Neuroscience and Biobehavioral Reviews 55 (2015) 247–267

Contents lists available at ScienceDirect

Neuroscience and Biobehavioral Reviews

journal homepage: www.elsevier.com/locate/neubiorev

Review

Reinforcement learning in depression: A review of computational research

Chong Chen [a,*], Taiki Takahashi [b], Shin Nakagawa [a], Takeshi Inoue [a,1], Ichiro Kusumi [a]

[a] Department of Psychiatry, Hokkaido University Graduate School of Medicine, Sapporo 060-8638, Japan
[b] Department of Behavioral Science/Center for Experimental Research in Social Sciences, Hokkaido University, Sapporo 060-0810, Japan

ARTICLE INFO

Article history:
Received 23 October 2014
Received in revised form 20 April 2015
Accepted 4 May 2015
Available online 12 May 2015

Keywords:
Anhedonia
Computational psychiatry
Depression
Dopamine
Incentive salience
Learning rate
'Liking'
Model-free
Model-based
Motivation
Prediction error
Reinforcement learning
Reward sensitivity
Stress
'Wanting'

ABSTRACT

Despite being considered primarily a mood disorder, major depressive disorder (MDD) is characterized by cognitive and decision making deficits. Recent research has employed computational models of reinforcement learning (RL) to address these deficits. The computational approach has the advantage in making explicit predictions about learning and behavior, specifying the process parameters of RL, differentiating between model-free and model-based RL, and the computational model-based functional magnetic resonance imaging and electroencephalography. With these merits there has been an emerging field of computational psychiatry and here we review specific studies that focused on MDD.

Considerable evidence suggests that MDD is associated with impaired brain signals of reward prediction error and expected value ('wanting'), decreased reward sensitivity ('liking') and/or learning (be it model-free or model-based), etc., although the causality remains unclear. These parameters may serve as valuable intermediate phenotypes of MDD, linking general clinical symptoms to underlying molecular dysfunctions. We believe future computational research at clinical, systems, and cellular/molecular/genetic levels will propel us toward a better understanding of the disease.

Contents

* Corresponding author at: Department of Psychiatry, Hokkaido University Graduate School of Medicine, Kita 15, Nishi 7, Kita-ku, Sapporo 060-8638, Japan. Tel.: +81 11 706 5160; fax: +81 11 706 5081.
E-mail address: cchen@med.hokudai.ac.jp (C. Chen).
[1] Present address: Department of Psychiatry, Tokyo Medical School, Tokyo 160-0023, Japan.

http://dx.doi.org/10.1016/j.neubiorev.2015.05.005
0149-7634/

1. Depression and reinforcement learning

Due to its hallmark feature of anhedonia, the inability to experience reward, major depressive disorder (MDD) has been considered primarily a mood disorder (Beck and Alford, 2009; American Psychiatric Association, 2014). However, deficits in cognitive function (McDermott and Ebmeier, 2009; Gotlib and Joormann, 2010; Lee et al., 2012b; Snyder, 2013; Belzung et al., 2014b) and decision making (Treadway and Zald, 2011; Paulus and Yu, 2012; Must et al., 2013) also occur in MDD, which reduces response to treatment, increases the risk of relapse and recurrence, and impairs social adaptation. Recently several lines of research have suggested dysfunctional reinforcement learning (RL) in MDD. RL, the process of maximizing reward and minimizing loss by modifying the behavior as a consequence of experience with the environment, plays a central role in decision making (Sutton and Barto, 1998; Schultz, 2006; Doll et al., 2012; Glimcher and Fehr, 2013).

Firstly, patients with MDD perform poorly on the Iowa Gambling Task which relies on RL (Must et al., 2013). In this task, subjects are asked to choose cards from four decks followed by monetary reward or punishment. Two of the four decks are disadvantageous because even though picking from them may be followed by a big gain, a bigger punishment occasionally occurs which eventually leads to inferior outcomes. Subjects have to learn this underlying rule by trial and error experiences. Research has consistently found that MDD patients tend to choose from the disadvantageous decks (Must et al., 2013).

Secondly, research using the signal-detection task also suggests impaired RL in MDD (Ragland et al., 2009; Pizzagalli, 2014). Similarly subjects have to learn a hidden rule in a probabilistic reward task and signal-detection analysis has shown that healthy volunteers would eventually demonstrate a bias toward frequently rewarded choices. The bias is likely based on implicit learning as it is rather difficult to infer the hidden rule explicitly (Ragland et al., 2009). On the other hand, MDD patients generally fail to show this bias and the impairment is correlated with anhedonia symptoms (Ragland et al., 2009; Pizzagalli, 2014).

Thirdly, consistent with these behavioral dysfunctions, in reward related tasks fMRI research has consistently found a low striatal (but see Knutson et al., 2008. Experimental design may have contributed to this inconsistency, which will be discussed later in Section 3.4) and high medial prefrontal cortex (PFC, although this is still somewhat controversial) response during monetary anticipation and/or outcomes stages in MDD (Steele et al., 2007; Kumar et al., 2008; Gradin et al., 2011. See a review by Forbes and Dahl, 2012; Zhang et al., 2013; Kerestes et al., 2014; Pizzagalli, 2014). Notably, striatum and medial PFC are both implicated in RL (Daw et al., 2011; Doll et al., 2012; Garrison et al., 2013; Chase et al., 2015).

Therefore a close examination of RL in depression may provide promising new insights into the underlying behavioral, cognitive and neural pathophysiology of the disease.

2. The computational approach and its merits

In the past decades, arising from two fast-advancing (but overlapping) fields of computational neuroscience (Sutton and Barto, 1998; Schultz, 2006; Niv, 2009; Doll et al., 2012) and neuroeconomics (Hasler, 2012; Sharp et al., 2012; Glimcher and Fehr, 2013), the computational theory of RL has been a major framework accounting for decision making. Under this framework, subjects choose actions according to mathematical value functions, which define the expected value of each action. Value functions can be updated through historical trial and error experience (i.e. prediction errors (PEs), the difference between received and expected values) or by prospective planning based on internal cognitive maps or learned models of the environment. The former is known as model-free RL and is habitual and slow to change. The latter is known as model-based RL, is goal-directed, and allows subjects to update value functions more flexibly (Daw et al., 2005; Dayan, 2009; Doll et al., 2012; Daw and O'Doherty, 2013; Dolan and Dayan, 2013; Nakahara, 2014). In model-based RL, subjects may simulate the consequences of potential actions that they may choose and then use the hypothetical outcomes to update their value functions. Therefore it is also known as counterfactual thinking (Lee et al., 2012a). There are at least four advantages of this computational approach.

Firstly, the mathematical analysis of RL allows explicit testable predictions about learning and behavior, thus linking learning and behavior to psychological processes and providing a useful normative framework to study them (Niv, 2009; Wiecki et al., 2015).

Secondly, the computational approach specifies the process parameters of RL, such as PE, learning rate, reward sensitivity, and memory of previous reinforcement (which will be discussed in detail below). For instance, in the field of depression, this reduces the broadly defined symptom of anhedonia to more specific and refined constructs (Treadway and Zald, 2011; Der-Avakian and Markou, 2012. See our later discussion of 'wanting', 'liking' and learning in Section 3.4). This moves the field from general symptom-based descriptions to direct and precise identification of impaired latent functions (Wiecki et al., 2015).

Thirdly, the recent differentiation of model-free and model-based RL brings the two most important but separated fields, cognition and reward learning/decision making together. This enables novel synthesis and may generate valuable insights regarding learning and behavior. The differentiation also perfectly resembles popular dual-process theories of decision making, learning and memory, such as the models of automatic and controlled processing (Shiffrin and Schneider, 1977), feeling and thinking (Zajonc, 1982), experiential and cognitive processing (Epstein, 1994), system 1 and system 2 (Kahneman, 2011), habitual and goal-directed control (Dickinson and Charnock, 1985; Balleine, 2005; Balleine and O'Doherty, 2010) and nondeclarative and declarative memory (Squire and Zola, 1996) (for an insightful review, see Dayan, 2009). Thus the computational approach provides a powerful and parsimonious tool to address these theories.

Fourthly, computational modeling makes model-based functional magnetic resonance imaging (fMRI) (O'Doherty et al., 2007. See also Huettel, 2012) and electroencephalography (EEG) (Larsen and O'Doherty, 2014) possible, which moves observations from the behavioral and cognitive level down to the underlying neural level. Briefly speaking, in model-based fMRI and EEG, a computational model is first fit to observed behaviors, and the best-fitting model is regressed against the fMRI data of blood-oxygen-level dependent (BOLD) or EEG signal changes over time. In this way,