A computational account of borderline personality disorder: impaired predictive learning about self and others through bodily simulation

Sarah K. Fineberg[1], Matthew Steinfeld[1], Judson A. Brewer[1,2][†] and Philip R. Corlett[1][†]*

[1] Department of Psychiatry, Yale University, New Haven, CT, USA
[2] Center for Mindfulness, University of Massachussetts Medical School, Worcester, MA, USA

***Correspondence:**
Philip R. Corlett, Department of Psychiatry, Connecticut Mental Health Center, Yale University School of Medicine, 34 Park Street, New Haven, CT 06519, USA
e-mail: philip.corlett@yale.edu

[†]These authors co-directed the project.

Social dysfunction is a prominent and disabling aspect of borderline personality disorder. We reconsider traditional explanations for this problem, especially early disruption in the way an infant feels physical care from its mother, in terms of recent developments in computational psychiatry. In particular, social learning may depend on reinforcement learning though embodied simulations. Such modeling involves calculations based on structures outside the brain such as face and hands, calculations on one's own body that are used to make inferences about others. We discuss ways to test the role of embodied simulation in BPD and potential implications for treatment.

Keywords: associative learning, computational psychiatry, borderline personality disorder, embodied simulation, attachment

INTRODUCTION

Social dysfunction in borderline personality disorder (BPD) is profound and persistent (1–3). In her memoir about BPD, Merri Lisa Johnson writes that having relationships was like "bleeding out," evoking both the blurring of boundaries between herself and others and the bodily seriousness of the problem (4). Psychological formulations of BPD have also focused on the body.

Current theories describing BPD are complex and treatment paradigms are time intensive and expensive. Given the large psychological and healthcare burden of BPD, theoretical and treatment innovations are essential (5–7).

We aim here to present a hypothesis: that people with BPD have deficits in embodied simulation, which is a way of computing information about others that uses data from ones' own body (see further explanation below). We discuss the evidence for this position, possible experimental approaches to testing it, and potential implications for treatment.

COMPUTATIONAL PSYCHIATRY

We outline our hypothesis using a computational psychiatry approach. Given that this is a new but growing field, it is worth outlining what we mean by computational psychiatry. Computational psychiatry involves applying advances in computational neuroscience to psychiatry (8–11). More specifically, we think of computational psychiatry as an extension of cognitive neuropsychiatry (12). Cognitive neuropsychiatry involves specifying a cognitive system or process with relevance to a psychiatric symptom and examining how aberrations of that system or process

might give rise to the symptom. For example, monitoring of inner speech may be relevant to auditory verbal hallucinations (13). Computational psychiatry develops this approach more formally, specifying mathematically how a problem is solved and then describing and quantifying how mental symptoms may arise in departures from that formalization. Those departures are delineated in terms of model parameters and the magnitude of change in those parameters required to engender mental symptoms.

Computational psychiatry may be useful in generating and testing hypotheses about mental illness. Furthermore, it may be possible to use computational psychiatry to generate and apply more data-driven nosologies for diagnosis and treatment. For example, by specifying which processes (neural, psychological, and behavioral) might be driving an individual's problems, we can use computational psychiatry to allocate them to particular treatment strategies. More deeply, it may be possible to use computational neuroscience to bridge mental symptoms with neural processes. We make this assertion by considering David Marr's levels of analysis (14), which we think may be key to achieving what Fodor thought to be impossible (15), a consilient understanding of cognition and its implementation in the brain.

After David Marr and Thomas Poggio, we believe that the computational problems solved by a system (e.g., a brain) (14) and their malfunctions (e.g., in mental illnesses like BPD) can be described at three levels of analysis:

(1) Computational – what are the representations being used by the system to solve the problem?

(2) Algorithmic – how are those representations combined and manipulated to solve the problem?

(3) Implementational – how does the hardware (neurons, circuits, brain, and body) implement the computations to achieve the end?

Thus far, most computational psychiatry has been algorithmic. It has used that level of analysis to discover what is happening at the implementational (neural) level and how that implementation may be disrupted in illness. We find this to be a worthy and efficient approach. But it is not the only approach. It is also, we feel, not the only way that a computational framework might be useful to psychiatry.

Here, we begin at the computational level (14), roughing out the problem at hand like a sculptor (16) – seeing whether there is any use for a computational explanation of BPD at the representational level. We think the utility of this approach is at least twofold: first, we can generate a language to explain to patients what their problems are and how they might be dealt with and second, we can generate testable hypotheses that will increase our understanding of BPD and hence facilitate better treatment.

As we progress in our work, we aim to fill in the finer and finer detail.

We believe the computational problem with BPD involves predicting the intentions of others using our own experiences as a template. We do not present any formal simulations nor any analysis of behavioral data in these computational terms – this will be the topic of our future work. We aim for an algorithmic, equation driven, explanation for the symptoms of BPD. This may be useful clinically, since parameter values from the computational model may be useful in clustering groups of patients into pathophysiological categories (10, 17, 18). Such clustering may be a useful therapeutic guide toward treatments that addressed underlying pathophysiology.

In this hypothesis and theory article, we will state our predictions. Our subsequent work will test those predictions.
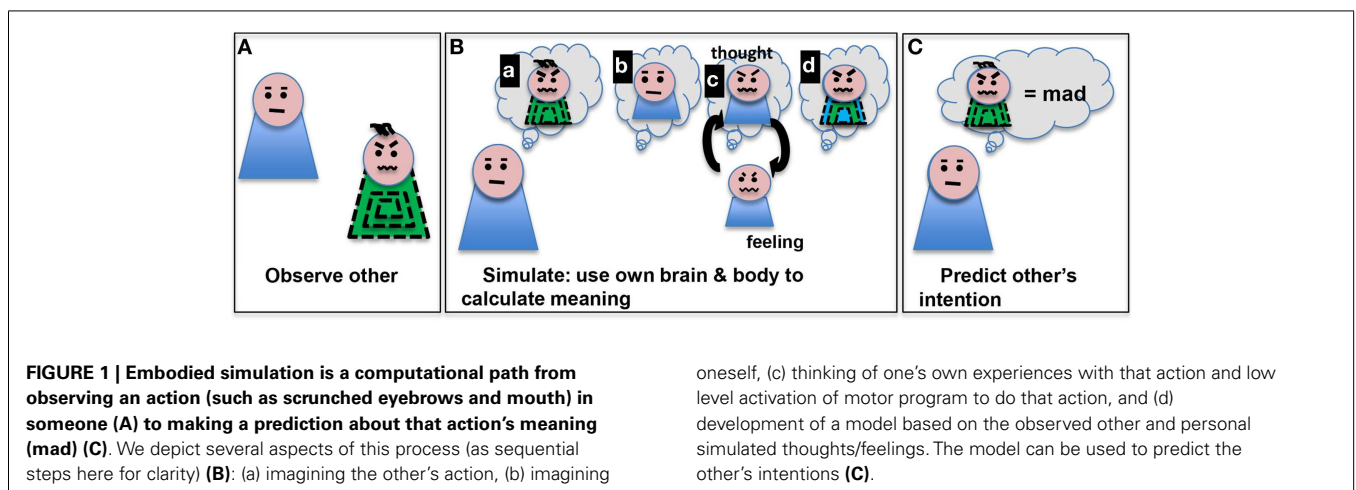
BPD AND THE BODY

Many have ascribed BPD pathology to disturbance in early attachment relationships (19, 20), specifically disruptions between mother and infant occurring at the skin surface, which regulate bodily needs such as being nourished and soothed (19). Current diagnostic criteria for BPD also emphasize the physical self: feelings of (physical) emptiness, dissociation (feeling physically apart from the current situation), self-harm, and suicidality. Studies of tactile experience in BPD have found decreased pain sensitivity (21–24), and change in discussion of scars (25) and tattoos (26). Psychotherapies for BPD already engage the body: DBT begins with physical safety assurances and continues with body-grounded mindfulness practices (27), and schema-focused therapy asks patients to trade chairs to physically take another's perspective (28).

EMBODIED SIMULATION THEORY

In a ground-breaking 1996 paper, Vittorio Gallese and colleagues described the observation that certain neurons in the macaque area F5 (ventral pre-motor cortex) are activated both by doing an action and by observing someone else do the same action (29). They have termed these cells "mirror neurons." Though single cells with mirror properties have not been much studied in humans (30), a system of brain regions with mirror properties have been mapped (31). These include pre-motor cortex, supplementary motor areas, primary somatosensory cortex, inferior parietal cortex, anterior cingulate cortex, and insula (31). Experiments transiently blocking mirror-neuron regions in human with transcranial magnetic stimulation (TMS) have shown deficits in mirroring activity (32, 33). These observations led to an embodied simulation explanation of the roles these mirror-neurons might play in cognition (34). This explanation suggests two important things (see **Figure 1**); first, observing others performing actions engages our own representation of those actions (mirroring) and second, we make inferences of others intentions using that evoked representation of our own action. The embodied simulation hypothesis takes an "extended mind" position (35): that the mind does computations in the brain, body, and perhaps the near environment. To make these inferences, we simulate what our own intention would have been, given their action kinematics and the local context (36). Hence, we use our representation of bodily self to make predictions about others, based on our own beliefs about intention, agency, and responsibility.

FIGURE 1 | Embodied simulation is a computational path from observing an action (such as scrunched eyebrows and mouth) in someone (A) to making a prediction about that action's meaning (mad) (C). We depict several aspects of this process (as sequential steps here for clarity) **(B)**: (a) imagining the other's action, (b) imagining oneself, (c) thinking of one's own experiences with that action and low level activation of motor program to do that action, and (d) development of a model based on the observed other and personal simulated thoughts/feelings. The model can be used to predict the other's intentions **(C)**.

Of course, mirroring in neural activity may be a product of simpler associative learning mechanisms (37). We focus here on the possibility that these learning mechanisms may have gone awry in BPD and how that might manifest in terms of disrupted inference about ones' body and other people.

Learning environmental contingencies (cues that predict rewards) and social contingencies (whether to trust someone's advice) can be described by the same mathematical models (38). Causal model theories are an important new direction in cognitive science, based in Bayesianism. Thomas Bayes' doctrine of probabilities formalizes reasoning about data using hypotheses and captures the probabilistic nature of many of the tasks faced by organisms. Organisms must predict their environments and respond appropriately. Thus, the brain forms and maintains a set of prior expectancies (predictive associations) about incoming stimuli in order to minimize uncertainty about subsequent inputs. This prediction error minimization is evident in hierarchical neuroanatomy down to the level of single neurons (39, 40).

The development of associationism has highlighted the value of a reductionist approach to the mind and brain to explain complex cognitive processes. As we predict our world and coordinate adaptive responses, prediction errors generated in response to a maladaptive environment can lead to strongly learned and difficult to change associations that persist into new circumstances. Also, people who erroneously signal prediction errors can learn strong (sometimes maladaptive) associations that may manifest as symptoms.

These predictive models are likely developed at least in part by embodied simulations (41, 42): we make inferences regarding other people's intentions by generating a model of our own intentions in similar contexts and using that model to predict what other people will do (36). Object relations theorists describe these predictions about new people in terms of referencing, or perhaps even overlaying, known models of people ("objects" developed through early experience) onto novel people (20). Associative learning theory agrees that social predictions would be based upon past experiences. If those predictions fail, healthy adults can update their beliefs into a more accurate model. It may be that the social difficulties that attend psychiatric illness represent difficulty with this process of making, testing, and updating accurate social models.

Although others have raised concern about whether embodied simulation is used just for mimicry, for some form of understanding others, or for mentalizing in the sense of being able to describe another's state of mind [for reviews of this issue, please see Ref. (35, 43)], leaving this debate unresolved for the time being, we can, however, assume that any of these uses of embodied simulation are important for social cognition, and that therefore deficits in embodied simulation do impact on real world social functioning.

COMPUTATIONAL MODELS OF LEARNING: RELEVANCE TO BPD

Computational models of learning distinguish between two neuro-computational systems that can control behavior (44). First, there is a model-free system that learns and caches the values of being in particular states. In this system, behavior is guided toward the choices that lead to subsequent states with the highest cached value. This system is robust because it is computationally simple, but it is insensitive to sudden changes in those values (e.g., if a food source is poisoned or a relationship becomes toxic). This cached value system has been associated with the function of the dorsolateral striatum via human imaging studies, recording and lesion studies in rodents (44). On the other hand, there is a computationally intensive Bayesian tree-search mechanism that constructs model of the environmental contingencies and explores those models to arrive at the best course of action. This system has been related to the functioning of the pre-frontal cortex (44). Behavioral control is ceded to the system that is most certain about the next option to take. In a stable predictable world, this is most often the computationally efficient striatal habit system.

Tony Dickinson and Bernard Balleine have proposed an associative-cybernetic model of learning that foreshadowed this computational framework. The associations between cues, rewards, and actions we highlighted above may well be the representational basis for model-based and model-free learning (45, 46).

Aberrant association as a result of inappropriate social interaction and feedback will bias the competition between these systems and the simple, robust habit system will be favored. We argue that when model-based mentalizing fails people with BPD learn inappropriate (but computationally efficient) social habits. For example, someone with BPD might have learned that other people are unpredictable and untrustworthy, and they would find it difficult to update this belief, even in the face of confounding experience.

Fonagy has argued that in BPD, genetic and early environmental factors undermine mentalized affectivity – the ability to represent either people's emotional states (a second order belief task). The stress and discomfort associated with this disruption (according to Fonagy) disrupts the development of cognitive control (47). It also contributes to poor attachment and an interactive cycle of further weakening attachment (since poor affect-recognition can make sensitive caregiving more challenging). Disorganized attachment also disrupts self-concept, creating incoherence, and splitting that is hard to manage and leads to frantic attempts to avoid abandonment, as well as the characteristic intense pattern of interpersonal relationships that escalate rapidly (47).

Fonagy points to the importance of the opioid and dopamine systems in developing attachment (47). He highlights the work of Jaak Panksepp in identifying a common neurobiology for mother–infant, infant–mother, and romantic attachments that involve opioids expressed in the frontostriatal dopaminergic circuitry (48). This same circuitry is involved in belief-based inference – in particular, that circuitry is engaged when evidence violates expectation (49) and in first-time mothers when they see an image of their own infant. People with disorganized attachment have a sensitized stress response (47) and are less likely to engage this inference circuitry and mentalizing processes when viewing pictures of their own child in distress (50, 51).

We posit that the balance between model-based and model-free reinforcement learning, mediated by uncertainty-based competition (44) is instructive here. Due to deficits in dopamine and opioid signaling (resultant from genetic and environmental insults) BPD individuals have attachment issues. Because of

these issues, they find it hard to learn a self-concept and to use that self-concept to mentalize (using the more computationally intensive and uncertainty sensitive model-based learning system in pre-frontal cortex). This generates stress that biases processing even more in favor of the habitual model-free controller. In the parlance of Regina Pally, through the model-free system, the individual learns a pattern of predictive responses in infancy that are recapitulated later in life (52, 53). The enhanced role of striatally mediated, relatively inflexible model-free responding is also consistent with the increased susceptibility to substance abuse in this population – since addictive substances also hijack this brain and behavioral circuitry.

Habitual model-free responding is enhanced with decreased opioid signaling in PFC (54) or with increased striatal dopamine (55) signaling in experimental animals. We believe these two processes are candidate pathophysiological markers in BPD and they highlight potential therapeutic avenues for future work.

Bringing these ideas back to our original predictive coding scheme – the control of behavior can be model-based (flexible and representationally rich) or model free (inflexible and representationally lean). Which system controls behavior depends on the systems' relative degrees of certainty about upcoming decisions (56). Within the predictive coding scheme that certainty is conveyed as precision (57, 58), the degree of confidence, we have in prediction errors (mismatches between expectation and experience). Dopamine enhances precision in this scheme, much like the uncertainty that biases instrumental control (trade-off between the two control systems – model based vs. model free).

HYPOTHESIS: BPD SOCIAL DYSFUNCTION RELATES TO DIFFICULTY WITH EMBODIED SIMULATION

In addition to feeling personally disembodied, as described in early psychological theory and the DSM, people with BPD may also have difficulty imagining embodied others – using the flexible, model-based pre-frontal tree-search system.

We believe that BPD pathology may arise from a deficit in representing bodily self. It may also involve a dysfunction in associating our action representations with those of others that we observe (59). People with BPD may have difficulty inverting representations of personal actions to make correct inferences about others (36). These predictions can be tested in a computational psychiatry framework that combines theoretical predictions with neural and behavioral data. It may be that each of these dysfunctions represents an endophenotype or sub-population of patients who meet BPD diagnostic criteria. Knowing more about the mechanisms that drive their particular dysfunction will allow us to devise targeted therapeutic approaches and allocate individuals toward those approaches based on their particular difficulties.

We predict that the diminished subjective sense of embodied self in BPD is associated with decreased use of embodied predictive simulations to generate beliefs about others, and decreased use of prediction errors to update those beliefs. Without the ability to incorporate new information about other people, social models may be applied inflexibly, leading to confusing and unstable interpersonal interactions.

EXPERIMENTAL METHODS: LINKING PSYCHOLOGICAL THEORY TO SOCIAL LEARNING AND EMBODIED SIMULATION IN BPD

Combining neuroscience theory (including mirror-neuron theory) and psychoanalytic theory is not new. To pick just a few among many examples, Nicola Diamond considers Winnicott's mirroring between mother and baby against the phantom body of Ramachandran (60), Vittorio Gallese considers mirror neurons as a potential substrate for developing interpersonal attunement (61), and Shantel Ehrenberg relates the process of learning dance movements in the studio to mirroring oneself and others in the frameworks of Lacan and the mirror-neuron system (62).

Nonetheless, these ideas are not widespread on the ground among BPD clinicians and researchers. The ability to enact social simulations has not been directly assayed in BPD patients, but several related functions have been studied. Zhao *et al.* have reviewed recent work describing self in mental illness, including BPD (63). Women with BPD had difficulty in giving detailed accounts of their own lives, and this correlated with difficulty figuring out social problems (64). Another study found that adolescents with BPD completing a social task "hypermentalize"; they indulge in too many hypotheses about others' intentions with counter-productive results. They are ultimately unable to make *useful and use-able* hypotheses (65). BPD subjects also had more trouble than did control subjects with cooperating in an economic game where sharing might be advantageous (66). Tasks that engage social processes from recognizing facial emotions (67) to looking at (versus intentionally distancing from) negative social images (68) to activating empathy (69) all resulted in *decreased* activity of mirror-neuron regions in BPD versus control subjects. We might even speculate that dissociation could represent failure to be able to simulate very upsetting situations *within oneself*, perhaps giving rise to the subjective sense of *not being in one's own body.*

We predict that a more globally effective intervention would engage the brain regions that integrate new data into preexisting models. For example, dopamine levels in pre-frontal cortex can increase model-based reinforcement learning in rodents (70). Also, Tim Behrens, Laurence Hunt, and Matthew Rushworth (38, 71) have found that developing a model about how much to trust another person's advice engages the superior temporal sulcus and temporoparietal junction (mirror-neuron regions), and that weighing the advice with one's personal factual models engages the ACC and ventromedial pre-frontal cortex.

Novel therapies for BPD may aim to directly alter the activity of these brain regions. For example, a recent study found that people's calculations about sharing money with a partner could be titrated by direct current stimulation of the right lateral pre-frontal cortex (72). These non-invasive safe techniques, practiced to effect, may have promise for modulating both mirror neuron and motor regions.

Furthermore, psychological tasks may allow us to quantify clinical progress. For example, the social valuation task developed by Behrens, Hunt, and Rushworth described above allows use to quantify relative use of social data (38, 71). Furthermore, holding a specific facial emotion modulates our reaction time to understand novel emotional sentences (73). These and other

measures could be used to assess the efficacy of novel socially directed interventions.

Finally, embodied robots, programed with the algorithms we outline, may be used to explore our hypotheses. There are already studies of robots interacting with one another (74). Aberrant prediction errors (and hence inappropriate association) can be introduced into a single robot. This engenders inappropriate behaviors by disrupting the learning and application of top-down priors (75). The same processes could be examined in two interacting robots. It would be useful to study the effect of different initial conditions and biases in the interactant (second robot) as well as the effect on the interactant of engaging with the BPD model robot.

Further research could help to clarify the relative contribution of specific cognitive mechanisms to symptom profiles for individual patients, and to develop (or assign already available) focused treatments to our patients.

SOCIAL LEARNING IN PSYCHIATRY

Of course, BPD is not the only mental illness characterized by social difficulties. Autism spectrum disorder (ASD) and psychotic illnesses are other salient examples of illnesses in which social interactions are profoundly challenging. In the spirit of the Research Domain Criteria initiative, it is important for us to consider whether the model that we have outlined is specific to BPD or may be applied to social dysfunction more broadly. Are we addressing something that occurs across diagnoses and might even speak to the debilitating and stigmatizing effects of having a diagnosis of mental illness?

Inappropriate habit learning has been implicated in psychotic illness and likewise, excessive precision (or certainty) has been associated with ASD (76). Both of these illnesses then would involve social dysfunction because of impaired model-based pre-frontal learning in the context of spared and perhaps even enhanced striatal responding. However, we feel that the manner in which the dysfunction arises is critical. While trauma may play a role in the genesis of psychosis, we feel there is something specific about the perturbed infant-parent interaction that culminates in BPD. By gathering the same data and performing the same model analyses in groups from each diagnostic category, we will be able to assess whether this prediction holds.

CONCLUDING REMARKS

Our prediction is that people need to be able to make and use predictive models to do well in social interactions. Therapies that engage one's own body (such as the mindfulness modules of DBT) may help to repair embodied simulations. This could be a useful pre-requisite to treatments that directly promote flexibility and precision in social simulations. Predictive models of social situations depend on understanding others' emotions and actions by simulating them in one's own mind *as though they are actually happening* in one's own body. This may only partially overlap with being able to verbalize other's intentions. Feeling those intentions is also key. Re-acquainting oneself with own body experience will facilitate simulating the experiences of others. Subsequently, when exposed to other patients in group therapy, our "body-centered patients" may be better able to use physical and verbal emotional displays from others to simulate and update their social predictions.

REFERENCES

1. Zanarini MC, Frankenburg FR, Reich DB, Fitzmaurice G. Attainment and stability of sustained symptomatic remission and recovery among patients with borderline personality disorder and axis II comparison subjects: a 16-year prospective follow-up study. *Am J Psychiatry* (2012) **169**(5):476–83. doi:10.1176/appi.ajp.2011.11101550
2. Gunderson JG, Stout RL, McGlashan TH, Shea MT, Morey LC, Grilo CM, et al. Ten-year course of borderline personality disorder: psychopathology and function from the Collaborative Longitudinal Personality Disorders study. *Arch Gen Psychiatry* (2011) **68**(8):827–37. doi:10.1001/archgenpsychiatry.2011.37
3. Clarkin JF, Levy KN. Psychotherapy for patients with borderline personality disorder: focusing on the mechanisms of change. *J Clin Psychol* (2006) **62**(4):405–10. doi:10.1002/jclp.20238
4. Johnson ML. *Girl in Need of a Tourniquet: Memoir of a Borderline Personality.* Berkeley, CA: Seal Press (2010). 247 p.
5. Zanarini MC. Psychotherapy of borderline personality disorder. *Acta Psychiatr Scand* (2009) **120**(5):373–7. doi:10.1111/j.1600-0447.2009.01448.x
6. Paris J. The treatment of borderline personality disorder: implications of research on diagnosis, etiology, and outcome. *Annu Rev Clin Psychol* (2009) **5**:277–90. doi:10.1146/annurev.clinpsy.032408.153457
7. Paris J. Pharmacological treatments for personality disorders. *Int Rev Psychiatry* (2011) **23**(3):303–9. doi:10.3109/09540261.2011.586993
8. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci* (2012) **16**(1):72–80. doi:10.1016/j.tics.2011.11.018
9. Huys QJ, Moutoussis M, Williams J. Are computational models of any use to psychiatry? *Neural Netw* (2011) **24**(6):544–51. doi:10.1016/j.neunet.2011.03.001
10. Stephan KE, Mathys C. Computational approaches to psychiatry. *Curr Opin Neurobiol* (2014) **25**:85–92. doi:10.1016/j.conb.2013.12.007
11. Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. *Nat Neurosci* (2011) **14**(2):154–62. doi:10.1038/nn.2723
12. Halligan PW, David AS. Cognitive neuropsychiatry: towards a scientific psychopathology. *Nat Rev Neurosci* (2001) **2**(3):209–15. doi:10.1038/35058586
13. Frith C. The neural basis of hallucinations and delusions. *C R Biol* (2005) **328**(2):169–75. doi:10.1016/j.crvi.2004.10.012
14. Marr D, Poggio T. From understanding computation to understanding neural circuitry. *Neurosci Res Prog Bull* (1977) **204**:301–28.
15. Fodor J. *The Modularity of Mind: An Essay on Faculty Psychology.* Harvard: MIT Press (1983).
16. Dennett DC. *Elbow Room: The Varieties of Free Will Worth Wanting.* Cambridge: MIT Press (1984). 200 p.
17. Brodersen KH, Schofield TM, Leff AP, Ong CS, Lomakina EI, Buhmann JM, et al. Generative embedding for model-based classification of fMRI data. *PLoS Comput Biol* (2011) **7**(6):e1002079. doi:10.1371/journal.pcbi.1002079
18. Brodersen KH, Deserno L, Schlagenhauf F, Lin Z, Penny WD, Buhmann JM, et al. Dissecting psychiatric spectrum disorders by generative embedding. *Neuroimage Clin* (2014) **4**:98–111. doi:10.1016/j.nicl.2013.11.002
19. Bowlby J. *A Secure Base: Parent-Child Attachment and Healthy Human Development.* New York: Basic Books (1988). xii,205 p.
20. Clarkin JF, Yeomans FE, Kernberg OF. *Psychotherapy for Borderline Personality: Focusing on Object Relations.* 1st ed. Arlington, VA: American Psychiatric Pub (2005). 397 p.

21. Pavony MT, Lenzenweger MF. Somatosensory processing and borderline personality disorder: pain perception and a signal detection analysis of proprioception and exteroceptive sensitivity. *Personal Disord* (2013) **27**(2):208–21. doi:10.1037/per0000017

22. Magerl W, Burkart D, Fernandez A, Schmidt LG, Treede RD. Persistent antinociception through repeated self-injury in patients with borderline personality disorder. *Pain* (2012) **153**(3):575–84. doi:10.1016/j.pain.2011.11.021

23. Rosenthal MZ, Ahn R, Geiger PJ. Reactivity to sensations in borderline personality disorder: a preliminary study. *J Pers Disord* (2011) **25**(5):715–21. doi:10.1521/pedi.2011.25.5.715

24. Niedtfeld I, Schulze L, Kirsch P, Herpertz SC, Bohus M, Schmahl C. Affect regulation and pain in borderline personality disorder: a possible link to the understanding of self-injury. *Biol Psychiatry* (2010) **68**(4):383–91. doi:10.1016/j.biopsych.2010.04.015

25. Dyer A, Hennrich L, Borgmann E, White AJ, Alpers GW. Body image and noticeable self-inflicted scars. *J Nerv Ment Dis* (2013) **201**(12):1080–4. doi:10.1097/NMD.0000000000000057

26. Hohner G, Teismann T, Willutzki U. Tattoos and piercings: motives for body modification in women suffering from borderline symptomatology. *Psychother Psychosom Med Psychol* (2014) **64**(2):63–9. doi:10.1055/s-0033-1347255

27. Linehan M. *Cognitive-Behavioral Treatment of Borderline Personality Disorder*. New York: Guilford Press (1993). xvii,558 p.

28. Young JE, Klosko JS, Weishaar ME. *Schema Therapy: A Practitioner's Guide*. New York: The Guilford Press (2003). xii,436 p.

29. Gallese V, Fadiga L, Fogassi L, Rizzolatti G. Action recognition in the premotor cortex. *Brain* (1996) **119**(Pt 2):593–609. doi:10.1093/brain/119.2.593

30. Keysers C, Gazzola V. Social neuroscience: mirror neurons recorded in humans. *Curr Biol* (2010) **20**(8):R353–4. doi:10.1016/j.cub.2010.03.013

31. Kilner JM, Neal A, Weiskopf N, Friston KJ, Frith CD. Evidence of mirror neurons in human inferior frontal gyrus. *J Neurosci* (2009) **29**(32):10153–9. doi:10.1523/JNEUROSCI.2668-09.2009

32. Pobric G, Hamilton AF. Action understanding requires the left inferior frontal cortex. *Curr Biol* (2006) **16**(5):524–9. doi:10.1016/j.cub.2006.01.033

33. Candidi M, Urgesi C, Ionta S, Aglioti SM. Virtual lesion of ventral premotor cortex impairs visual perception of biomechanically possible but not impossible actions. *Soc Neurosci* (2008) **3**(3–4):388–400. doi:10.1080/17470910701676269

34. Gallese V. Embodied simulation: from mirror neuron systems to interpersonal relations. *Novartis Found Symp* (2007) **278**:3–12; discussion 9, 89–96, 216–21. doi:10.1002/9780470030585.ch2

35. Goldman A. A moderate approach to embodied cognitive science. *Rev Philos Psychol* (2012) **3**:71–88. doi:10.1007/s13164-012-0089-0

36. Kilner JM, Friston KJ, Frith CD. Predictive coding: an account of the mirror neuron system. *Cogn Process* (2007) **8**(3):159–66. doi:10.1007/s10339-007-0170-2

37. Cooper RP, Cook R, Dickinson A, Heyes CM. Associative (not Hebbian) learning and the mirror neuron system. *Neurosci Lett* (2013) **540**:28–36. doi:10.1016/j.neulet.2012.10.002

38. Behrens TEJ, Hunt LT, Woolrich MW, Rushworth MFS. Associative learning of social value. *Nature* (2008) **456**(7219):245–U45. doi:10.1038/Nature07538

39. Friston K. A theory of cortical responses. *Philos Trans R Soc Lond B Biol Sci* (2005) **360**(1456):815–36. doi:10.1098/rstb.2005.1622

40. Friston K. The free-energy principle: a rough guide to the brain? *Trends Cogn Sci* (2009) **13**(7):293–301. doi:10.1016/j.tics.2009.04.005

41. Kahneman D, Slovic P, Tversky A. *Judgment under Uncertainty: Heuristics and Biases*. Cambridge, NY: Cambridge University Press (1982). xiii,555 p.

42. Gilovich T, Griffin DW, Kahneman D. *Heuristics and Biases: The Psychology of Intuitive Judgement*. Cambridge, NY: Cambridge University Press (2002). xvi,857 p.

43. Jacob P. How from action-mirroring to intention-ascription? *Conscious Cogn* (2013) **22**(3):1132–41. doi:10.1016/j.concog.2013.02.005

44. Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci* (2005) **8**(12):1704–11. doi:10.1038/nn1560

45. Balleine BW, Dickinson A. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* (1998) **37**(4–5):407–19. doi:10.1016/S0028-3908(98)00033-1

46. de Wit S, Dickinson A. Associative theories of goal-directed behaviour: a case for animal-human translational models. *Psychol Res* (2009) **73**(4):463–76. doi:10.1007/s00426-009-0230-6

47. Fonagy P, Bateman A. The development of borderline personality disorder – a mentalizing model. *J Pers Disord* (2008) **22**(1):4–21. doi:10.1521/pedi.2008.22.1.4

48. Panksepp J, Nelson E, Siviy S. Brain opioids and mother-infant social motivation. *Acta Paediatr Suppl* (1994) **397**:40–6. doi:10.1111/j.1651-2227.1994.tb13264.x

49. Corlett PR, Aitken MR, Dickinson A, Shanks DR, Honey GD, Honey RA, et al. Prediction error during retrospective revaluation of causal associations in humans: fMRI evidence in favor of an associative model of learning. *Neuron* (2004) **44**(5):877–88. doi:10.1016/j.neuron.2004.11.022

50. Gillath O, Bunge SA, Shaver PR, Wendelken C, Mikulincer M. Attachment-style differences in the ability to suppress negative thoughts: exploring the neural correlates. *Neuroimage* (2005) **28**(4):835–47. doi:10.1016/j.neuroimage.2005.06.048

51. Strathearn L, Fonagy P, Amico J, Montague PR. Adult attachment predicts maternal brain and oxytocin response to infant cues. *Neuropsychopharmacology* (2009) **34**(13):2655–66. doi:10.1038/npp.2009.103

52. Pally R. Non-conscious prediction and a role for consciousness in correcting prediction errors. *Cortex* (2005) **41**(5):643–62; discussion 731–4. doi:10.1016/S0010-9452(08)70282-X

53. Pally R. The predicting brain: unconscious repetition, conscious reflection and therapeutic change. *Int J Psychoanal* (2007) **88**(Pt 4):861–81. doi:10.1516/B328-8P54-2870-P703

54. Wassum KM, Cely IC, Maidment NT, Balleine BW. Disruption of endogenous opioid activity during instrumental learning enhances habit acquisition. *Neuroscience* (2009) **163**(3):770–80. doi:10.1016/j.neuroscience.2009.06.071

55. Nelson A, Killcross S. Amphetamine exposure enhances habit formation. *J Neurosci* (2006) **26**(14):3805–12. doi:10.1523/JNEUROSCI.4305-05.2006

56. FitzGerald TH, Dolan RJ, Friston KJ. Model averaging, optimal inference, and habit formation. *Front Hum Neurosci* (2014) **8**:457. doi:10.3389/fnhum.2014.00457

57. Mathys C, Daunizeau J, Friston KJ, Stephan KE. A bayesian foundation for individual learning under uncertainty. *Front Hum Neurosci* (2011) **5**:39. doi:10.3389/fnhum.2011.00039

58. Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HE, et al. Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* (2013) **80**(2):519–30. doi:10.1016/j.neuron.2013.09.009

59. Cook R, Dickinson A, Heyes C. Contextual modulation of mirror and countermirror sensorimotor associations. *J Exp Psychol Gen* (2012) **141**(4):774–87. doi:10.1037/a0027561

60. Diamond N. *Between Skins: The Body in Psychoanalysis – Contemporary Developments*. Oxford: John Wiley & Sons (2013). vi, 234 p. doi:10.1002/9781118321102

61. Gallese V, Eagle MN, Migone P. Intentional attunement: mirror neurons and the neural underpinnings of interpersonal relations. *J Am Psychoanal Assoc* (2007) **55**(1):131–76.

62. Ehrenberg S. Reflections on reflections: mirror use in a university dance training environment. *Theatre Dance Perform Train* (2010) **1**(2):172–84. doi:10.1080/19443927.2010.505001

63. Zhao W, Luo L, Li Q, Kendrick KM. What can psychiatric disorders tell us about neural processing of the self? *Front Hum Neurosci* (2013) **7**:485. doi:10.3389/fnhum.2013.00485

64. Maurex L, Lekander M, Nilsonne A, Andersson EE, Asberg M, Ohman A. Social problem solving, autobiographical memory, trauma, and depression in women with borderline personality disorder and a history of suicide attempts. *Br J Clin Psychol* (2010) **49**(Pt 3):327–42. doi:10.1348/014466509X454831

65. Sharp C, Ha C, Carbone C, Kim S, Perry K, Williams L, et al. Hypermentalizing in adolescent inpatients: treatment effects and association with borderline traits. *J Pers Disord* (2013) **27**(1):3–18. doi:10.1521/pedi.2013.27.1.3

66. King-Casas B, Sharp C, Lomax-Bream L, Lohrenz T, Fonagy P, Montague PR. The rupture and repair of cooperation in borderline personality disorder. *Science* (2008) **321**(5890):806–10. doi:10.1126/science.1156902

67. Mier D, Lis S, Esslinger C, Sauer C, Hagenhoff M, Ulferts J, et al. Neuronal correlates of social cognition in borderline personality disorder. *Soc Cogn Affect Neurosci* (2013) **8**(5):531–7. doi:10.1093/scan/nss028

68. Koenigsberg HW, Fan J, Ochsner KN, Liu X, Guise KG, Pizzarello S, et al. Neural correlates of the use of psychological distancing to regulate responses to negative social cues: a study of patients with borderline personality disorder. *Biol Psychiatry* (2009) **66**(9):854–63. doi:10.1016/j.biopsych.2009.06.010

69. Dziobek I, Preissler S, Grozdanovic Z, Heuser I, Heekeren HR, Roepke S. Neuronal correlates of altered empathy and social cognition in borderline

personality disorder. *Neuroimage* (2011) **57**(2):539–48. doi:10.1016/j. neuroimage.2011.05.005

70. Hitchcott PK, Quinn JJ, Taylor JR. Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. *Cereb Cortex* (2007) **17**(12):2820–7. doi:10.1093/cercor/bhm010

71. Behrens TE, Hunt LT, Rushworth MF. The computation of social behavior. *Science* (2009) **324**(5931):1160–4. doi:10.1126/science.1169694

72. Ruff CC, Ugazio G, Fehr E. Changing social norm compliance with noninvasive brain stimulation. *Science* (2013) **342**(6157):482–4. doi:10.1126/science. 1241399

73. Havas DA, Glenberg AM, Rinck M. Emotion simulation during language comprehension. *Psychon Bull Rev* (2007) **14**(3):436–41. doi:10.3758/BF03194085

74. Hoffman G. Embodied cognition for autonomous interactive robots. *Top Cogn Sci* (2012) **4**(4):759–72. doi:10.1111/j.1756-8765.2012.01218.x

75. Yamashita Y, Tani J. Spontaneous prediction error generation in schizophrenia. *PLoS One* (2012) **7**(5):e37843. doi:10.1371/journal.pone.0037843

76. Lawson RP, Rees G, Friston KJ. An aberrant precision account of autism. *Front Hum Neurosci* (2014) **8**:302. doi:10.3389/fnhum.2014.00302