risk confer utility and have asymmetric effects on observers' choices. *Nat Neurosci* 18:912–916.
- Lee H, Chung D (2022): Characterization of the core determinants of social influence from a computational and cognitive perspective. *Front Psychiatry* 13:846535.
- Moll J, Krueger F, Zahn R, Pardini M, de Oliveira-Souza R, Grafman J (2006): Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci U S A* 103:15623–15628.
- Olsson A, Knapska E, Lindström B (2020): The neural and computational systems of social learning. *Nat Rev Neurosci* 21:197–212.
- Zhang L, Gläscher J (2020): A brain network supporting social influences in human decision-making. *Sci Adv* 6:eabb4159.
- Speranza L, di Porzio U, Viggiano D, de Donato A, Volpicelli F (2021): Dopamine: The neuromodulator of long-term synaptic plasticity, reward and movement control. *Cells* 10:735.
- Montague PR, Dolan RJ, Friston KJ, Dayan P (2012): Computational psychiatry. *Trends Cogn Sci* 16:72–80.
- Cervenka S, Gustavsson JP, Halldin C, Farde L (2010): Association between striatal and extrastriatal dopamine D2-receptor binding and social desirability. *Neuroimage* 50:323–328.
- Yafai A-F, Verrier D, Reidy L (2014): Social conformity and autism spectrum disorder: A child-friendly take on a classic study. *Autism* 18:1007–1013.
- Shamay-Tsoory SG, Saporta N, Marton-Alper IZ, Gvirts HZ (2019): Herding brains: A core neural mechanism for social alignment. *Trends Cogn Sci* 23:174–186.
- von der Lüh T, Manera V, Barisic I, Becchio C, Vogetley K, Schilbach L (2016): Interpersonal predictive coding, not action perception, is impaired in autism. *Philos Trans R Soc Lond B Biol Sci* 371:20150373.
- Kanai R, Bahrami B, Duchaine B, Janik A, Banissy MJ, Rees G (2012): Brain structure links loneliness to social perception. *Curr Biol* 22:1975–1979.
- Pauli R, Lockwood PL (2023): The computational psychiatry of anti-social behaviour and psychopathy. *Neurosci Biobehav Rev* 145:104995.
- Wasylshyn N, Hemenway Falk B, Garcia JO, Cascio CN, O'Donnell MB, Bingham CR, et al. (2018): Global brain dynamics during social exclusion predict subsequent behavioral conformity. *Soc Cogn Affect Neurosci* 13:182–191.
- Eisenberger NI, Lieberman MD, Williams KD (2003): Does rejection hurt? An fMRI study of social exclusion. *Science* 302:290–292.
- Huys QJM, Browning M, Paulus MP, Frank MJ (2021): Advances in the computational understanding of mental illness. *Neuropsychopharmacology* 46:3–19.
- Klucharev V, Munneke MAM, Smidts A, Fernández G (2011): Down-regulation of the posterior medial frontal cortex prevents social conformity. *J Neurosci* 31:11934–11940.
- Izuma K, Adolphs R (2013): Social manipulation of preference in the human brain. *Neuron* 78:563–573.
- Chang LJ, Sanfey AG (2013): Great expectations: Neural computations underlying the use of social norms in decision-making. *Soc Cogn Affect Neurosci* 8:277–284.
- Lieberman MD, Eisenberger NI (2015): The dorsal anterior cingulate cortex is selective for pain: Results from large-scale reverse inference. *Proc Natl Acad Sci U S A* 112:15250–15255.
- Wittfoth M, Schröder C, Schardt DM, Dengler R, Heinze HJ, Kotz SA (2010): On emotional conflict: Interference resolution of happy and angry prosody reveals valence-specific effects. *Cereb Cortex* 20:383–392.
- Liuzza MT, Macaluso E, Chiesa PA, Lingardi V, Aglioti SM (2019): An fMRI study on the neural correlates of social conformity to a sexual minority. *Sci Rep* 9:4691.
- Saxe R, Wexler A (2005): Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia* 43:1391–1399.
- Kahn T, Mahmoodi A, Nili H, Bang D, Mehring C, Bahrami B (2022): Distinct neurocomputational mechanisms support informational and socially normative conformity. *PLoS Biol* 20:e3001565.
- Volkow ND, Morales M (2015): The brain on drugs: From reward to addiction. *Cell* 162:712–725.
- Koob GF, Le Moal M (2005): Plasticity of reward neurocircuitry and the 'dark side' of drug addiction. *Nat Neurosci* 8:1442–1444.
- Chechko N, Kellermann T, Augustin M, Zvyagintsev M, Schneider F, Habel U (2016): Disorder-specific characteristics of borderline personality disorder with co-occurring depression and its comparison with major depression: An fMRI study with emotional interference task. *Neuroimage Clin* 12:517–525.
- Kohn N, Eickhoff SB, Scheller M, Laird AR, Fox PT, Habel U (2014): Neural network of cognitive emotion regulation - An ALE meta-analysis and MACM analysis. *Neuroimage* 87:345–355.
- Liu Y, Hanna GL, Carrasco M, Gehring WJ, Fitzgerald KD (2014): Altered relationship between electrophysiological response to errors and gray matter volumes in an extended network for error-processing in pediatric obsessive-compulsive disorder. *Hum Brain Mapp* 35:1143–1153.
- Ursu S, Stenger VA, Shear MK, Jones MR, Carter CS (2003): Overactive action monitoring in obsessive-compulsive disorder: Evidence from functional magnetic resonance imaging. *Psychol Sci* 14:347–353.
- Maltby N, Tolin DF, Worhunsky P, O'Keefe TM, Kiehl KA (2005): Dysfunctional action monitoring hyperactivates frontal-striatal circuits in obsessive-compulsive disorder: An event-related fMRI study. *Neuroimage* 24:495–503.
- Fitzgerald KD, Welsh RC, Gehring WJ, Abelson JL, Himle JA, Liberzon I, Taylor SF (2005): Error-related hyperactivity of the anterior cingulate cortex in obsessive-compulsive disorder. *Biol Psychiatry* 57:287–294.

45. Fitzgerald KD, Liu Y, Johnson TD, Moser JS, Marsh R, Hanna GL, Taylor SF (2018): Development of posterior medial frontal cortex function in pediatric obsessive-compulsive disorder. *J Am Acad Child Adolesc Psychiatry* 57:397–406.
46. Cisler JM, Koster EHW (2010): Mechanisms of attentional biases towards threat in anxiety disorders: An integrative review. *Clin Psychol Rev* 30:203–216.
47. Takami K, Haruno M (2019): Behavioral and functional connectivity basis for peer-influenced bystander participation in bullying. *Soc Cogn Affect Neurosci* 14:23–33.
48. Browning M, Paulus M, Huys QJM (2023): What is computational psychiatry good for? *Biol Psychiatry* 93:658–660.
49. Kelly C, Toro R, Di Martino A, Cox CL, Bellec P, Castellanos FX, Milham MP (2012): A convergent functional architecture of the insula emerges across imaging modalities. *Neuroimage* 61:1129–1142.
50. Dyck M, Loughhead J, Kellermann T, Boers F, Gur RC, Mathiak K (2011): Cognitive versus automatic mechanisms of mood induction differentially activate left and right amygdala. *Neuroimage* 54:2503–2513.
51. Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schweren LSJ, *et al.* (2017): Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: A cross-sectional mega-analysis. *Lancet Psychiatry* 4:310–319.
52. Xu W, Li YH, Tan BP, Luo XJ, Xiao L, Zheng XG, *et al.* (2012): Inhibition of the acquisition of conditioned place aversion by dopaminergic lesions of the central nucleus of the amygdala in morphine-treated rats. *Physiol Res* 61:437–442.
53. Bloodgood DW, Sugam JA, Holmes A, Kash TL (2018): Fear extinction requires infralimbic cortex projections to the basolateral amygdala. *Transl Psychiatry* 8:60.
54. Wolf RC, Hildebrandt V, Schmitgen MM, Pycha R, Kirchner E, Macina C, *et al.* (2020): Aberrant gray matter volume and cortical surface in paranoid-type delusional disorder. *Neuropsychobiology* 79:335–344.
55. Ganella DE, Barendse MEA, Kim JH, Whittle S (2017): Prefrontal-amygdala connectivity and state anxiety during fear extinction recall in adolescents. *Front Hum Neurosci* 11:587.
56. Huys QJM (2020): Computational Psychiatry Series. *Biol Psychiatry Cogn Neurosci Neuroimaging* 5:835–836.
57. Wang X-J, Krystal JH (2014): Computational psychiatry. *Neuron* 84:638–654.
58. Yip SW, Barch DM, Chase HW, Flagel S, Huys QJM, Konova AB, *et al.* (2023): From computation to clinic. *Biol Psychiatry Glob Open Sci* 3:319–328.
59. Pan Y, Wen Y, Jin J, Chen J (2023): The interpersonal computational psychiatry of social coordination in schizophrenia. *Lancet Psychiatry* 10:801–808.
60. Constantino JN, Todd RD (2000): Genetic structure of reciprocal social behavior. *Am J Psychiatry* 157:2043–2045.
61. Friston K, Frith C (2015): A duet for one. *Conscious Cogn* 36:390–405.
62. Cui X, Bryant DM, Reiss AL (2012): NIRS-based hyperscanning reveals increased interpersonal coherence in superior frontal cortex during cooperation. *Neuroimage* 59:2430–2437.
63. Schilbach L (2016): Towards a second-person neuropsychiatry. *Philos Trans R Soc B Biol Sci* 371:20150081.
64. Pan Y, Wen Y, Wang Y-J, Schilbach L, Chen J (2023): Interpersonal coordination in schizophrenia: A concise update on paradigms, computations, and neuroimaging findings. *Psychoradiology* 3:kkad002.
65. Heggli OA, Cabral J, Konvalinka I, Vuust P, Kringelbach ML (2019): A Kuramoto model of self-other integration across interpersonal synchronization strategies. *PLoS Comput Biol* 15:e1007422.
66. Wang Y-J, Wen Y, Zheng L, Chen J, Lin Z, Pan Y (2025): A computational and multi-brain signature for aberrant social coordination in schizophrenia. *Prog Neuropsychopharmacol Biol Psychiatry* 136:111225.
67. Friston K (2010): The free-energy principle: A unified brain theory? *Nat Rev Neurosci* 11:127–138.
68. Friston K, Kiebel S (2009): Predictive coding under the free-energy principle. *Philos Trans R Soc B Biol Sci* 364:1211–1221.
69. Bogacz R (2017): A tutorial on the free-energy framework for modelling perception and learning. *J Math Psychol* 76:198–211.
70. 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* 79:313–322.
71. Henco L, Diaconescu AO, Lahnakoski JM, Brandt M-L, Hörmann S, Hennings J, *et al.* (2020): Aberrant computational mechanisms of social learning and decision-making in schizophrenia and borderline personality disorder. *PLoS Comput Biol* 16:e1008162.
72. Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999): Measuring phase synchrony in brain signals. *Hum Brain Mapp* 8:194–208.
73. Couppis MH, Kennedy CH (2008): The rewarding effect of aggression is reduced by nucleus accumbens dopamine receptor antagonism in mice. *Psychopharmacology (Berl)* 197:449–456.
74. Zhukov IS, Karpova IV, Krotova NA, Tissen IY, Demin KA, Shabanov PD, *et al.* (2022): Enhanced aggression, reduced self-grooming behavior and altered 5-HT regulation in the frontal cortex in mice lacking trace amine-associated receptor 1 (TAAR1). *Int J Mol Sci* 23:14066.
75. Zubareva OE, Sinyak DS, Kalita AD, Griflyuk AV, Diespirov GP, Postnikova TY, Zaitsev AV (2023): Antiepileptogenic effects of anakinra, lamotrigine and their combination in a lithium-pilocarpine model of temporal lobe epilepsy in rats. *Int J Mol Sci* 24:15400.
76. Zhang D, Ao X, Zheng Z, Shen J, Zhang Y, Gu R (2022): Modulating social feedback processing by deep TMS targeting the medial prefrontal cortex: Behavioral and electrophysiological manifestations. *Neuroimage* 250:118967.
77. Zhou H, Zhu J, Jia J, Xiang W, Peng H, Zhang Y, *et al.* (2022): The antidepressant effect of nucleus accumbens deep brain stimulation is mediated by parvalbumin-positive interneurons in the dorsal dentate gyrus. *Neurobiol Stress* 21:100492.
78. Yan N, Chen N, Zhu H, Zhang J, Sim M, Ma Y, Wang W (2013): High-frequency stimulation of nucleus accumbens changes in dopaminergic reward circuit. *PLoS One* 8:e79318.