Short Thesis for the Degree of Doctor of Philosophy (PhD)

The FNDC5/irisin/BDNF axis, as a modulator of reinforcement learning and mood in health and disease

by: Dr. Csaba Papp

supervisor: Dr. Judit Zsuga

University of Debrecen
Doctoral School of Health Sciences
Debrecen, 2017

The FNDC5/irisin/BDNF axis, as a modulator of reinforcement learning and mood in health and disease

By Csaba Papp MD

Supervisor: Judit Zsuga MD, PhD

Doctoral School of Health Sciences (Preventive Medicine and Public Health program)
University of Debrecen

Head of the **Defense Committee**:	György Paragh MD, PhD, DSc
Reviewers:	Éva Szőke PhD
	Miklós Vecsernyés PharmD, PhD
Members of the Defense Committee:	Antal Bugán PhD
	Csaba Dózsa PhD

The PhD Defense takes place at the Lecture Hall of Bldg. A, Department of Internal Medicine, Faculty of Medicine, University of Debrecen

13 p.m., July 7, 2017

Table of contents

1 Introduction

1.1 General overview

The need to combat chronic non-communicable diseases (NCDs) is clearly undisputable, forcing high-income countries to develop measures to overcome this challenge, a challenge placed in the focus of numerous joint initiatives and international policies including the “Global action plan for the prevention and control of noncommunicable diseases 2013-2020”. Unhealthy diet, harmful consumption of alcohol, tobacco use and physical inactivity are considered as fundamental risk factors for chronic NCDs, e.g. diabetes mellitus, cancer, cardiovascular diseases, and chronic inflammatory lung disease, primarily sources of premature death globally, according to the World Health Organization (WHO). Nevertheless, instead of viewing these elements as risk factors, each may be conceptualized as risk behaviors, a perception underscored by the contemporary finding that behavioral modifications resulting in permanent change of lifestyle are among the most efficacious methods for preventing and halting the progression of major non-communicable diseases.

A common denominator for behavioral change is motivated behavior governed by reinforcement learning. Based on these considerations, basic understanding of the neuronal circuitry underlying behavioral change and its autocrine/paracrine/endocrine modulators is obligatory for developing novel diagnostic markers and identification of therapeutic targets.

Reinforcement learning is a concept underlying forms of associative learning governed by the use of a scalar reward signal, with learning taking place if expectations are violated. Stemming from computational accounts, the reinforcement learning paradigm is conceptualized as learning performed by an agent (e.g. the individual) that is interacting with its environment (referred to as a set ‘states’), an environment that is unknown and uncontrolled by said agent. Reinforcement learning works by utilizing Pavlovian learning to link states and rewards by learning the contingency of neutral state-related stimuli and stimuli predictive of reward, and instrumental learning to optimize action selection in order to maximize the reward obtained long run. Depending on the agent’s approach both accounts can be addressed with or without using the agent’s model of the environment. Accordingly, there are two alternate ways to address reinforcement learning related paradigms: the model-free and the model-based approaches, based on whether the agent attempts to build a model of the external and internal environment or not. A set of distinct neuronal structures have been linked to model-free and model-based reinforcement learning, with the hippocampus, amygdala and orbitofrontal cortex (OFC) being linked to model-based and the mesencephalic dopaminergic neurons being associated with model-free learning.

Additional to the significant role of the neurocircuitry of behavioral change, elucidation of its modulation by humoral factors is also important. One influential factor is brain derived neurotrophic factor (BDNF), a neurotrophin that has a substantial role regarding the plasticity of the CNS, modulates neurotransmitter, especially serotonin and dopamine release. Hence, decrease of BDNF level alters synaptic strength, as BDNF induces long-term potentiation, an integral process for reinforcement-learning, thus consequently influences the process of behavioral modification. BDNF is synthetized throughout the neurons of the CNS including the midbrain’s ventral tegmental area (VTA), hippocampus and peripheral organs such as skeletal muscle, liver and adipose tissue. BDNF is able to cross the blood-brain barrier. BDNF has a specific high affinity receptor, the tropomyosin-related kinase B (TrkB), a tyrosine kinase receptor mediating its effects.

The link between mesocortico-limbic function and BDNF is well established as BDNF is expressed and has direct influence on neuronal circuits activated by reinforcement learning and related reward-processing. BDNF’s role concerning the development of the dopaminergic system is underscored by its efficacy as a trophic factor in human and rat mesencephalic dopaminergic neuron cultures, indicated by increase of neuronal survival and elevation of neuronal tyrosine hydroxylase and dopamine content. Furthermore, Previous findings show the expression of TrkB receptors by mesencephalic dopaminergic neurons, hence making these neurons susceptible to the endocrine/paracrine and autocrine effects of BDNF.

1.2 Model-free and model-based reinforcement learning

Model-free approaches such as temporal difference learning (or its variants e.g. the actor-critic model, SARSA a policy- and Q-learning a value-iteration method) work by making predictions about the future value of states (in Pavlovian learning) or state-action pairs (instrumental learning) based on direct interaction with the environment e.g. using experience to sample the environment. Model-free learning is governed by the utility of a stimulus (Pavlovian learning) or outcome (instrumental learning) with respect to its predicted cumulative future value discounted as a function of time, embodied by the reward prediction error.

Compelling evidence from varying fields of neuroscience depicted dopaminergic neurons of the mesocortico-limbic system as candidate neural substrates for model-free information processing. Indeed, to date the most elaborate framework accounting for the role of dopaminergic (DA) neurons in reinforcement learning is the RPE hypothesis. Accordingly, phasic DA neuronal activity is triggered by either unexpected reward, or sensory signals for unexpected rewards. The difference between the predicted and actual value of rewards are depicted yielding a response pattern that is in alignment with a model-free appetitive RPE. Dopaminergic neurons emanating from the ventral tegmental area and pars compacta substantiae nigrae project to the ventral striatum (VS), forming the key neuronal pathway of the mesocortico-limbic circuit. Additionally, VTA DA neurons project to the amygdala, hippocampus and the OFC and dorsal striatum. To account for learning, RPE serves as plasticity-modulating teaching signal that governs future behavior in order to maximize predicted future reward and consequently minimize RPE. During learning the dopamine teaching signal is transferred from the unconditioned primary reinforcer to its preceding sensory cue in a way that allows for the back-propagation of the teaching signal so reward may be predicted at the earliest time possible. Burst generation of VTA DA neurons depend on intact synaptic inputs, as the presence of glutaminergic synaptic drive from the subthalamic nucleus (STN) and laterodorsal tegmentum gated glutaminergic and cholinergic activity of the pedunculo-pontine tegmental nucleus (PPTgN) are prerequisites for burst firing to occur. The PPTgN offers one of the strongest excitatory drives to the VTA.

Model-based approaches work by building a model of the internal and external environment e.g. of distinct states encompassing rewards (described by the reward function), their connections and probabilities governing state transitions (characterized by the transition function). Model-based methods use the knowledge of sequential contingencies describing state transitions and the reward function for either making predictions pertaining to future rewards, or planning e.g. use of the model for making forward-looking mental simulations relating to actual or imagined states (Pavlovian learning) or state-action pairs (instrumental learning). By running serial computations concerning the immediate consequences of state transitions (Pavlovian learning) and/or action sequences (instrumental learning) the utility of a state (Pavlovian learning) or state-action pair (instrumental learning) is obtained indirectly based on the model.

There are numerous reports dealing with the involvement of specific brain structures in model-based reinforcement learning, however propositions concerning the unambiguous neurobiological underpinning of the model seem to be missing. In the following section utilizing the proactive brain concept we elaborate that the brain's default network might be the structure that has the function to continuously work on modeling the environment. In addition, we will show that several components of the default network (e.g. the hippocampus, the amygdala and the OFC) have been linked to model-based reinforcement learning, thus it is convenient to suggest that the model constructed by the default network is accessed by the reinforcement learning system (both by Pavlovian and instrumental learning) to allow model-based computations.

1.3 The proactive brain

The previously known concept of proactive brain provides a framework for understanding how the brain makes predictions. In focus of this theory stands the proposition that the brain in its default mode is continuously making predictions by means of associations driven by external gist information obtained from sensory stimuli or alternatively by thought mobilizing relevant memories. These associations as elemental building blocks of thought enable taking advantage of frequent trends in the environment to help interpret and anticipate immediate future events therefore help cope with uncertainty, resolve ambiguity, generate reward-related perceptions, and govern action selection as these predictions offer information pertaining to what to expect in a given environment. Accordingly, one of the key functions of the default network (as indicated by its overlap with contextual associative areas) is to organize typical arrangements of the environment into context frames that contain typical, generic representations such as the probable objects clustered together, their relations and affective and reward value and offer a set of expectations that can govern attention and action selection. The structure of context frames enables pattern activation in associative memory e.g. automatic co-activations based on feature similarity, spatiotemporal contiguity. By organizing the external world and internal states into context frames, information pertaining to valence and reward value of stimuli is accessible in a way that it may assist making predictions facilitating action selection for maximizing reward.

1.4 Modulation of the reward circuity: the FNDC5/irisin/BDNF axis

Irisin was discovered by Boström and colleagues in mice and humans. It is a highly conservative polypeptide (12 kDa) with its amino acid sequence showing 100% homology in most mammals, reflecting highly conserved function. Irisin is released by proteolysis from the transmembrane fibronectin type III domain-containing 5 protein (FNDC5). FNDC5 expression is regulated by the transcriptional co-activator, peroxisome proliferator-activated receptor-gamma coactivator protein-1α (PGC1α), a known regulator of oxidative processes in BAT (Phillips et al., 2014). Expression of both FNDC5 and PGC1α are positively correlated with physical activity and with each other, evidenced by their decreased and increased expression in response to sedentary lifestyle or sustained physical training, respectively. Irisin is released into the circulation following proteolytic cleavage of FNDC5. Although most abundant in skeletal muscle, FNDC5/irisin expression is present in other tissues as well, including the adipose tissue, (hence irisin is an adipokine too), the rectum, the tongue, and the brain.

Additional to irisin's role as a link between skeletal muscle and adipose tissue, its effect in the central nervous system (CNS) is also being acknowledged. FNDC5 mRNA was isolated from distinct structures linked to reinforcement learning processes e.g. the hippocampus (model-based learning), and midbrain (model-free learning) in rodents. Furthermore, it should be noted that irisin produced in the periphery readily crosses the blood-brain barrier. The most relevant central effect of neuronal FNDC5/irisin regarding its contribution to reinforcement learning processes may be its effect to induce BDNF expression in relevant brain areas. Forced hippocampal expression of FNDC5, for example resulted in BDNF expression, and conversely, parallel to the increase of irisin's level in the systemic circulation led to increased hippocampal BDNF expression. Moreover, alteration of the expression of FNDC5 by RNAi, small noncoding RNA transcripts, regulators of mRNA expression-mediated knockdown in cortical neurons, resulted in a parallel reduction of cortical BDNF expression.

1.5 Mood disturbance and the reinforcement learning paradigm

Several disorders have been mapped onto the reinforcement learning paradigm including depression, with distinct attributes of value-based decision making being altered. For example, higher discounting rates for delayed rewards reflective of hopelessness and unwillingness to invest in the future was shown in major depressive disorder. In a different study, anhedonia, one of the cardinal symptoms of major depressive disorder, was associated with diminished primary sensitivity to rewards. Accordingly, the dysfunction of VTA - VS axis have been specifically associated with anhedonia and anergy, also characteristic of depression.

Summarizing, it may be proposed that change of reward processing in mood disorders may be accompanied by alteration of the irisin/BDNF axis.

1.6 Mood disturbance, a highly co-morbid disorder of COPD

Chronic inflammatory pulmonary disease (COPD), by profoundly impacting the patient’s quality of life, poses great socio-economic burden for individual patients, their families and society. COPD primarily worsens quality of life by developing chronic, progressive dyspnea and consequent limitation of physical activity. Moreover, co-existing mental health problems show higher prevalence in COPD patients than in the general population, with depression and anxiety being present in approximately 20-40% and 30-50% of COPD cases, respectively. Disturbance of mood not only causes disability per se, but changes the course of the disease by altering how patients experience and manage their disease thus worsening their functional and health status. Therefore, the quest to elucidate the potential mechanisms underlying mood disturbances in COPD is ever so pressing.

1.7 Goals

During our investigations, we – directly or indirectly – put forward the hypothesis that the common denominator of risk behaviors underlying the most burdening NCDs may be the dysfunction (or untoward function) of the mesocortico-limbic system and consequent alteration of the reinforcement learning system and that a number of humoral factors appear to play a significant role in its evolution. According to this our aims were

1. To develop a unified, integrated model for reinforcement learning that has the ability to account for the model-based and model-free accounts of reinforcement learning
2. To offer a neurobiological substrate for the model used by the model-based reinforcement learning system.
3. To clinically validate the hypothesis that humoral mediators, such as irisin and BDNF, can contribute to the development of disorders e.g. mood disorders in which the reinforcement learning paradigm has been previously described to be altered.

2 Materials and methods

2.1 Conceptualization

Based on theoretical premises we conceptualized a model that links together concepts of reinforcement learning (e.g. both Pavlovian and instrumental learning), model-based and model-free accounts by utilizing functional anatomical considerations, concepts of machine learning (e.g. the reinforcement learning agent) and the theoretical framework of ‘proactive brain’ deploying the brain’s default network. The merit of the resultant concept termed “proactive model of reinforcement learning” is that it gives rise to several testable hypotheses and offers a representational architecture carrying clinical implications. The current work is deeply rooted in the conceptual and experimental findings of others, cited throughout the thesis.

2.2 Study design for investigating COPD patients

This investigation was designed in agreement with the STROBE statement for cross-sectional studies and is in line with the principles established by the Declaration of Helsinki. Approval of the Ethical Committee of the University of Debrecen (DEOEC RKEB/IKEB 3632-2012) was obtained in advance. Informed consent was obtained from each participant.

In this study, data of a COPD patient population was analyzed. Briefly, every COPD patient, attending the outpatient unit of the Department of Pulmonology (University of Debrecen) between September 1, 2012 and October 15, 2013 for the management of COPD, were screened by attending pulmonologists, who were unaware of the research hypothesis and study protocol. Patients suffering from any acute inflammatory disease over the preceding one month and those having benign or malignant tumors in their case history were excluded. Patients meeting the entry criteria were referred to the study nurse who explained the details of the study and obtained informed consent. Every patient referred by the pulmonologists consented to study participation. Overall, 74 COPD patients were recruited. At the time of inclusion, patients were managed for COPD according to the relevant Hungarian practice guideline and the GOLD initiative. Patients received therapy at the time of inclusion as clinically warranted. Whole-body plethysmography was performed for every patient to obtain lung function parameters. Demographic, anthropometric, anamnestic, laboratory and quality of life data were also acquired. Cumulative measure of smoking exposure was described by pack-years (accounted for both past and current smoking exposure). To assess disease-specific quality of life, the official Hungarian version of Saint George's Respiratory Questionnaire (SGRQ) was used with the permission of the proprietor (Paul Jones, University of London, London, UK).

2.3 Pulmonary function testing

Whole-body plethysmography was performed according to the ATS/ERS criteria (Miller et al., 2005; Wanger et al., 2005) with Piston whole-body plethysmograph (PDT-111/p, Piston Medical, Budapest, Hungary). Plethysmography was performed while patients were receiving long-term therapy for COPD. The best of three technically sound maneuvers was selected in case of each participant. Regarding resistance curves, at least two separate and technically appropriate measurements were performed and results were accepted only if these were the same for both measurements.

2.4 Blood samples and routine laboratory tests

Blood samples were obtained in the morning of the examination, after an overnight fast. Routine laboratory investigations were performed following standard procedures. Serum or plasma samples were used to characterize carbohydrate homeostasis, lipid homeostasis, kidney function, liver function, status of skeletal muscles, thyroid-stimulating hormone-sensitive (sTSH) and systemic inflammation. Serum samples used to determine irisin and BDNF were frozen within 60 minutes and stored at -80 °C until further analysis.

2.5 Determination of serum irisin and BDNF

Serum BDNF and irisin levels were measured using commercial ELISA kits, compliant with the manufacturer's instructions (Sigma-Aldrich, MO, USA and Phoenix

Pharmaceuticals, Burlingame, CA, USA, respectively). All measurements were performed in duplicate. A standard curve showing linear relationship between optical density and concentration of irisin as well as BDNF were obtained with each plate. The detection limit for BDNF was less than 80 pg/ml. In line with the manufacturer's informaiton, the irisin standard curve was linear from 1.34 to 29.0 ng/ml, and the detection limit was 1.34 ng/ml. For the stratification of the final multiple regression model, serum BDNF levels were dichotomized according to their median value.

2.6 St. George's Respiratory Questionnaire (SGRQ)

The official Hungarian version of SGRQ validated for a 1-month recall period was used according to the SGRQ manual supplied by the proprietor. SGRQ quantifies health impairment with three component scores (Activity, Symptoms and Impact component score) and one total score of which the Impacts score also strongly correlates with disturbances of mood (e.g. depression). The Scores are provided as a percentage, thus 100% indicates the worst and 0% represents the best subjective health status. Two independent raters recorded data by diligently following data entry guidelines, and scoring was done using the score calculation algorithm provided by the developer of the SGRQ. Inter-rater variability assessed by Spearman correlation was 0.99 ($p<0.001$), 0.988 ($p<0.001$), 0.999 ($p<0.001$) and 0.999 ($p<0.001$) for the Symptoms, Activity, Impacts and Total scores of SGRQ, respectively. Both raters were blinded to raters were blinded to patients' irisin and BDNF levels.

2.7 Statistical analysis

Disturbances of mood were quantified with the Impacts score reflective of mood disorders and overall psycho-social dysfunction.

The mean of the Impacts score was used as cutoff for dichotomization of the COPD cohort, so patients with Impacts score < 32.65% were put into the lower Impacts score group (n=40), while patients with Impacts score ≥ 32.65% formed the higher Impacts score group (n=34), corresponding to less or more pronounced mood disturbances, respectively. Normality of continuous variables was checked by the Shapiro-Wilk test. For variables following Gaussian distribution, two datasets were compared using Student’s t-test, while Mann-Whitney U test was carried out for those not showing normal distribution. Frequencies were compared with Pearson’s $\chi 2$ test.

The correlation of mood disturbance and serum irisin concentration was established using Spearman’s correlation. The relationship between mood disturbance and serum irisin level was further investigated with simple as well as multiple linear regression.

Simple linear regression was carried out with traditional confounding factors (age, gender, height, disease duration in years), lung function parameters and routine laboratory parameters obtained from the serum or plasma samples. Missing data were omitted. To eliminate effects of potential confounders, multiple linear regression modeling was performed. First, a multiple model was compiled including all significant regressors identified by means of simple linear regression and a priori variables (age and gender). Variables were introduced into the initial multiple model simultaneously, then factors not contributing significantly to the model were deleted (except for the a priori variables). The final model contained (in addition to the a priori parameters) FEV1% pred, body mass index, weight and (log) triglyceride levels. Furthermore, the final model was stratified with respect to BDNF levels. Heteroskedasticity and goodness of fit for the model was assessed by Cook-Weisberg and Ramsey test.

Statistical analysis was performed with Stata 13.0 software (Stata Corporation). Values are given as mean ± SD or median (with the interquartile range: IQR), and regression coefficients are presented with their 95% confidence interval (CI).

3 Results

3.1 The proactive model of reinforcement learning

Our conceptualization yielded the proactive model of reinforcement learning that articulates two premises. On one hand based on neuroanatomical observations of others, we propose that the ventral striatum compiles the value function component of the reinforcement learning agent by integrating model-free and model-based inputs about rewards in a way that value is computed. Hence, VS supplies value information for both types of reinforcement learning paradigms, e.g. Pavlovian and instrumental learning. On the other hand, the proactive model of reinforcement learning utilizes the default mode network as the model for the model-based system with the amygdala, hippocampus and the orbitofrontal cortex assuming specific reward-related roles. The default network builds on episodic memory to function in a way that allows future-oriented predictions and planning. The amygdala couples with VS to code stimulus-outcome contingencies. The hippocampus couples with VS to code context-outcome contingences while the OFC is driven by hippocampus and amygdala to integrate reward-related information into context frames thus OFC will provide information about expected rewards. The model-free system (e.g. the PPTgN, VTA and the VS) together compute reward prediction error. Given that the OFC has strong connections with the VTA and its afferent PPTgN, structures relevant for dopaminergic burst firing, the OFC may offer modulatory input that further interferes with the model-free system. Summarizing the expected reward information provided by the OFC gets fed back to the model-free system, in a way that the VS may integrate model-based reward information with the model-free reward prediction errors to compute the value signal emitted by the VS.

3.2 Association between serum levels of irisin and BDNF and mood disturbances in COPD

3.2.1 The COPD patient population involved

The baseline characteristics of our COPD patient cohort were as follows. The mean age of the patient population is 62.15±9.70 years, 27 patients were female and 25 patients were smokers. The median disease duration in years was 5 years (3.00 to 10.00 years). The median of the SGRQ component scores were: SGRQ Symptoms score 32.66 (13.64-58.28), SGRQ Impacts score 29.64 (15.44-49.79), SGRQ Activity score 57.32 (47.24-72.08), while the mean of the Total score was 41.08±20.99. Serum irisin and BDNF levels were 7.22 (6.63-8.10) ng/ml, 345.6 (294.20-387.90) ng/ml, respectively.

3.2.2 Comparison of patients with respect to mood disturbance

The two groups of COPD patients, dichotomized with respect to the mean of Impacts score, proved to be homogenous regarding most of the parameters investigated. Nevertheless, in the group with higher Impacts score (showing more pronounced impairment of mood), dyslipidemia (9/40; 17/34 p=0.017 for lower vs. higher impacts score) and hypertension (18/40; 25/34 p=0.013 for lower vs. higher impacts score) (as anamnestic data) were more frequent, and differences were seen with respect to serum LDL (3.25 (2.65-4.00) mmol/L; 2.80 (2.40-3.30) p=0.030 for lower vs. higher impacts score), serum irisin (7.37 (6.99-8.12) ng/ml; 6.85 (6.49-7.99) p=0.030 for lower vs. higher impacts score), FEV1% pred (71.37±19.26; 60.85±20.25 p=0.025 for lower vs. higher impacts score), FVC% pred (86.3±16.92; 78.12±17.25 p=0.043 for lower vs. higher impacts score) and serum glucose (4.08 (4.00-5.50); 5.30 (4.70-6.60) p=0.015 for lower vs. higher impacts score.

3.2.3 Associations among SGRQ’s Impacts score, serum irisin and BDNF levels

Upon assessing the correlation between the Impacts score and reciprocal of irisin, we found a significant positive correlation in the whole COPD cohort (Spearman correlation coefficient: 0.26, p=0.02;), in agreement with the finding that the irisin concentration was smaller in the higher Impacts score group. This correlation became stronger (and remained almost statistically significant) in the stratum with lower BDNF level, while it was weaker (and non-significant) in the stratum with higher BDNF (Spearman correlation coefficient: 0.32 and 0.22, p=0.055 and p=0.19, respectively). Albeit the relationship between the Impacts score and reciprocal of serum irisin level characterized by simple linear regression was not significant (p=0.08) after correction for possible confounders by multiple linear regression the

Impacts score and reciprocal of irisin showed a strong, significant, positive association (β: 419.97; CI: 204.31, 635.63; p<0.001). This association became even more distinct among patients with lower BDNF levels (β: 434.11; CI: 166.17, 702.05; p=0.002), while a considerably weaker and statistically non-significant association was present in case of patients with higher BDNF concentrations (β: 373.49; CI: -74.91, 821.88; p=0.10). The Cook-Weisberg test showed no heteroskedasticity for the full model and strata with lower and higher BDNF (p=0.92, p=0.67 and p=0.82, respectively). Furthermore, all three models showed good fit reflected the Ramsey test (p=0.82; p=0.53 and p=0.79 for the whole data set and strata with lower and higher BDNF, respectively).

Based on the final multiple linear regression model (built for the Impacts score), body mass index (BMI) (β: 3.68; CI: 2.01, 5.34; p<0.001), (log)triglyceride (β: -8.70; CI: -16.38, -1.02; p=0.027) and body weight (β: -0.78; CI: -1.31, -0.25; p=0.004) were significantly associated with mood disturbances among COPD patients. In addition, severity of airflow limitation, characterized by FEV1% pred, showed a significant negative association with the Impacts score (β: -0.52; CI: -0.71, -0.32; p<0.001).

4 Discussion

4.1 Components of the proactive model

4.1.1 Default network for building a model of the environment

Decreased brain activity seen in response to cognitive tasks as opposed to the heightened metabolic activity encountered in the same areas upon passive mental states (e.g. when left to think undisturbed) is the hallmark of the default mode network. Based on the ubiquity of its involvement in mental processes, the default network was proposed to have an essential, adaptive function. More interestingly the proactive brain concept has attributed an integrative function to the default network, based on the elaborate overlap seen between the contextual associative network and that the default network (e.g. that the default network continuously creates and updates the model of the environment by means of generating contextual associations). The core regions associated with this network include the ventro-medial and ventro-lateral prefrontal cortex (including the OFC), the posterior cingular/retrosplenial cortex, the inferior parietal lobule, the lateral temporal cortex, the hippocampal formation (including the entorhinal and the parahippocampal cortex) and the amygdala. The connection of the default network with striatal reward pathways has also been established, in fact there is evidence for the default network supporting goal-directed simulations.

In our view this function pertains to the use of a model in reinforcement learning for the purpose of planning and optimizing policy in a way to maximize the sum of future discounted reward. The following subsections will overview the component structures of the default network that are involved in reward-related processes used in model-based learning.

4.1.2 Model-based Pavlovian learning: amygdala-VS computes stimulus-reward contingencies

Intact amygdala function is necessary for delivering reward-related information about discrete stimulus. The amygdala is highly important for extracting the affective properties of a stimulus and as such assumes a central role in processing aspects that are in agreement with actual goals. In other words, amygdala responds to motivationally relevant information with relevance indicating the congruence of the affective component of the stimulus and goal.

4.1.3 Model-based Pavlovian learning: hippocampus-VS compute context-reward contingencies

The hippocampus is postulated to be a key in contextual learning as its role in encoding and remembering contexts is well described. The hippocampus is able to encode the spatial properties of the environment, it encodes interoceptive contexts (such as hunger, thirst and timing) and accordingly is able to influence motivated behavior by integrating information about internal and external state related information into response-outcome relationships. Furthermore, it has been demonstrated that VS-hippocampal connections have a role in learning place-reward associations. The possible contribution of the hippocampus to model-based reinforcement learning has been postulated by accounts which give evidence for hippocampal representation of anticipated future states, and that these representations reflect the statistics of the environment, fundamental properties inherent of model-based systems.

4.1.4 Model-based Pavlovian learning: OFC in the integrates reward-related attributes into default-network derived context-frames

According to our framework, the proactive model of reinforcement learning, we propose that OFC is a likely candidate for integrating stimulus-reward and context-reward information (coming from the integration of amygdalal and hippocampal input) in a way that reward-related information (spatio-temporal, affective, motivational attributes) are integrated into context frames yielded by the function of the default network. As a result, the OFC provides expected reward information to its downstream structures the PPTgN, VTA and VS. This assumption rests on several grounds. First the afferent connectivity of OFC differs from that of the surrounding prefrontal areas, as it receives multimodal sensory input, and afferentation from the anterior cingulate cortex, the dorsolateral prefrontal cortex, the hippocampus, the amygdala and the VS (structures also known to be part of the default network). Accordingly, the OFC is a heterogenous associative area that integrates external and internal information in order to embed multimodal representations in a spatio-temporal context reflecting the monetary and affective value of stimuli. Its efferents are highly intertwined with structures of the reward systems as its glutaminergic neural outflow targets the VS, the VTA, and the PPTgN, structures canonical for model-free learning (e.g. VS, VTA and PPTgN). A role for OFC in driving reinforcement learning may be proposed based on the fact that it is embedded in distinct circuits related to outcome-specific and value-based behaviors, behaviors that depend on instrumental learning and Pavlovian learning, respectively.

4.1.5 Ventral striatum: the role of value function component of reinforcement learning agent in the integration of model-based and model-free input

According to the proactive model of reinforcement learning the VS receives varying afferent inputs, to yield the value function component of the reinforcement learning agent by emitting a value signal that encompasses both reward-related signals and RPE, provided by model-based and model-free inferences, respectively.

As a key nucleus of the mesocortico-limbic system, VS assumes a central role according to the reward prediction error hypothesis that hinges on dopaminergic mesolimbic pathways. Moreover, model-based inferences to the function of VS may be deduced upon the considering the connectivity of VS; showing abundant glutaminergic afferentation rising from the hippocampus, the amygdala and the OFC.

Based on these considerations we suggest the strong interconnection of model-free and model-based learning processes based on the functional connectivity of the VS, as the VS - a likely candidate for computing value - receives input from the dopaminergic neurons of the VTA/ substantia nigra pars compacta (in the form of reward prediction error emitted by model-free account) as well as the hippocampus, amygdala and orbitofrontal cortex (offering reward-related input from the model-based account). This value signal may be embodied by the GABA-ergic efferents emanating through the ventral pallidum, the limbic final common pathway for mesocortico-limbic processing of many rewards.

4.2 Antecedents of the proactive model of reinforcement learning

Our model, the proactive model of reinforcement learning posits that model-free and model-based reinforcement learning are linked via the integrative function of the VS that yields the value function component of reinforcement learning. This concept utilizes the former proactive brain concept, and organizes hippocampal, amygdalal and OFC functioning into a ‘model’, by use of which model-based Pavlovian learning paradigms may be interpreted. Furthermore, we put forward the hypothesis that the default mode network builds the model utilized by model-based reinforcement learning.

Possible support for the default network’s ability to offer a model upon which model-based reinforcement-learning related computations are performed comes from a recent fMRI study, showing the co-operation of the default network and reward-processing regions in generating simulations of desired future outcomes to facilitate decision-making about future goals. The authors interpreted their finding as evidence for the involvement of the default network in goal directed episodic simulations that may help with selecting the most beneficial long-term goals.

Further attribute of the framework is that the VS integrates these varying afferent inputs, and to compile the value function component of the reinforcement learning agent by emitting a value signal that encompasses both reward-related signals and RPE, provided by model-based and model-free inferences, respectively. This proposition is supported by unit recordings in rats that failed to reflect reward prediction error coding in the VS. Instead it was suggested that the dopaminergic signal emitted by the VTA is transformed by ventral striatal processing in a way that it results in a value signal. Actor-critic accounts of model-free learning are also congruent with this proposition that the VS computes value of states, as the function of the critic (known for computing value of states or state-actions pairs and using this value to update the actor’s function) is often confined to the VS.

Summarizing, we offered an integrative account, the proactive model of reinforcement learning, to conceptualize model-free and model-based reinforcement learning by laying out a set of propositions. Also, we articulate that the brain’s default network underlies modeling the environment in a way that it may be accessed by structures involved in model-based learning. Accordingly, the neurobiological connectivity of the DA reward system (e.g. VTA-VS) was overviewed with the aim of accentuating its centrality with respect to afferentation by the amygdala, hippocampus and OFC. To account for making predictions for elucidating what is expected the concept of proactive brain was elaborated in light of the brain's default network function. We theorized that the orbitofrontal cortex offers an input for computing the reward expectations by integrating reward information into context frames developed as a function of the default network and forwards this information to the model-free reinforcement learning structures, and as a result RPE and model-based reward expectations are combined to yield the value signal in the VS.

4.3 The impact of BDNF and irisin on mood disturbance in COPD

The clinical validation of the hypothesis that altered irisin/BDNF axis accompanies disorders conceptualized as the dysfunction of reinforcement learning e.g. mood disorders was done by analyzing data of our COPD cohort. Our present analysis showed a very strong positive association between the Impacts score reflective of depressive mood disturbances in COPD and the reciprocal of serum irisin that was substantially more remarkable in the stratum with BDNF levels lower than the sample median. Analysis by means of multiple linear regression that corrects for possible confounders has confirmed the significant association between Impacts score and serum irisin, furthermore revealed four other significant determinants of Impacts score: BMI, weight, triglyceride level and FEV1% pred, an index of the severity of airflow limitation in COPD.

In addition to reciprocal of irisin, regression coefficients remained significant for BMI, log triglyceride and FEV1% pred in the stratum with lower BDNF levels. However, BMI, weight and FEV1% pred showed significant contribution to the final model in the stratum with higher BDNF levels. These results suggest the presence of an interaction between serum irisin and serum BDNF levels regarding their influence on Impacts score and they underscore our previous hypothesis that serum irisin may exert a peripheral effect reflected by the alteration of metabolic parameters (BMI, weight and serum triglyceride levels) and a central effect related to mood and motivation based on BDNF's action. To the best of our knowledge, this is the first time that the irisin-BDNF axis was assessed with respect to its possible influence on mood disturbances in COPD patients. The ability of FNDC5/irisin pathway to induce BDNF expression in the hippocampus has been reported previously in mice. Our present results seem to support these findings in humans for the first time.

The role of BDNF in depressive disorders has well been articulated by the neurotrophic hypothesis of depression. According to this, depression is based on neurotrophin deficiency of the limbic system, an effect that may be reversed by long-term administration of antidepressants. This hypothesis is closely linked to the neural plasticity hypothesis which postulates that environmental factors (e.g. stress) cause dysfunction of signal transduction cascades involved in neuronal adaptation and plasticity. A candidate pathway is that containing BDNF-cAMP response element-binding protein (CREB), a transcription factor. Change of BDNF plasma levels as well as tissue levels from post mortem biopsies of hippocampus have been described in depressed patients. Furthermore, the cause-effect relationship between BDNF and major depressive disorder was established by a case-control study nested in a cohort of 1276 women aged 75 to 84 years. Using incident cases and controls over the four-year observation period, it was concluded that BDNF is a state marker of major depressive disease based on the longitudinal decrease of serum BDNF levels in this cohort. Corroborating evidence from a systemic review and meta-analysis of twenty publications including 1504 participants, furthermore showed significant correlation between changes of BDNF level and depression score as well as significant increases of BDNF levels accompanied therapy with antidepressants.

Several studies have corroborating evidence for the relationship between markers of disease severity and BMI. Recently, the COPDGene investigators analyzed the data of 3631 spirometry-confirmed COPD patients obtained from a multicenter prospective cohort study. The investigators found significant association between obesity, characterized by higher BMI and worse outcomes including poorer quality of life, dyspnea and reduced 6-minute walk distance. Conversely others have also reported significant correlation between FEV1% (FEV1/FVC) and BMI (Spearman's correlation coefficient 0.255, $p<0.01$). In another study, the influence of metabolic syndrome and its components on the 5-year mortality was assessed in COPD. The authors found that 100 mg/dL increase of plasma triglyceride concentration increases the probability of death over the 5 years by 39% (translating into a hazard ratio of 1.39, CI: 1.06, 1.83). This finding corroborates our result that log triglyceride levels significantly associated with the Impacts score in our final multiple linear regression model.

5 New findings

1. The proactive model of reinforcement learning was developed a novel, integrative framework that accounts for the interaction between model-free and model-based reinforcement learning.

2. Using this new model, the proactive model of reinforcement learning a neurobiological substrate was offered for the model used by model-based reinforcement learning by proposing that this model develops as the result of the continuous function of the default network.

3. Starting from the integrative model to link mood disorder to reinforcement learning it was shown that humoral modulation of reinforcement learning associated neuronal structures, by altered irisin/BDNF my underlie the mood disturbances accompanying COPD.

6 Summary

Interventions focusing on the prevention and treatment of chronic non-communicable diseases are on rise. An integrative theoretical framework, the proactive model of reinforcement learning is proposed as it may contribute to understanding the etiopathogenesis of certain risk behaviors leading to NCDs.

The proactive model of reinforcement learning posits that based on the connectivity of structures attributed to model-based and model-free account, the two approaches for solving the reinforcement learning problems are closely linked and interact. On one hand based on the functional connectivity of VS, model-free and model based RL systems center on the VS that by integrating model-free signals (received as reward prediction error) and model-based reward related input to compute the main substrate of reinforcement learning the value signal. The proactive model of reinforcement learning also offers a neurobiological substrate for the model utilized by model-based reinforcement learning by suggesting that the default network showing extensive functional overlap with contextual associative areas functions in a way that the environment is continuously organized into context frames enabling the formulation of analogy-based association that are turned into predictions of what to expect.

Additional to the neural inferences humoral factors are also important, with irisin, a contraction-regulated myokine formed primarily in skeletal muscle but also in the brain and its downstream mediator BDNF possibly assuming a role. Starting from this we set out to elucidate the possible alteration of the irisin/BDNF axis in a clinical population who suffer from mood disorder, a disorder known to be linked with alteration of the reinforcement learning paradigm in a cohort of COPD patients. Case history, laboratory parameters, serum irisin and BDNF, pulmonary function and disease-specific quality of life (SGRQ) were determined in a cohort of COPD patients (n=74). Simple and then multiple linear regression was used to evaluate data.

We found that mood disturbances are associated with lower serum irisin levels (SGRQ's Impacts score and reciprocal of irisin showed a strong positive association; β: 419.97; CI: 204.31, 635.63; $p<0.001$). This association was even stronger among patients in the lower 50% of BDNF levels (β: 434.11; CI: 166.17, 702.05; $p=0.002$), while it became weaker for patients in the higher 50% of BDNF concentrations (β: 373.49; CI: -74.91, 821.88; $p=0.1$). These results suggest that irisin exerts beneficial effect on mood in COPD patients, possibly by inducing the expression of BDNF in brain areas associated with reward-related processes involved in by depression.

7 Acknowledgements

This research was supported by the Hungarian Brain Research Program (**KTIA_13_NAP-A-V/2**), the Scientific Research Grant of the Hungarian Foundation for Pulmonology (awarded in 2015), the Ministry for National Economy (**GINOP-2.3.2-15-2016-00062** and **AGR-PIAC-13-1-2013-0008**). Writing the thesis was supported by the Ministry for National Economy (**GINOP-2.3.2-15-2016-00005**).

UNIVERSITY OF DEBRECEN
UNIVERSITY AND NATIONAL LIBRARY

Registry number: DEENK/166/2017.PL
Subject: PhD Publikációs Lista

Candidate: Csaba Papp
Neptun ID: EKCKVB
Doctoral School: Doctoral School of Health Sciences

List of publications related to the dissertation

1. **Papp, C.**, Pák, K., Erdei, T., Juhász, B., Seres, I., Szentpéteri, A., Kardos, L., Szilasi, M., Gesztelyi, R., Zsuga, J.: Alteration of the irisin / brain-derived neurotrophic factor axis contributes to disturbance of mood in chronic obstructive pulmonary disease patients.
   Int. J. Chronic Obstr. Pulm. Dis. "Accepted by Publisher", 2017.
   IF: 3.046 (2015)
2. Zsuga, J., Bíró, K., **Papp, C.**, Tajti, G., Gesztelyi, R.: The "proactive" model of learning: integrative framework for model-free and model-based reinforcement learning utilizing the associative learning-based proactive brain concept.
   Behav. Neurosci. 130 (1), 6-18, 2016.
   DOI: http://dx.doi.org/10.1037/bne0000116
   IF: 2.69 (2015)

List of other publications

3. Zsuga, J., Erdei, T., Szabó, K., Lampé, N., **Papp, C.**, Pinter, Á., Szentmiklósi, J. A., Juhász, B., Szilvássy, Z., Gesztelyi, R.: Methodical Challenges and a Possible Resolution in the Assessment of Receptor Reserve for Adenosine, an Agonist with Short Half-Life.
   Molecules. 22 (5), 839-855, 2017.
   DOI: http://dx.doi.org/10.3390/molecules22050839
   IF: 2.465 (2015)
4. Tajti, G., Gesztelyi, R., Pák, K., **Papp, C.**, Kéki, S., Szilasi, M. E., Mikáczó, A., Fodor, A., Szilasi, M., Zsuga, J.: Positive correlation of airway resistance and serum asymmetric dimethylarginine level in COPD patients with systemic markers of low-grade inflammation.
   Int. J. Chronic Obstr. Pulm. Dis. 12, 873-884, 2017.
   DOI: http://dx.doi.org/10.2147/COPD.S127373
   IF: 3.046 (2015)
5. Zsuga, J., Bíró, K., Tajti, G., Szilasi, M. E., **Papp, C.**, Juhász, B., Gesztelyi, R.: 'Proactive' use of cue-context congruence for building reinforcement learning's reward function.
   BMC Neurosci. 17 (70), 70, 2016.
   DOI: http://dx.doi.org/10.1186/s12868-016-0302-7
   IF: 2.304 (2015)
6. Zsuga, J., Tajti, G., **Papp, C.**, Juhász, B., Gesztelyi, R.: FNDC5/irisin, a molecular target for boosting reward-related learning and motivation.
   Med. Hypotheses. 90, 23-28, 2016.
   DOI: http://dx.doi.org/10.1016/j.mehy.2016.02.020
   IF: 1.136 (2015)
7. Pák, K., **Papp, C.**, Galajda, Z., Szerafin, T., Varga, B., Juhász, B., Haines, D. D., Szentmiklósi, J. A., Tósaki, Á., Gesztelyi, R.: Approximation of A1 adenosine receptor reserve appertaining to the direct negative inotropic effect of adenosine in hyperthyroid guinea pig left atria.
   Gen. Physiol. Biophys. 33 (2), 177-188, 2014.
   DOI: http://dx.doi.org/10.4149/gpb_2013079
   IF: 1.173
8. Bíró, K., Bányai, G., Nádházy, Z., **Papp, C.**, Zsuga, J.: Bevezetés az egészségügyi menedzsmentbe. Digitális Tankönyvtár, Budapes, 50 p., 2014.

9. Gesztelyi, R., Kiss, Z. M., Wachal, Z., Juhász, B., Bombicz, M., Csépányi, E., Pák, K., Zsuga, J., **Papp, C.**, Galajda, Z., Branzaniuc, K., Pórszász, R., Szentmiklósi, J. A., Tósaki, Á.: The surmountable effect of FSCPX, an irreversible A1 adenosine receptor antagonist, on the negative inotropic action of A1 adenosine receptor full agonists in isolated guinea pig left atria.
   Arch. Pharm. Res. 36 (3), 293-305, 2013.
   DOI: http://dx.doi.org/10.1007/s12272-013-0056-z
   IF: 1.751

10. Gesztelyi, R., Kiss, Z. M., Zsuga, J., Pák, K., **Papp, C.**, Galajda, Z., Branzaniuc, K., Szentmiklósi, J. A., Tósaki, Á.: Thyroid hormones decrease the affinity of 8-cyclopentyl-1,3-dipropylxanthine (CPX), a competitive antagonist, for the guinea pig atrial A(1) adenosine receptor.
    Gen. Physiol. Biophys. 31 (4), 389-400, 2012.
    DOI: http://dx.doi.org/10.4149/gpb_2012_043
    IF: 0.852

11. Szerafin, T., Leny, A., Palotás, L., Veres, L., Homolay, Á., **Papp, C.**: Hasfali sérv műtéte No-React R eljárással készült borjúpericardium-folt beültetésével.
    Magyar Seb. 61, 61-65, 2008.

12. Miskolczi, S., Vaszily, M., **Papp, C.**, Péterffy, Á.: Tapasztalataink a rekombináns faktor VIIa-val (NovoSeven R) a magas kockázatú szívműtétek vérzéses szövődményeinek kezelésében.
    Magyar Seb. 61 (Suppl.), 45-47, 2008.

13. Szentkirályi, I., Galajda, Z., Jagamos, E., **Papp, C.**, Bodnár, F., Fülöp, T., Péterffy, Á.: Ebstein-anomália - a felnőttkorban végzett műtétekkel kapcsolatos tapasztalatok.
    Orv. Hetil. 145 (45), 2289-2292, 2004.

Total IF of journals (all publications): 18,463
Total IF of journals (publications related to the dissertation): 5,736

The Candidate's publication data submitted to the iDEa Tudóstér have been validated by DEENK on the basis of Web of Science, Scopus and Journal Citation Report (Impact Factor) databases.

01 June, 2017