Active Inference on OpenAI Gym: A Paradigm for Computational Investigations into Psychiatric Illness

Maell Cullen[1+], Ben Davey[1+], Karl J. Friston[2], Klaas Enno Stephan[3], Rosalyn J. Moran*[1,4]

[1]Department of Engineering Mathematics, Merchant Venturers School of Engineering, University of Bristol, 75 Woodland Rd, Bristol BS8 1UB, UK.

[2]Wellcome Trust Centre for Neuroimaging, University College London, 12 Queen Square, London WC1N 3BG, UK.

[3]Translational Neuromodeling Unit, Institute for Biomedical Engineering, University of Zurich and ETH Zurich, 6 Wilfriedstrasse, Zurich 8032, Switzerland.

[4]Department of Neuroimaging, Institute of Psychiatry, Psychology and Neuroscience, King's College London, De Crespigny Park, London, SE5 8AF, UK.

[+] Equal Contributions

*To whom correspondence should be addressed:

Rosalyn Moran,
Department of Neuroimaging,
Institute of Psychiatry, Psychology and Neuroscience,
King's College London,
De Crespigny Park,
London, SE5 8AF,
UK
rosalynjmoran@gmail.com

Artificial Intelligence (AI) has recently attained human-like performance in a number of 'game-like' domains. These advances have been spurred by brain-inspired architectures and algorithms, most notably by deep (hierarchical) filtering and reinforcement learning. OpenAI Gym is an open-source platform to train, test and benchmark algorithms – and provides a range of tasks including classic arcade games such as 'doom'. In this article, we describe how the platform might be used as a simulation, test and diagnostic paradigm for psychiatric conditions. Specifically, we describe how Active Inference (a new AI) can be implemented to 'play' on the OpenAI Gym platform. We show that Active Inference, through epistemic and value-based goals, enables simulated subjects to develop more complete representations of gaming environments as compared to reward only approaches. We show that Active Inference approximates the performance of real human players and can reproduce age-related changes. Crucially, we use Active Inference's proposed mappings from computation to neural implementation to investigate pathophysiology associated with anhedonia. To mimic anhedonia – and its neural substrates – we simulated flat prior beliefs about outcomes that reflected a diminished sensitivity to reward and compared the resulting choice behaviour using motivated agent who believes they will win the game. These simulations produced a particular time course of differences in prefrontal local field potentials and prefrontal BOLD responses during the course of the game. These results speak to a novel game-based imaging biomarker of anhedonia or incipient depression; for example, in large-scale early intervention programs.

Introduction

Recent perspectives on psychiatric illness highlight the crucial role of computational assays in deciphering the complexities of mental illness (Montague, Dolan et al. 2012, Friston, Redish et al. 2017). Computational assays of cognitive and behavioural abnormalities in psychiatric illness provide formal mappings from complex thought disorders to putative neural substrates (Fletcher and Frith 2009). The central proposition here is that cognition and behaviour are emergent features of biological processes and that by capturing these processes formally, we may better access the origins of psychiatric disease. For the purposes of stratifying, understanding and treating neuropsychiatric diseases, several computational frameworks have been deployed recently; including Bayesian learning (Paliwal, Petzschner et al. 2014), drift diffusions processes (Pedersen, Frank et al. 2017) and temporal difference models (Redish 2004). Hierarchical Bayesian models of learning, for example, have been used to differentiate regional brain activations in psychosis with and without auditory hallucinations (Powers, Mathys et al. 2017). At present, clinicians categorize disorders based on symptom lists with broad diagnostic criteria to determine treatments for individual patients. Computational psychiatry aims to refine nosology and treatment, wherein diagnosis and treatment relies not only upon observable traits but should be informed by hidden pathophysiological psychopathological processes (Friston, Redish et al. 2017). In other words, computational approaches aim to access these 'hidden' brain processes. Within the context of psychiatric disorders, this concept is important since, diagnosis and understanding are constructs that refer not to fixed states, but to complex interactions between behavioural phenomena and the world (Rosenman and Nasti 2012). Although observable behaviours may be related conceptually, they may be fundamentally heterogeneous in their aetiology – and show a degenerate mapping with mental processes and empirically observable neuronal features of the disease.

Neuroimaging advances have provided useful biological insights into regional and connectivity deficits associated with mental illness (Hyett, Breakspear et al. 2015). These methods require further development to translate into pragmatic clinical tools for diagnostic and prognostic classifications at the individual level (Tognin, Pettersson-Yeo et al. 2014). Mathematical models that provide both individual brain and behavioural predictions are potentially even more powerful candidates to advance the field of computational psychiatry with appreciable clinical utility. This is because mathematical models that provide both algorithmic (information-processing) and biophysical insights may herald new implementation categories for linking psychopathology and pathophysiology within a single framework (Friston, Redish et al. 2017). Crucially, a prerequisite is that the mechanisms of these putative algorithmic models should have inherent biological plausibility (Doya 2007).

Reinforcement learning (RL) models with model-based fMRI (O'Doherty, Hampton et al. 2007) have predominated in this area, with dopaminergic substrates and decision-making circuits in the striatum highlighted as crucial neural substrates linking aberrant decision making and learning to psychiatric disease symptoms (Daw, Kakade et al. 2002). Using this technique, simulated events, based upon an RL model, are convolved with a hemodynamic response function and correlated against BOLD fMRI signals to associate decision-making processes with the brain regions in which they originate (Guitart-Masip, Huys et al. 2012). The aim is to produce regressors with powerful explanatory capabilities in terms of adaptive reward-seeking and punishment-avoiding behaviours. The appeal of RL models lies predominantly in the expression of *valence* prediction errors and their association with mental disorders such as anhedonia in depression (Huys, Pizzagalli et al. 2013) and impulsivity in ADHD (Sonuga-Barke 2003). Similarly, hierarchal Bayesian analysis has been used in conjunction with model-based fMRI to model prediction errors (Iglesias, Mathys et al. 2013). Here, the focus is not specifically on reward or punishment prediction errors but errors in relation to prior beliefs about states in the world more generally. For example, Ahn et al (Ahn, Krawitz et al. 2011) correlate choice behaviours with decision-time activation in the ventromedial prefrontal cortex (Ahn, Krawitz et al. 2011). Interestingly, when compared with non-Bayesian methods, this technique was found to provide more accurate individual and group estimates and modelling predictions.

The (Bayesian) formalism provided by the free energy principle and active inference provides an account of how the brain models, learns, infers and acts within its environment to minimize long-term surprise (i.e., maximise long-term marginal likelihood or model evidence). In the current setting, resolving the moment to moment free energy of the brain, – comprised of precision-weighted prediction errors – ensures the minimisation of long term surprise (Friston and Kiebel 2009). The free energy principle thus appeals to the dual goals of computational psychiatry as the neurobiological circuits required for this form of 'active inference' overlap with key anatomical features; e.g., of canonical cortical microcircuitry in the sensory domains (Bastos, Usrey et al. 2012). Recently brain regions and models associated with decision-making have been accommodated under active inference for free energy minimization (Friston and Buzsáki 2016).

The main advantage of applying a free energy scheme – for phenotyping behaviour in neuropsychiatric diseases like dementia or schizophrenia – is that it subsumes perception, learning, memory and action; meaning that the parameterizations of a given individual in one task should be able to predict behaviour in many (most) cognitive protocols; providing for consistent comparisons across different tasks and laboratories. This formulation has recently been applied to neuroeconomic games, demonstrating that individual decisions can be best explained by a combination of individual preferences, mathematically represented by prior beliefs and individual confidence metrics. These confidence measures are correspond, mathematically, to a precision or inverse temperature parameter (Friston, Schwartenbeck et al. 2014). Importantly, elements of the update procedure within this modelling framework have been mapped to putative neurobiological components or processes (Schwartenbeck, FitzGerald et al. 2014). This attendant process theory suggests that prior preferences may be encoded in specific cortical regions; e.g., ventromedial prefrontal cortex (Friston, Schwartenbeck et al. 2014), while the inverse temperature or precision parameter may be encoded by neuromodulatory systems, such as dopamine or acetylcholine (Schwartenbeck, FitzGerald et al. 2014). Thus, given a sufficiently simple design for patient populations, where the task structure can be formalised, distinctions between cortical and subcortical effects on behaviour can be distinguished. Importantly the combined algorithmic and process framework afforded by the Free Energy Principle can be applied at an individual level, to characterize individuals with respect to their prior beliefs and preferences by fitting their choice behaviours to a computational model (Schwartenbeck and Friston 2016). Such models deconstruct pathological behaviour in terms of an individual's generative model of the task at hand (usually focusing on their prior beliefs that are a major part of the generative model). The usefulness of this formal approach to computational psychiatry is that psychopathologies are likely to arise from dysfunctional models of the environment, maladaptive learning or failures of inference (Schwartenbeck and Friston 2016).

This paper is concerned with constructing and demonstrating the use of generative probabilistic models that can explain psychopathology – under the free energy formalism – to produce behavioural and imaging features that can be tested empirically at an individual level. Specifically, we aim to demonstrate the contribution of model variables and parameters to perception, inference, learning and behaviour, within game environments that have served as benchmarks in the reinforcement learning community. We use the OpenAI Gym platform as it provides standardized tasks and computational environments, allowing for comparative models of behaviour to be shared across the community. Following the same approach to computational phenotyping described in (Schwartenbeck and Friston 2016), we use generative Markov models of the OpenAI Gym environments; where the form of the model embodies prior beliefs and active inference entails behaviour. By altering the parameters of the generative model and therefore of the inference procedure, we aim to demonstrate how decision making can be altered over the lifespan and under neurobiological changes associated with psychiatric illness. Below we will select 'abnormal' prior expectations about outcomes of a simple arcade game to characterize pathological phenotypes.

In what follows we describe the OpenAI Gym platform and how active inference-based agents can be configured as simulated players. We then compare agents that apply active inference to reward maximising agents and to real human participants of different ages. We finally show how abnormal beliefs can be used to simulate putative pathological signatures from prefrontal cortical neurons, in a simulated model of anhedonia.

Modelling & Methods

The DOOM Environment on OpenAI Gym

Here, we present the DOOM environment provided by the OpenAI Gym (Brockman, Cheung et al. 2016) toolkit. DOOM is a well-known pseudo-3d game that has been used as a platform for reinforcement learning (Kempka, Wydmuch et al. 2016) and computer vision (Mahendran, Bilen et al. 2016). DOOM was chosen to demonstrate the versatility and potential for gaming environments and platforms such as Gym for computational psychiatry. It also provides a simple game scenario for comparisons of free energy and reinforcement-learning, reward-maximising based schemes. The requisite state space is simple to define, while the action space is small and involves a stationary target state. As such, DOOM is appropriate for an active inference-based deconstruction since this formalism is applicable to any paradigm where an agent within a closed environment must perceive, act and make decisions.

Observations are returned from the DOOM environment in 480x640 pixel space as shown in Figure 1a. This observation is split into N 100x640/N sized frames where Nx2 defines the size of our state space (Figure 1b). The location of the target in this space is determined by the Harris Corner detection operator (Harris and Stephens 1988), under the assumption that the target is present within the frame that exhibits the largest variation in (global) pixel intensity Figure 1a. We defined discrete ‘states’ of the environment by the location of the target (monster) relative to the player and whether the agent is currently shooting. In this work, we investigate both a simple 6-state model with 3 positional states and a larger 10-state model with 5 positional states (Figure 1b). In other words, we considered models that carve states of the world with different resolutions.

During play, the agent starts in the centre of the screen at the beginning of each episode. An episode is defined as the period between initialization of a game and the end of the game due to shooting the target or 10 seconds (350 frames) of game time elapsing. The target can be positioned anywhere in front of the agent but constrained to the back wall. The agent can then move through the environment with left and right movements or can emit a shot, what we denote as a ‘fire’ action, i.e. there are three possible actions.

The game metrics used to evaluate the performance of an agent are based on the total amount of reward obtained during an episode and the amount of time it takes to solve the episode. When a ‘fire’ action is submitted from the middle position, i.e. killing the monster, a reward value of 100 is returned. A negative reward penalty (-1) is imposed on every time step, meaning the agent loses points for staying alive while not solving the environment. The optimal solution to the task is thus to move in front of the target and then fire, resulting in a low number of steps and high reward score. In addition, the game returns a ‘survival’ score which serves to index the time taken for the monster to be killed.

Where comparisons between multiple agents are drawn in the following sections, each episode has been seeded such that each agent is presented with the same environment.

Active Inference Agents playing DOOM and Comparison with Reward Maximising Agents

Active inference is a corollary of the free energy principle that provides a normative model of human cognition and behaviour based on Bayesian principles of belief updating and information theoretic constructs of surprise (a.k.a. self information or negative log model evidence). The main tenet of active inference is that action, perception and learning can all be explained in terms of the minimization of variational free energy. An agent operating under this imperative implicitly minimizes its surprise or maximizes its Bayesian model evidence by actively inferring the causal and statistical structure of the world around it (Friston, Mattout et al. 2011). What emerges is an adaptive agent that moves – purposefully – through an environment to solicit outcomes that it believes are most likely (i.e., the least surprising). For example, an agent might believe rewards are likely outcomes and therefore act to minimise surprise by maximising reward. More generally, minimising the surprise expected following an action corresponds to resolving uncertainty. This contextualises goal-directed, reward-seeking behaviour and imposes epistemic, information and -seeking aspect. Crucially, prior beliefs – that determine surprise – implicit in active inference can be used to simulate analogous game play under a putative neuropathological belief. Below we use the prior beliefs about the final state of the game to model anhedonia; i.e., whether one wins or loses.

Active inference rests on a generative model of observed outcomes that can be optimized with respect to free energy. Recent formulations of decision making assume a partially observable Markov Decision Model (POMDP) form of the generative model (Friston, FitzGerald et al. 2016, Pezzulo, Cartoni et al. 2016, Friston, FitzGerald et al. 2017, Friston, Lin et al. 2017). This model defines the joint probability distribution over the observations, hidden states, policies (a sequence of available actions) and precision (or degree of belief in controllability of the environment). A graphical representation of the generative model is illustrated in Figure 1c. Here, a likelihood term establishes a mapping between hidden states and observations, defining the probability of being in a state after making an observation. This is referred to throughout this paper as the A or observation matrix. For our game play we assumed an identity matrix for A, and leave uncertainty about states to emerge in the transition probabilities. These state transition probabilities are represented by the B matrix. Figure 1d illustrates potential trajectories within the DOOM state space where odd numbers denote positional states, while even numbers represent firing within a positional state. For our simulations below, we set each entry in each column of B to have equal values, i.e. the agents did not know the state transitions but had to learn them.

Finally, the mapping between policies and hidden states are also influenced by the agent's prior preferences, which determine how likely or rewarding a given outcome is. This critical quantity is denoted as the C vector and profiles the prior preferences or 'end goal' of the game. In DOOM it should be maximum at the position in front of the monster firing – if prior preferences are defined over as hidden states or simply being rewarded – if prior preferences, defined over outcomes. The probability of a policy also depends on precision γ and its hyperparameters α and β. This quantity is related to the inverse temperature parameter from statistical physics and softmax response functions economics (but is optimized with respect to free energy over play) and determines an agent's confidence in its decisions.

Based upon the current form of the generative model, an action is chosen from a particular policy ($\boldsymbol{\pi}$) where that policy minimizes the expected free energy of the agent (Eqn. 1). To evaluate the expected free energy of a policy and select an action; i.e., whether to move left, move right or fire, the agent first must estimate its current, past and future hidden states under each available policy. In our simulations, we allow the agent to entertain short horizon policies (3-action sequences) and allow all combinations of such 3-action policies to give a total number of policies of 9 e.g. one policy may be {'left', 'left', 'left'}, while another {'right', 'right', 'fire'}. State estimation also minimizes free energy according to Figure 1c. After a set of 16 iterative updates to estimate the states under each policy, a Bayesian Model Averaging procedure is used to construct the final expected states in the past, currently and in the future $\boldsymbol{q}(\boldsymbol{s})$. Using this estimate, the (negative) expected Free Energy of each policy is calculated and passed through a softmax operator to select the current best policy and hence the current optimal action: 'left', 'right' or 'fire'. Within the softmax operation the precision parameter determines the 'temperature' of the decision (where the precision itself is updated at the end of each policy evaluation cycle (Friston, Schwartenbeck et al. 2014)).

Under active inference a policy, π at time t, is valuable (has a high negative expected free energy, $Q(\pi, t)$) when it maximises the expected information (Shewry and Wynn 1987) about the true state of the environment (i.e. maximizing epistemic value – first term expected under current state estimate $\langle \quad \rangle_{q(s)}$) while maximizing extrinsic value (reward of getting to the preferred or believed final state – second term expected under current state estimate).

$$Q(\pi, t) = \langle lnP(o|s, \pi) - lnP(o|\pi) \rangle_{q(s)} + \langle lnP(o) \rangle_{q(s)} \qquad \text{Eqn. 1}$$

Given this form it is easy to contrast active inference with reinforcement learning or simply 'reward maximising' agents. For this comparison (Free Energy Minimizing vs. Reward Maximising), we simply remove the first term (epistemic value) from the evaluation of the policy.

To play the game, we seed the first state, s, using the pixel-based estimate from the DOOM environment and supply the DOOM environment with only the first optimal action from the chosen 3-action policy and repeat. We call this a 'trial'. After each trial, we update the initial state estimate using a new pixel grab from the current Gym frame. A trial will thus form part of an 'episode', defined above, which ends when the monster is fired upon or the game times out.

To simulate DOOM play under active inference, we allow the agent to learn the optimal state transitions. Our learning scheme treats the model's B parameters, encoding transition probabilities as unknown and establishes initial beliefs over these unknowns in the form of prior distributions. Defined as Dirichlet distributions, these B matrices are updated after an action observation cycle by adding a '1' to the correct state to state entry and renormalizing the transition matrix (Figure 1c). Using a set of 'flat' entries as priors (i.e. each element of the matrix is set to 0.167 for a six-state model and 0.1 for the ten-state model) enables us to establish how a simulated agent plays the game with no de-novo knowledge about the environment. As noted above the A matrix is set to the identity, while for the prior belief in final states, we set to C = [0.1, 0, 0.2, 0.6, 0.1, 0] for the six-state model and C = [0.05, 0, 0.05, 0, 0.2, 0.6, 0.05 , 0, 0.05, 0]. These prior beliefs about preferred states might correspond to how the game would be described by a human player; i.e., seek to be in-front of the target (states 3 and 5 for the 6-state and 10-state agents) then fire to win, (states 4 and 6 for the 6-state and 10-state model), while non-firing states are preferable to wasting ammunition (e.g. the zero belief in firing when the monster is on the left or right; states 2 and 6 in the six-state model, Figure 1b). For each artificial agent we simulated learning over 128 episodes and performed 50 runs to evaluate the average behaviour of the agent.

Human Play of DOOM

In order to assess whether our simulated agents could attain 'human-like' performance we collated survival scores from a sample of real players. We measured the performance of 16 players, 8 female aged 37 ± 17 years (mean ± std). To further address the 'human-like' capacity of our synthetic subjects, we assessed the effects of aging on game play. In particular, we collected data from across the lifespan and within our cohort and qualified younger (n=9, 22 ± 1 years) and older (n=7, 56 ± 5 years) adult players for comparison with our 6 and 10-state agents, designed to represent and contrast simpler (older) generative models of the world with more complex (younger) models of the 'DOOM' world.

Each player played 64 consecutive episodes where each episode consisted of self-timed button-press responses and ended either when the monster was killed or when the game timed out (though in practice no human player was timed out). Each episode resulted in a survival score commensurate with the simulated agent's play. The players were told that the goal was to shoot the monster whilst conserving ammunition and that a button would either move the shooter left, right or would result in a 'fire' output. To coincide with the learning simulations of the *in silico* agents, players were not instructed which of three buttons would result in a left, right or fire action. For statistical comparison of younger and older survival scores over time we applied a repeated measures analysis of variance using the measures of survival from each 64 episodes of the game, a post-hoc comparison was used to test for age-group differences.

Ethics was approved by the Faculty of Biomedical Sciences Research Ethics Committee (FREC) at the University of Bristol. No player had a history of neurological or psychiatric conditions.

Manipulations to simulate features of Anhedonia

This framework allows us to investigate pathological or maladaptive decision-making by altering the generative model and simulating play as well as putative neural correlates. Specifically, we will select prior expectations about outcomes or 'goal states' (alterations in the *C* vector) in order to characterize pathological phenotypes. As before we assume uniform prior beliefs about the transitions amongst hidden states of the environment, represented by uniform state transition matrices and the identity matrix maps states to outcomes.

The rationale for the model re-parameterisation is as follows: Anhedonia is a behavioural trait of individuals with depression characterized by a lack interest in rewarding or pleasurable activities (Gard, Kring et al. 2007, Treadway and Zald 2011). The anhedonic aspect of depression has been previously examined in a large reinforcement learning meta-analysis, which compared learning mechanisms with reward sensitivity (Huys, Pizzagalli et al. 2013). There they found that among two alternative hypotheses of anhedonic responses (abnormal learning vs. diminished reward sensitivity), that diminished reward sensitivity most parsimoniously explained a large literature on reward-based reinforcement paradigms in depression. Under active inference, diminished reward sensitivity can be easily represented in our model priors by adjusting the expected outcomes to reflect a less optimistic view of the world and where one may 'finish', in the game of DOOM. By reducing the difference between desired and undesirable states in our *C* vector, we obtain a more uniform distribution. This emphasises potentially 'rewarding' outcomes (i.e. winning the game) and aims to mimic diminished sensitivity to reward. Furthermore, this implicit change in prior preferences may recapitulate the symptoms of depression related to an apparent inability to make decisions – given that all outcomes are now deemed by our agent as equally probable. In summary, the comparative analysis comprises a 'motivated agent' and an 'anhedonic agent' with identical parameters apart from the *C* vector (of prior beliefs in outcomes or 'goals'), illustrated in Figure 5a. We simulated 4 such agents from each class and tested their behaviour and neural responses.

Simulating Neural Responses

Active inference represents both a normative account of brain function ('what is the goal of the brain', Eqn. 1) and a process theory – describing the putative neuronal mechanisms that may perform the computations required. To generate neurobiological predictions related to symptoms of anhedonia – and potential imaging biomarkers for depression – we used the simulations from the agents above and harvested the belief updates over trials across 128 episodes. Each action is based upon inferred states (past, current and future), which then inform policy evaluation and action selection. This requires policies, possible sequences of actions as well as past and potential future states to be 'kept in mind'. Hence, these state updates under the variational scheme may be represented in online, working memory areas such as the prefrontal cortex. We sought to test how alterations in goal states or expected beliefs about outcomes would alter state estimation and the prefrontal responses that could subtend this inference under a 'motivated' and 'anhedonic' set of goals. In our results we illustrate putative local field potentials (LFPs) within the PFC over the iterative updates within and over trials. We assume that each iteration within the state estimation scheme takes 15 msec and filter the state estimation signal from 4 to 32 Hz. To simulate the associated BOLD responses we submitted the LFPs to a Balloon model (Buxton, Wong et al. 1998), using the standard parameter settings in SPM's dynamic causal modelling software (Stephan, Harrison et al. 2004): note that only the differences in goal states (C vector) could influence state estimation and hence the BOLD response.

Results

Active Inference Builds More Complete Representations of DOOM

Figure 2 illustrates how our simulated free energy agent plays the game DOOM (Figure 1a) and how it compares to a classic reward (goal) maximizing scheme, Eqn. 1. In these simulations we report the 'reward scores' (higher better) and 'survival' (lower better) metrics returned by Gym over each episode. We assumed that the agent holds a 6-state representation of the game in mind, (Figure 1b). From 50 runs or instances over 128 episodes we examined how the simulated agents played the game. The agents began at episode 1 with the assumption that states map directly to outcomes (identity *A* matrix), that selecting an action of 'move left', 'move right' or 'fire' will result in moving from any state to any other state with equal probability (equal entries of 0.167 in the *B* matrices) and that the desired state is state 4 (in the middle firing), with firing outside of range (right firing; state 2 and left firing; state 6, see full *C* vector in modelling and methods) the least desirable states (see C vector in modelling and methods). Only policy selection depended on whether we used epistemic and extrinsic value (free energy minimizing) or only the extrinsic value (reward maximising, Eqn. 1) to guide behaviour.

Figure 1a shows one such instance from an agent that is driven by the imperative to minimize free energy and an agent driven by the imperative to maximise reward. Specifically, we illustrate the *B* matrices at episodes 4, 16, 64, and 128. Under free energy, as the agent moves through the episodes we visualise its emerging understanding of the environment and state-action dynamics. Interestingly, around trial 75 (Figure 1b), the agent begins to lose. Here the agent has learned an incorrect state transition; most likely due to a failure in the feature extraction. By the end of the trial, the agent has learned a full and correct representation of the environment, overcoming the earlier erroneous state transitions. The learned state transitions under reward maximization are however demonstrably less robust. Figure 2a shows the reinforcement learning agent's model structure at the end of episode 128. The agent has learned little about the causal structure of the environment, indicated by uniformity of the transition matrices amongst some of its columns. The agent has a lower average reward on this particular instance due to this inability to form an optimal policy for navigating the environment (Figure 1b). This particular instance is reflected across our 50 runs and the lack of environmental structure learned by the reward-only agent is demonstrated at the end of 128 episodes. Here reward 'scores' are significantly lower over instances for reward-based decision making as compared to free energy-based decision making, $p = 0.05$ and for 'survival' scores where free energy agents exit the games earlier compared to reward maximizing agents; $p = 0.04$.

Active Inference vs. Humans in DOOM

In order to assess the ability of a free energy based agent to 'compete' with a human player we enrolled 16 participants to play the game. These participants were instructed to shoot the monster while conserving ammunition – and played 64 consecutive episodes. The participants were also told that three buttons would lead to a right move, left move, or fire response, but were not told the button mapping. In figure 3, we show the survival scores from these real game plays and compare these with the free energy agent, taking every other episode from the 128 simulations above. We found that during the very early trials, the free energy agent is exploring and learning the structure of the environment before engaging in exploitative behaviour. This, alongside learning and then unlearning (Figure 2) maladaptive behaviours, explains the slower transition to behaving optimally. However, matching human performance was remarkably fast, with the free energy agent attaining human-like performance after only 12 actions (Figure 3). This indistinguishable performance was also retained throughout all of the remaining trials ($p > 0.05$), Figure 3.

Complex and Simple Models of the DOOM world and Aging

Given the importance of development and aging in psychiatric disease onset (DeLisi 1997) and recovery (Jeste, Twamley et al. 2003), as well as the notion of model simplification with aging (Moran, Symmonds et al. 2014), we asked whether free energy agents can mimic age-dependent play in our human player cohort. We first described two free energy models that represent complex and simple models. The complex model divided the pixel-based image into a discrete state space that contained representations of 'far left' and 'far right' as well as left and right. This ten-state model was simulated with the goal states described above (see modelling and methods) where the 'preferred' state was to be in from of the monster firing. Our simpler model was the six-state model, which simply represented 'left' and 'right' of the monster without recourse to distance. Above, we show that this simple model performs well (and equally well after 12 decisions) when compared to a human agent.

From our human sample we compared play from 9 younger to 7 older participants, collated from participants in their third decade to those in their sixth and seventh respectively. Using two-way repeated measures ANOVA, we obtained the average survival score from episodes 2-5 (Early 1), 6-10 (Early 2), 31-35 (Late 1) and 36-40 (Late 2), for each run. We found a significant effect of episode ($p < 0.0001$) as well as a significant effect of the size of the state space representation, with the ten-state model outperforming the six-state model on average ($p= 0.026$), Figure 4a. Interestingly, we also observed an interaction of early vs. late epochs with the size of the state space ($p = 0.02$), Figure 4b; where the simpler model performed better on early trials and the more complex model outperforming the six-state model after a protracted period of learning. Interestingly, we found similar effects in our human players. Though an interaction of early and late with age (which we hypothesised might be reflected in model complexity) reached only trend level significance ($p = 0.09$ two-sided student t test), we observed the main effects of learning from early to late trials ($p <0.0001$), as well as an effect of age ($p = 0.29$), Figure 4c and 4d.

Simulating Game Play under Anhedonic Priors

To simulate features of depression in simulated play, we developed a new agent whose belief in final outcomes was relatively flat (Figure 5). This agent represented 10 states and believed they would shoot the monster only marginally more than it believed it would remain to its left or right (Figure 5a). In contrast, our ‘healthy’ or ‘motivated’ agent, while also representing 10 states, retained similar preferences to above, believing that it would end up in front of the monster, shooting it (Figure 5a). From 4 simulations of each agent over 64 episodes, we found that on average, the motivated agent outperformed the anhedonic agent ($p = 0.04$). However, interestingly the anhedonic agent still learned the structure of the environment and sought out wins in later trials, indicating intact learning (data not shown).

Importantly for neuroimaging predictions, we were able to simulate putative neural correlates of these behaviours. Using the variational updates that subtend state inference, where states must be considered in light of all policy options in the past, the present and the future, we simulated the ‘prefrontal cortex’ of these anhedonic and motivated simulated players. Specifically, when examining the state estimate updates within a trial we used filtered time series from the variational updates to represent LFPs and then passed these LFPs through a balloon model to simulate BOLD responses. We found that the amplitude of LFPs from the prefrontal cortex demonstrated a particular temporal excursion in motivated compared to anhedonic agents (Figure 5). Overall LFPs had similar patterns within and over trials, with triphasic potentials for both anhedonic and motivated agents. Importantly this triplet reduced in amplitude for later potentials at a discrete point of learning over the episodes (Figure 5b).

Crucially, the anhedonic agents displayed this qualitative change earlier in the learning episodes. Thus, the difference potentials exhibit an excursion around episode 20, with motivated agents retaining the larger potential triplet for a further 5 trials (Figure 5b). This side-by-side comparison of two agents with different belief structures was replicated in 3 further exemplars (Supplemental Figure 1), suggesting a consistent alteration in state inference strategy and a concomitant change in LFPs that can be systematically predicted and verified; e.g. using a time-frequency analysis of EEG or MEG for a particular player based upon a set of beliefs and behaviour). We then used these LFPs to generate BOLD responses within the PFC. Here the excursion is also marked, with the BOLD response exhibiting a second small peak at around 22 seconds after the beginning of the game, consistent with the timing of when the LFP response exhibits its qualitative change (Figure 5c). Overall, we can compare individuals in terms of their neural responses for alternative beliefs or goals, while recapitulating similar forms at the group level; i.e. over different instances of these healthy and pathological agents (Supplemental Figure 1).

Discussion

Here, we present a treatment of game play using a benchmarking framework – OpenAI Gym – and simulate changes in behaviour and brain responses associated with healthy neurological aging and neuropsychiatric disease. Our simulations are based on the theory that living creatures, including humans, seek to minimise free energy (Friston, Mattout et al. 2011). Importantly, both neurobiological and algorithmic components are interpretable within this framework and so alterations in abstract cognitive constructs such as 'belief' can be mapped directly to their putative neural substrates. This is important for modern computational psychiatry where a key assumption driving many computational deconstructions is that that psychopathologies such as depression or addiction are likely to arrive from maladaptive alterations in neural circuitry which then subtend dysfunctional models of the environment, maladaptive learning or failures of inference (Williams and Dayan 2005, Montague, Dolan et al. 2012).

From our simulations, we find first that unlike an artificial agent that simply seeks to maximise reward, a free energy minimizing agent can develop an internal model of the environment in which it is placed (Figure 2). This is a crucial point, since it is often implicitly assumed that schemes such as reinforcement learning actually 'learn' contingencies in a game or environment, while our results support that this may not be the case. Rather our reward maximization scheme find ‘holes’ in the gaming environment to finish the game in the desired state (Figure 2), without really knowing what is going on. We also compared our simulations to real human players. And although the performance of humans in this scenario will be somewhat trivial, it is important that the performance of free energy-minimizing (and reward maximizing) agents can be put into context in AI Gym in order to assess whether it is, in general, fit-for-purpose. Less trivial environments may be used to stress-test human performance and may also be implemented using our free energy strategy where optimal strategies unknown or difficult to infer and learn. Nevertheless, even in a simple environment we are able to identify learning trends that may be reflected in aging (Figure 4) and neural biomarkers that may represent frank or incipient depression (Figure 5).

It is interesting that our agents perform well compared to human players (Figure 3), despite a simplified 3-dimensional pixel-based view. It is not a trivial task to divide a domain, at any level of abstraction into discrete states *a-priori* but MDPs can be combined with continuous space and time generative models (Friston, Parr et al. 2017), which may offer a scope for future active inference applications. We have manually discretised the game environments used in this work but the discretisation of internal representations through perceptual inference could also be the subject of future work.

Games can provide a basis for testing memory, reasoning, sensory-motor capabilities and attention in individuals regardless of their physical and cognitive abilities or their age, gender or culture. The emergent nature of game play thus presents a constrained complexity that can be used to understand interactions between various determinants of an individual's behaviour. Furthermore, comparing neuroimaging data with behavioural models rather than behaviour allows us to deconvolve complex psychiatric phenotypes that may be used as proxies for the hidden causes that drive aberrant behaviours; where a cause may conceptually encompass anything from an individual's prior beliefs and experiences to their neurobiological idiosyncrasies. In stroke, substance abuse and general-purpose motor rehabilitation (Cevasco, Kennedy et al. 2005, Burke, Tobler et al. 2010, Saposnik, Teasell et al. 2010) game environments have been employed for recovery. Efforts are also being made to translate evidence-based interventions such as behavioural and exposure therapies to computer game formats (Hudlicka 2016). If game environments can be shown to facilitate changes in behaviours it follows that changes in behaviour can be captured and potentially used to identify and monitor patterns of behaviour related to disease onset and progression. Indeed, early clinical signs of psychiatric illness are difficult to distinguish from normal experience and the risk and frequency of 'false positive' diagnoses have been discussed extensively (McGorry, Killackey et al. 2008, Wakefield 2014). Thus, a gaming platform where participants have high engagement and compliance may be a useful adjunct in early intervention programs.

Our simulations of aging were designed to illustrate the general lifespan changes that may need to be considered in applying this sort of platform. The rationale behind the comparison of a 6-state agent and a 10-state agent to potentially recapitulate game play from older and younger adults respectively is based on our earlier work on the free energy principle and aging (Gilbert and Moran 2016). Theory and data support the idea that synaptic loss over the lifespan may offer adaptive pruning, where simpler models instantiated in older brains are driven by top-down prior beliefs (Moran, Symmonds et al. 2014, Gilbert and Moran 2016). This in turn will make older brains more resilient to short term changes in environmental input and provide a more general purpose brain where ‘the gist’ of an environmental challenge is readily identified. This is in contradistinction to younger brains which may over learn unimportant details of the environment’s structure. In our simulation we do indeed find such a pattern with simpler models performing better early in the game, since less structure needs learning (Figure 4).

In simulating features of depression, we choose anhedonia where previous meta-analytic work had highlighted the importance of diminished reward sensitivity but intact learning from reward and punishment (Huys, Pizzagalli et al. 2013). We find that we can simulate a small behavioural deficiency in playing doom, by flattening the prior belief structure. This is a close correlate of diminished reward sensitivity but casts the phenomenon in the future, not the present. Of course, patients with depression do have a diminished optimism about the future (Sharot 2011) though here we aim to demonstrate that goal states represent not only desired outcomes but also believed outcomes. This links the diminished optimism observed in patients to their diminished capacity to perceive alterative outcomes during decision making (Ambady and Gray 2002) – under our framework they are the same thing.

Overall, we demonstrate that active inference may be a more sensitive model of mind as compared to more traditional reinforcement learning models in the literature. We also show that as a dual normative and process theory of the brain, active inference under the free energy principle can reveal structure in behaviour and imaging markers that would allow clinicians and patients to gain a more comprehensive description at the algorithmic and mechanistic level, of mental illness.

Figure Legends

Figure 1. Game structure and State Space Definition

a) Observation from the gym DOOM environment in 480x640 pixel space corresponding to state 1 (top). This observation is cropped to 100x640 pixels, removing image features such as the ceiling and game information to allow more efficient processing of the pixel data (top-middle). Output from Harris Corner detection algorithm with local maxima of the corner response function highlighted in yellow (bottom-middle). An overlay of the positional states on preprocessed image for both 6 and 10 state environments (bottom), note that the size of the centre state is constrained to the size of the target so that the fire action remains effective. b) A graphical representation of the possible state transitions within the 6 (left) and 10 (right) state environments. Green lines denote optimal transitions from each state while red arrows denote possible but sub optimal transitions. Any connection not shown is not possible within the DOOM environment, for example, it is not possible to move from a 'right and firing' state to a 'middle and firing' state without transitioning through a new positional state. c) Separation between the DOOM environment (outer) and the agent's generative model (inner), formulated as an MDP. The inner figure demonstrates how the state transition matrix *B*, observation matrix *A* and prior expectations *C* influence action selection, belief updating and learning. Policy selection depends on the agent's prior preferences *C* and the prescribed precision **γ.** Policies are sampled from a softmax function of their log probability and passed to the DOOM environment as an action, which generates a new observation. The *B* matrix provides a mapping between hidden states under the given action. The *A* matrix maps hidden states to observations and defines the probability of being in a given state after receiving an observation. The state and observation matrix hyperparameters are updated at each state transition in a way that resembles classical Hebbian plasticity. Each update comprises two terms; a digamma function of the accumulated product of expected (post-synaptic) outcomes and their (pre-synaptic) causes. Learning thus equates to updating *B* by accumulating evidence for real state transitions in consequence of action. d) Composition of a 10-state transition matrix that might reflect ‘accurate’ beliefs about the environmental contingencies.

Figure 2. Adaptive Behaviours and Learned Contingencies

a) Simulated agents underwent a single trial (128 episodes) of learning. The *B* matrices shown correspond to the 'Fire', 'Move Right', 'Move Left' actions at t=4, t=16, t=64 and t=128 under the free energy minimization (upper) and reward-maximizing (lower) paradigms. Each matrix represents the agent's belief about how the environment will change after making the respective action. For example, at trial 4 the Reinforcement Learning agent strongly believes that a 'fire' action will bring it from state 3 to state 4. The uniformity of the 'Move Right' matrix at the same time step implies that the agent has no knowledge of the consequence of a 'Move Right' action. After 128 epochs of learning the B matrices of the free energy agent have converged to those of the optimized agent presented in 1B. The B matrix of the reward-based agent is much sparser by comparison, reflecting a lack of knowledge about the environmental contingencies. b) The returned score for each agent in a. c) A further 49 trials of learning were simulated for both agents. The mean total reward achieved by the free energy agents was significantly greater than that of the reward maximizing agent; p = 0.05. Plot shows mean +/- s.e.m.

Figure 3 Comparison of free energy agents and human players

Human players played 64 episodes of the doom game. The upper panel shows the average survival scores (lower better) +/- s.e.m. These were compared to the free energy agents from figure 2 – where alternating trials from the 128 episodes were compared to the human’s 64. The lower panel shows a ‘Manhattan plot’, of statistical difference (-log(p-value) for each episode. After 6 epochs (12 decisions) the free energy agents attain human-like performance.

Figure 4

a) Comparison of 6-state (black) and 10-state (red) free energy agents shown again for every other episode over 128 episodes. Plots denote mean +/- s.e.m. Repeated measures ANOVA revealed a significant effect of time (early vs. late trials) as well as a main effect of model and an interaction between time in episodes and model (see results). b) Comparison with aging effects in human play. Similar to the simulated agents older participants on average, outperform younger participants in early trials (trend in difference between Early 1 and Late 2, p = 0.09). We also observe a significant effect of time and age with the younger participants on average outperforming the older participants.

Figure 5

a) Prior belief structure for the ‘motivated’: green compared to the ‘anhedonic’: blue agents who carry a 10 state model. The anhedonic agent displays a flattened prior belief in the final state of the game, believing with less magnitude that it will kill the monster compared to the motivated agent. Behavioural performance was significantly worse for the anhedonic agents ($p < 0.05$); however, performance matched the motivated agents later across the 64 trials. Example of imposed trial timings for simulated agents. B) Local field potentials derived from state updates that evaluate previous current and future states under all allowable policies. Plotted LFPs are proposed to thus represent the prefrontal cortex. When comparing a single motivated to anhedonic agent the LFPs exhibit large differences around trial 20 and persist over ~5 trials. This finding was replicated in 3 other agent comparisons. c) Based on these LFPs we simulated the associated BOLD response from the PFC and show a second increase in the HRF around 20 seconds for the motivated compared to the anhedonic agents. Anhedonic agents exhibit a more protracted HRF.

References

Ahn, W.-Y., A. Krawitz, W. Kim, J. R. Busemeyer and J. W. Brown (2011). "A model-based fMRI analysis with hierarchical Bayesian parameter estimation." Journal of neuroscience, psychology, and economics **4**(2): 95.

Ambady, N. and H. M. Gray (2002). "On being sad and mistaken: Mood effects on the accuracy of thin-slice judgments." Journal of personality and social psychology **83**(4): 947.

Bastos, A. M., W. M. Usrey, R. A. Adams, G. R. Mangun, P. Fries and K. J. Friston (2012). "Canonical microcircuits for predictive coding." Neuron **76**(4): 695-711.

Brockman, G., V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang and W. Zaremba (2016). "Openai gym." arXiv preprint arXiv:1606.01540.

Burke, C. J., P. N. Tobler, M. Baddeley and W. Schultz (2010). "Neural mechanisms of observational learning." Proceedings of the National Academy of Sciences **107**(32): 14431-14436.

Buxton, R. B., E. C. Wong and L. R. Frank (1998). "Dynamics of blood flow and oxygenation changes during brain activation: the balloon model." Magnetic resonance in medicine **39**(6): 855-864.

Cevasco, A. M., R. Kennedy and N. R. Generally (2005). "Comparison of movement-to-music, rhythm activities, and competitive games on depression, stress, anxiety, and anger of females in substance abuse rehabilitation." Journal of music therapy **42**(1): 64-80.

Daw, N. D., S. Kakade and P. Dayan (2002). "Opponent interactions between serotonin and dopamine." Neural Networks **15**(4): 603-616.

DeLisi, L. E. (1997). "Is schizophrenia a lifetime disorder of brain plasticity, growth and aging?" Schizophrenia research **23**(2): 119-129.

Doya, K. (2007). "Reinforcement learning: Computational theory and biological mechanisms." HFSP journal **1**(1): 30.

Fletcher, P. C. and C. D. Frith (2009). "Perceiving is believing: a Bayesian approach to explaining the positive symptoms of schizophrenia." Nature Reviews Neuroscience **10**(1): 48-58.

Friston, K. and G. Buzsáki (2016). "The functional anatomy of time: what and when in the brain." Trends in cognitive sciences **20**(7): 500-511.

Friston, K., T. FitzGerald, F. Rigoli, P. Schwartenbeck, J. O'Doherty and G. Pezzulo (2016). "Active inference and learning." Neurosci Biobehav Rev **68**: 862-879.

Friston, K., T. FitzGerald, F. Rigoli, P. Schwartenbeck and G. Pezzulo (2017). "Active Inference: A Process Theory." Neural Comput **29**(1): 1-49.

Friston, K. and S. Kiebel (2009). "Predictive coding under the free-energy principle." Philosophical Transactions of the Royal Society B: Biological Sciences **364**(1521): 1211-1221.

Friston, K., J. Mattout and J. Kilner (2011). "Action understanding and active inference." Biological cybernetics **104**(1-2): 137-160.

Friston, K., P. Schwartenbeck, T. FitzGerald, M. Moutoussis, T. Behrens and R. J. Dolan (2014). "The anatomy of choice: dopamine and decision-making." Phil. Trans. R. Soc. B **369**(1655): 20130481.

Friston, K. J., M. Lin, C. D. Frith, G. Pezzulo, J. A. Hobson and S. Ondobaka (2017). "Active Inference, Curiosity and Insight." Neural Comput: 1-51.

Friston, K. J., T. Parr and B. de Vries (2017). "The graphical brain: Belief propagation and active inference." Network Neuroscience **0**(0): 1-34.

Friston, K. J., A. D. Redish and J. A. Gordon (2017). "Computational nosology and precision psychiatry." Computational Psychiatry **1**: 2-23.

Gard, D. E., A. M. Kring, M. G. Gard, W. P. Horan and M. F. Green (2007). "Anhedonia in schizophrenia: distinctions between anticipatory and consummatory pleasure." Schizophrenia research **93**(1): 253-260.

Gilbert, J. R. and R. J. Moran (2016). "Inputs to prefrontal cortex support visual recognition in the aging brain." Scientific reports **6**: 31943.

Guitart-Masip, M., Q. J. Huys, L. Fuentemilla, P. Dayan, E. Duzel and R. J. Dolan (2012). "Go and no-go learning in reward and punishment: interactions between affect and effect." Neuroimage **62**(1): 154-166.

Harris, C. and M. Stephens (1988). A combined corner and edge detector. Alvey vision conference, Citeseer.

Hudlicka, E. (2016). Virtual affective agents and therapeutic games. Artificial Intelligence in Behavioral and Mental Health Care, Elsevier**:** 81-115.

Huys, Q. J., D. A. Pizzagalli, R. Bogdan and P. Dayan (2013). "Mapping anhedonia onto reinforcement learning: a behavioural meta-analysis." Biology of mood & anxiety disorders **3**(1): 12.

Hyett, M. P., M. J. Breakspear, K. J. Friston, C. C. Guo and G. B. Parker (2015). "Disrupted effective connectivity of cortical systems supporting attention and interoception in melancholia." JAMA psychiatry **72**(4): 350-358.

Iglesias, S., C. Mathys, K. H. Brodersen, L. Kasper, M. Piccirelli, H. E. den Ouden and K. E. Stephan (2013). "Hierarchical prediction errors in midbrain and basal forebrain during sensory learning." Neuron **80**(2): 519-530.

Jeste, D., E. Twamley, L. Eyler Zorrilla, S. Golshan, T. Patterson and B. Palmer (2003). "Aging and outcome in schizophrenia." Acta Psychiatrica Scandinavica **107**(5): 336-343.

Kempka, M., M. Wydmuch, G. Runc, J. Toczek and W. Jaśkowski (2016). Vizdoom: A doom-based ai research platform for visual reinforcement learning. Computational Intelligence and Games (CIG), 2016 IEEE Conference on, IEEE.

Mahendran, A., H. Bilen, J. F. Henriques and A. Vedaldi (2016). "ResearchDOOM and CocoDOOM: learning computer vision with games." arXiv preprint arXiv:1610.02431.

McGorry, P. D., E. Killackey and A. Yung (2008). "Early intervention in psychosis: concepts, evidence and future directions." World psychiatry **7**(3): 148-156.

Montague, P. R., R. J. Dolan, K. J. Friston and P. Dayan (2012). "Computational psychiatry." Trends in cognitive sciences **16**(1): 72-80.

Moran, R. J., M. Symmonds, R. J. Dolan and K. J. Friston (2014). "The brain ages optimally to model its environment: evidence from sensory learning over the adult lifespan." PLoS computational biology **10**(1): e1003422.

O'Doherty, J. P., A. Hampton and H. Kim (2007). "Model-based fMRI and its application to reward learning and decision making." Annals of the New York Academy of sciences **1104**(1): 35-53.

Paliwal, S., F. H. Petzschner, A. K. Schmitz, M. Tittgemeyer and K. E. Stephan (2014). "A model-based analysis of impulsivity using a slot-machine gambling paradigm." Frontiers in human neuroscience **8**: 428.

Pedersen, M. L., M. J. Frank and G. Biele (2017). "The drift diffusion model as the choice rule in reinforcement learning." Psychonomic bulletin & review **24**(4): 1234-1251.

Pezzulo, G., E. Cartoni, F. Rigoli, L. Pio-Lopez and K. Friston (2016). "Active Inference, epistemic value, and vicarious trial and error." Learning & Memory **23**(7): 322-338.

Powers, A. R., C. Mathys and P. Corlett (2017). "Pavlovian conditioning–induced hallucinations result from overweighting of perceptual priors." Science **357**(6351): 596-600.

Redish, A. D. (2004). "Addiction as a computational process gone awry." Science **306**(5703): 1944-1947.

Rosenman, S. and J. Nasti (2012). "Psychiatric diagnoses are not mental processes: Wittgenstein on conceptual confusion." Australian & New Zealand Journal of Psychiatry **46**(11): 1046-1052.

Saposnik, G., R. Teasell, M. Mamdani, J. Hall, W. McIlroy, D. Cheung, K. E. Thorpe, L. G. Cohen and M. Bayley (2010). "Effectiveness of virtual reality using Wii gaming technology in stroke rehabilitation: a pilot randomized clinical trial and proof of principle." Stroke **41**(7): 1477-1484.

Schwartenbeck, P., T. H. FitzGerald, C. Mathys, R. Dolan and K. Friston (2014). "The dopaminergic midbrain encodes the expected certainty about desired outcomes." Cerebral Cortex **25**(10): 3434-3445.

Schwartenbeck, P. and K. Friston (2016). "Computational phenotyping in psychiatry: a worked example." eneuro **3**(4): ENEURO. 0049-0016.2016.

Sharot, T. (2011). "The optimism bias." Current Biology **21**(23): R941-R945.

Shewry, M. C. and H. P. Wynn (1987). "Maximum entropy sampling." Journal of Applied Statistics **14**(2): 165-170.

Sonuga-Barke, E. J. (2003). "The dual pathway model of AD/HD: an elaboration of neuro-developmental characteristics." Neuroscience & Biobehavioral Reviews **27**(7): 593-604.

Stephan, K. E., L. M. Harrison, W. D. Penny and K. J. Friston (2004). "Biophysical models of fMRI responses." Current opinion in neurobiology **14**(5): 629-635.

Tognin, S., W. Pettersson-Yeo, I. Valli, C. Hutton, J. Woolley, P. Allen, P. McGuire and A. Mechelli (2014). "Using structural neuroimaging to make quantitative predictions of symptom progression in individuals at ultra-high risk for psychosis." Frontiers in psychiatry **4**: 187.

Treadway, M. T. and D. H. Zald (2011). "Reconsidering anhedonia in depression: lessons from translational neuroscience." Neuroscience & Biobehavioral Reviews **35**(3): 537-555.

Wakefield, J. C. (2014). "Wittgenstein's nightmare: why the RDoC grid needs a conceptual dimension." World Psychiatry **13**(1): 38-40.

Williams, J. and P. Dayan (2005). "Dopamine, Learning, and Impulsivity: ABiological Account of Attention-Deficit/Hyperactivity Disorder." Journal of Child & Adolescent Psychopharmacology **15**(2): 160-179.