Deficient reinforcement learning in medial frontal cortex as a model of dopamine-related motivational deficits in ADHD

Massimo Silvetti[1,2], Jan R. Wiersema[3], Edmund Sonuga-Barke[3,4], Tom Verguts[1,2]

[1] Department of Experimental Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium.

[2] GIfMI (Ghent Institute for Functional and Metabolic Imaging), Ghent University Hospital, De Pintelaan 185, 9000 Ghent, Belgium

[3]Department of Experimental Clinical and Health Psychology, Ghent University, Henri Dunantlaan 2, 9000 Ghent, Belgium.

[4] Department of Psychology, Shackleton Building (B44), University of Southampton, Highfield Campus, Southampton SO17 1BJ, UK.

Corresponding author: Massimo Silvetti

Department of Experimental Psychology,

Ghent University, Henri Dunantlaan 2,

9000 Ghent, Belgium.

GIfMI (Ghent Institute for Functional and Metabolic Imaging)

Ghent University Hospital, De Pintelaan 185B

9000 Ghent, Belgium

e-mail: massimo.silvetti@ugent.be

Keywords: ADHD, ACC, reinforcement learning, dopamine, prediction error, reward expectation

Abstract

Attention Deficit/Hyperactivity Disorder (ADHD) is a pathophysiologically complex and heterogeneous condition with both cognitive and motivational components. We propose a novel computational hypothesis of motivational deficits in ADHD, drawing together recent evidence on the role of anterior cingulate cortex (ACC) and associated meso-limbic dopamine circuits in both reinforcement learning and ADHD. Based on findings of dopamine dysregulation and ACC involvement in ADHD we simulated a lesion in a previously validated computational model of ACC (Reward Value and Prediction Model, RVPM). We explored the effects of the lesion on the processing of reinforcement signals. We tested specific behavioural predictions about the profile of reinforcement-related deficits in ADHD in three experimental contexts; probability tracking task, partial and continuous reward schedules, and immediate versus delayed rewards. In addition, predictions were made at the neurophysiological level. Behavioural and neurophysiological predictions from the RVPM-based lesion-model of motivational dysfunction in ADHD were confirmed by data from previously published studies. RVPM represents a promising model of ADHD reinforcement learning suggesting that ACC dysregulation might play a role in the pathogenesis of motivational deficits in ADHD. However, more behavioral and neurophysiological studies are required to test core predictions of the model. In addition, the interaction with different brain networks underpinning other aspects of ADHD neuropathology (i.e., executive function) needs to be better understood.

1. Introduction

Attention-deficit/hyperactivity disorder (ADHD) is a prevalent neuropsychiatric disorder marked by persistent and pervasive symptoms of inattention, hyperactivity and impulsivity. It is associated with substantial negative impact on everyday life and academic performance (Taylor & Sonuga-Barke, 2008). It is a heterogeneous and complex condition at clinical (Martel, Roberts, Gremillion, von Eye, & Nigg, 2011), etiological (Brookes, et al., 2008) and pathophysiological (Sonuga-Barke, Bitsakou, & Thompson, 2010) levels. The acceptance of pathophysiological heterogeneity has led to the development of multiple pathway accounts (Durston, van Belle, & de Zeeuw, 2011; Nigg & Casey, 2005; Sonuga-Barke, 2002). The current paper is dedicated to understanding motivational deficits in ADHD linked to mesolimbic dopamine deficit (Volkow, Wang, & Baler, 2011). We used a neurocomputational approach to; construct an explicit theory of ADHD pathophysiology; explain part of the existing data; and generate explicit experimental predictions. Although the empirical "dust" on anterior cingulate cortex (ACC) dysfunctions in ADHD has not "settled", it is useful to generate such an explicit theory for categorizing existing data and guiding future empirical work. The role of neurocomputational models in ADHD research (and in neuroscience in general) is creating a bridge between the microscopic level of cellular-receptor findings and the macroscopic level of electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and behavioral findings. Such a bridge should help providing a systematic framework to help understand the pathogenesis of ADHD and to generate explicit predictions for experimental testing. Here we propose a novel computational hypothesis drawing together recent evidence on the role of ACC and associated

mesolimbic dopamine circuits in both reinforcement learning (RL) and ADHD. In particular we simulated a lesion in a previously validated RL computational model (Reward Value and Prediction Model, RVPM; Silvetti et al., 2011, 2012) on ACC and its interactions with brainstem dopaminergic nuclei. The lesion simulated a dopamine signaling deficit, based on pharmacological, neurophysiological and neuroimaging findings in ADHD patients (Swanson, et al., 2007).

1.1 RL impairment in ADHD

Theoretical accounts identifying RL-related deficits as being at the core of ADHD patients' altered motivation have received considerable attention recently. The supporting evidence implicating putative RL-related brain networks is growing (Luman, Tripp, & Scheres, 2010), though behavioral evidence from classical RL schedules is currently limited (Sonuga-Barke, 2011). Nonetheless, several studies have documented impaired performance in ADHD patients during RL tasks, like probability tracking (Frank, Santamaria, O'Reilly, & Willcutt, 2007a; Luman, et al., 2009) or reward temporal discounting (Scheres, Tontsch, Thoeny, & Kaczkurkin, 2010). Indeed, although the literature lacks complete consistency (Scheres, et al., 2006), ADHD patients typically prefer immediate small over delayed large rewards, showing a steeper temporal discount curve (Marco, et al., 2009; Sagvolden, Aase, Zeiner, & Berger, 1998; Scheres, Lee, & Sumiya, 2008; Scheres, et al., 2010). The latter characteristic is possibily related mainly to the hyperactive/impulsive dimension of ADHD (Scheres, et al., 2010), which could explain why some studies reported null results. Altered reinforcer sensitivity is reflected in altered electrophysiological activity related to RL processing (Groen, et al., 2008; Herrmann, et al., 2010; van

Meel, Heslenfeld, Oosterlaan, Luman, & Sergeant, 2011; van Meel, Oosterlaan, Heslenfeld, & Sergeant, 2005).

RL models of motivational dysfunction in ADHD have focused on the reward circuitry modulated by mesolimbic dopamine branches (Luman, et al., 2010). Ventral tegmental area (VTA), in the mesencephalic brainstem, is the major source of dopaminergic input to subcortical-limbic structures and medial frontal cortex (mesolimbic pathway) and to dorsolateral cortical areas (mesocortical pathway) (Oades & Halliday, 1987). The Mesolimbic pathway modulates reward processing and motivation (Rushworth, 2008; Wise, 2002, , 2004), while the mesocortical pathway is involved in executive processes (McNab, et al., 2009; Sonuga-Barke & Fairchild, 2012; Vijayraghavan, Wang, Birnbaum, Williams, & Arnsten, 2007). Within the reward circuit, most attention has focused on dysregulation of ventral striatum (VS) and orbitofrontal cortex (OFC). fMRI studies showed a hypofunctioning in ADHD of VS to cues predicting future rewards (Carmona, et al., 2011; Hoogman, et al., 2011; Scheres, Milham, Knutson, & Castellanos, 2007; Strohle, et al., 2008). Furthermore, in ADHD, OFC activation is altered during the delivery of signalled rewards (Strohle, et al., 2008) and during reinforcement in attentional tasks (Cubillo, Halari, Smith, Taylor, & Rubia, 2011).

1.2 Dopamine Impairment in ADHD

Consistent with the hypothesis of RL-related impairments in ADHD, PET studies (Volkow, et al., 2009) have demonstrated a dopaminergic transmission deficit in mesolimbic and nigrostriatal systems in untreated adult patients. In particular, a reduced density of both postsynaptic dopamine receptors (subtype D2/D3) and dopamine transporters has been demonstrated. Other PET studies have found

impaired functioning of dopamine systems in ADHD patients in the nigrostriatal pathway (Volkow, et al., 2007). Building on the work of Schultz and colleagues (Schultz, 1998), models of motivational deficits in ADHD have hypothesized a lack of dopaminergic signal transfer from actual rewards to preceding events that reliably predict the future reward (Dopamine Transfer Deficit theory, DTD) (Tripp & Wickens, 2008). According to the DTD theory, ADHD patients are impaired in assigning the correct value to events that predict future rewards. A second model postulates that a low level of tonic dopamine causes a steeper temporal discounting slope (Dynamic Developmental Theory, DDT) (Sagvolden, Johansen, Aase, & Russell, 2005). The DDT differs from the DTD theory in emphasizing the role of temporal delay in RL impairments in ADHD, rather than the disruption of the learning mechanisms themselves, leading individuals with ADHD to prefer small immediate rewards over large delayed ones. At the physiological level these models give rise to different predictions. In particular, the DTD theory is able to explain the phasic VS hypoactivation for reward predicting cues (Scheres, et al., 2007), while the DDT can account for both tonic and phasic hypodopaminergic state of VS (Volkow, et al., 2009; Volkow, et al., 2007).

1.3 ACC dysfunction in ADHD

Within these RL models of ADHD motivational deficits, little attention to date has been paid to a core element within the dopamine mesolimbic pathway: ACC. ACC and VS together comprise the key components of the cingulate loop through the basal ganglia (G. E. Alexander, DeLong, & Strick, 1986). ACC plays a pivotal role in RL (Botvinick, 2007; Rushworth, 2008). It is thought to be implicated in the computation of reward expectations linked to actions or environmental stimuli, and

calculate the difference between such expectations and the actual environmental outcomes (prediction error) (Amiez, Joseph, & Procyk, 2005; Jessup, Busemeyer, & Brown, 2010; Kennerley, Behrens, & Wallis, 2011; Matsumoto, Matsumoto, Abe, & Tanaka, 2007; Oliveira, McDonald, & Goodman, 2007). In ADHD patients, EEG studies have a diminished error related negativity (ERN) and error positivity (Pe) (Groen, et al., 2008; Herrmann, et al., 2010; Wiersema, van der Meere, & Roeyers, 2009). Given that ERN and Pe are typically ascribed to ACC (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993; Herrmann, Rommler, Ehlis, Heidrich, & Fallgatter, 2004; Posner & Dehaene, 1994), this suggests an impairment of ACC in ADHD patients. In addition, dopamine agonists restore the Pe amplitude (Groen, et al., 2008; Jonkman, van Melis, Kemner, & Markus, 2007). Moreover, neuroimaging studies document metabolic hypoactivation (Bush, et al., 1999; Rubia, et al., 1999) and hypotrophy of ACC (Makris, et al., 2007; Makris, et al., 2010; Seidman, et al., 2006). Finally, ACC hypotrophy extends beyond the cortex to the underlying white matter (Amico, Stauber, Koutsouleris, & Frodl, 2011), probably causing an impairment of ACC connectivity (Castellanos, et al., 2008).

Here we propose a new computational theory bridging the dopaminergic deficits in ADHD patients, the RL-related behavioral findings, and the role of ACC dysfunction. Our core hypothesis is that mesolimbic dopaminergic transmission impairment in ADHD leads to a malfunctioning of RL processing by ACC. In particular, it disrupts the ability of patients to predict future rewards by ACC and constrains the updating of such predictions according to environmental outcomes (the "Critic" function in the RL framework; Sutton & Barto, 1998). In the current paper this model is instantiated using simulated lesions of the RVPM (Silvetti, Seurinck, &

Verguts, 2011, , 2012). We have previously demonstrated that the RVPM provides a unified explanation of numerous and apparently heterogeneous experimental findings relating to diverse ACC functions. At the same time, it provides a computational explanation of the temporal difference (TD) signature of brainstem dopaminergic neurons during RL (Schultz, 1998). The RVPM belongs to a new class of models aimed to interpret ACC functions in terms of RL operations (W. H. Alexander & Brown, 2011). It is worth stressing that the RVPM was developed to account for ACC functions, and not specifically as a model of ADHD pathophysiology. For this reason, this work represents an interesting generalization of our model predictions.

In the following sections of this article, we; a) describe simulations with an impaired (reduced) dopaminergic transmission in RVPM; b) make a number of predictions about behavioural performance on specific tasks which we then test against existing experimental findings on ADHD; c) investigate the performance of the RVPM in three different experimental settings. The first setting concerns behavioral and neurophysiological consequences of dopaminergic impairment in RL in uncertain environments (probability tracking tasks with reinforcement schedule). In the second simulation we investigated the effect of dopaminergic impairment on temporal discounting of rewards (Sagvolden, et al., 1998). In the third simulation, we examined how the reward schedule (continuous vs. partial) influences the performance of the lesioned system (ADHD) (Luman, Oosterlaan, & Sergeant, 2005). Our goal is to generate predictions at both the behavioural and neurophysiological levels, test these against existing data where such are available, and to stimulate new empirical studies when these data are not available.

2. Model structure and dynamics

Figure 1 summarizes the RVPM structure. All technical details are reported in Silvetti et al. (2011) (see also Appendix). The parameter settings of the RVPM remained constant for all the simulations and were the same as in Silvetti et al. (2011). The RVPM consists of three neural modules. The first one (labeled CUE) codes for events external to ACC (or *cues*, neural units C1 and C2). These events can correspond either to stimuli, planned actions, or more generally, to options between which the individual can choose. ACC has consistently been shown to compute RL operations linked to both actions and stimuli (Amiez, Joseph, & Procyk, 2006). The second module simulates ACC itself. This module contains distinct neural units estimating reward expectations (V unit) (Amiez, et al., 2006; Kennerley, et al., 2011), and neural units coding for the difference between these expectations and actual environmental outcomes (prediction errors, δ units) (Kennerley, et al., 2011; Matsumoto, et al., 2007). The ACC module estimates the reward expectations linked to each external event (stimulus or action), so that the activation of each of the two neurons coding for cues (C1, C2) is followed by the response of the V neuron. This response will be proportional to the value (expected reward) of the corresponding event. Reward expectations coded by V are used as input to a decision making system (SOFTMAX) that selects one of the options (C1 or C2). Finally, the third module simulates the brainstem (VTA) dopamine neurons. The latter module consists of two different neural units. The first provides a dopamine signal coding for the "raw" reward (RW) (Ljungberg, Apicella, & Schultz, 1992). The second (temporal shifting neuron, TSN) receives input from the ACC module. It simulates dopamine neurons exhibiting a TD signature, i.e. shifting of the dopamine signal from reward onset to the onset of reward-predictive cues (Schultz, 1998). Although the latter has been

highlighted in earlier computational work (Montague, Dayan, & Sejnowski, 1996), both neuron types have indeed been observed in VTA, with a minority (10-20%) preserving reward-locked activity after several weeks of conditional training (Ljungberg, et al., 1992; Takikawa, Kawagoe, & Hikosaka, 2004). However, most biological VTA units show a mixed behavior between these two extremes, so the presence of units in the model coding purely for primary rewards could be debated. From the computational viewpoint this is not an issue for the RVPM. What is crucial for the model is the presence of a signal encoding primary rewards steadily along time. This means that even if the biological dopaminergic units responding to primary reward were actually units just showing a very slow and incomplete TD signature, it would have no impact on the main assumptions founding the RVPM. Figure 1 shows the interaction between the RVPM and environment (dark blue arrows). In addition, it displays the environmental variables we used in the three simulations (light blue box).

Figure 1 about here

2.1 Actor-Critic framework

In RL terminology, the RVPM is a *Critic*, i.e. a system that evaluates actions or stimuli in terms of reward expectations. To simulate behavior, we additionally need an *Actor* which selects options based on the Critic's evaluations. We implemented the *Actor* by means of a SOFTMAX function (Figure 1; see also Appendix).

2.2 Simulation methods

Each trial of each simulation consisted of a choice between two options. The model was required to learn by trial and error which option led to the highest long-term

reward (Figure 1). In each trial, one of the two options (units C1 or C2) was chosen by the SOFTMAX, based on their respective expected values (Equation A5, Appendix). In all the simulations we compared the performance of the intact model with that of the model having defective dopamine transmission (ADHD simulation). This lesion was simulated by decreasing the amplitude of the reward signal coded by the RW unit. The environmental variables we used in the three simulations are summarized in Figure 1 (light blue box).

3. Simulation Results

3.1 Simulation 1: Probabilistic choice tasks

ADHD subjects have difficulties discovering the best option based on environmental outcomes (Frank, et al., 2007a; Luman, et al., 2009). Further, during learning, ADHD patients switch between options more than controls (Frank, et al., 2007a; Luman, et al., 2009). Here we administered a probabilistic choice task, in which two possible options were rewarded with different probabilities (75% and 25%, see Appendix), to both our ADHD (RVPM with reduced dopaminergic signal) and control model (intact RVPM). At the same time we recorded the simulated neural activity from the RVPM, in order to compare it to the experimental neurophysiological data.

Figure 2a shows the performance of ADHD and control groups after extensive training. The simulated ADHD group was less efficient in choosing according to the reward contingencies, with frequent switches toward the less rewarded option (40% vs. 60% for the most rewarded option). The control group, on the contrary, showed a clear preference for the most rewarded option (20% vs. 80%; comparison between the percentage of selection of the most rewarded option in the two groups: $t(19)=9.72$, $p<0.0001$). Figure 2b shows the percentage of switches between options during the

learning phase. Consistent with experimental data (Frank, et al., 2007a; Luman, et al., 2009) the ADHD group switched more often ($t(19)=2.72$, $p=0.013$), as a consequence of difficulty in learning the reward contingencies. This result cannot be attributed to differences in the action-selection process itself (SOFTMAX), because in both groups the action selection system had the same parameters.

Figure 2 about here

According to our model, the impaired dopaminergic reward signal prevents complete learning of reward expectations from reward history (lower V unit activation). This mechanism can be documented by the analysis of RVPM neural activity during this task. Figure 2c shows the activations of the V unit after cue selection and before the outcome period (when the reward was given or omitted). In the model, ADHD subjects formulated lower reward expectations than controls ($t(19)=13.19$, $p<0.0001$). In humans this effect has been shown in VS (Scheres, et al., 2007), but not yet in ACC. More precisely, Scheres et al. (2007) did not find any ACC activation related to reward expectation, not even in normal controls. One reason may be that the cue in the paradigm used by Scheres and co-workers was independent of the action, whereas ACC seems primarily concerned with action value (Rushworth & Behrens, 2008). This remains an issue for future research. Figure 2d displays the simulated outcome-locked ACC activity in error trials (i.e. when the choice was unrewarded) in both groups (intact versus lesioned model, ($t(19)=14.21$, $p<0.0001$). Again the RVPM activation is consistent with experimental data from ADHD patients, who exhibit a reduced amplitude of error related activity in EEG studies (Groen, et al., 2008; Herrmann, et al., 2010; van Meel, et al., 2011), if we

assume error-related activity to be due to reward omission (Holroyd & Coles, 2002). This simulation result is due to impaired reward expectations; indeed, a violation of a low reward expectation leads to a lower negative prediction error (δ^-). The RVPM also had reduced ACC activation for correct trials in ADHD ($t(19)=22.40$, $p<0.0001$) (Figure 2e). This provides another novel experimental prediction to test in the future. In this case, the impaired encoding of correct outcomes (positive prediction error, δ^+) is a direct consequence of dopamine dysfunction, which prevents optimal learning of reward expectations (Figure 2c).

Finally, Figure 2f-g shows the activation of the TSN neuron in the VTA module. In the control case, there is a shift of a dopamine signal from reward to cue onset with progressive learning (Figure 2f), consistent with the classical data of Schultz et al (1998). In contrast, in the ADHD model, not only is there less signal at reward onset early during learning, but, in addition, the signal does not shift back in time with progressive learning (Figure 2g). This is shown by an interaction *group (ADHD, Controls) × epoch (cue, feedback)* comparing the VTA activity at the end of learning ($F(1,19)=54.39$, $p<0.0001$; red plots in Figure 2f-g). This finding is consistent with the DTD theory of Tripp and Wickens (2008).

3.2 Simulation 2: Reward temporal discounting in ADHD

ADHD patients discount delayed rewards at a higher rate than controls, showing a steeper decay of reward value estimation as a function of the delay to reward (i.e., temporal discounting) (Demurie, Roeyers, Baeyens, & Sonuga-Barke, 2012; Scheres, et al., 2008; Scheres, et al., 2010). We compared the performance of the ADHD (i.e., lesioned) and control RVPM model in a modified version of the probabilistic choice

task, in which the system chose between two options differing in delay prior to reward (Figure 1 box, see also Appendix).

Figure 3 shows the percentage of choices for the delayed reward as a function of the delay duration. At the first data point both choices were rewarded after 1s of delay. With this reward delay, the system did not show a preferred choice, and there was no difference between control and simulated ADHD patients (95% confidence interval). The lesioned ADHD RVPM displayed greater sensitivity to delay, as shown by the steeper discount curve (Figure 3), an effect that has been widely documented in ADHD patients (Sonuga-Barke, Sergeant, Nigg, & Willcutt, 2008).

Figure 3 about here

In the intact RVPM a reward-related delay evokes a negative prediction error (Silvetti, et al., 2011). This negative prediction error ($\bar{\delta}$) reduces the reward expectation linked to the option leading to delayed reward. This explains why in the intact RVPM (i.e., control model) the probability of selecting the delayed option decreased as a function of delay. The probability of selecting the delayed option does not fall to zero as the system (like biological agents) always maintains a certain amount of exploratory behavior because of the stochastic Softmax action selection module. The choice mechanism interacts with impairments in the reward expectation due to the disruption of dopaminergic input to the ACC module (ADHD group, see also Figure 2a) during learning. As a consequence, long delays create a further disruption of reward expectation in the lesioned RVPM. For this reason, the probability of choosing a delayed reward was much lower in the ADHD group than in the control group. Hence, the RVPM proposes that delay aversion is a consequence of

impairments during learning of the choice options, i.e. when the contingencies are experienced (Scheres, et al., 2008).

3.3 Simulation 3: Continuous vs. partial reward schedules

Some experimental findings have shown a differential impact of reinforcement different schedule types on ADHD performance. Though inconsistencies remain, some studies have documented improved performance when ADHD patients were trained using a continuous (i.e. when each correct choice is rewarded) compared to a partial reward schedule. Some of these showed a specific enhancement of ADHD performance for continuous schedules compared with control subjects (Barber, Milich, & Welsch, 1996; Douglas & Parry, 1994; Freibergs & Douglas, 1969; Luman, et al., 2005; Parry & Douglas, 1983).

In this simulation we compared the performance of the lesioned and intact RVPM under different reinforcement schedules (continuous versus partial). In the continuous reinforcement schedule, one choice was rewarded 100% and the other 0%. For the partial reinforcement schedule, we used the performance modeled in Simulation 1 (75% and 25% reward rates). Figure 4b shows the difference between control and ADHD simulations under the continuous reinforcement schedule; the difference is indeed smaller than in Simulation 1 (data reproduced in Figure 4a, partial reinforcement schedule). This result is documented by the 2-way interaction *group (*control *vs.* ADHD*)* × *reward schedule* (continuous vs. partial) ($F(1,159)=5.87$, $p=0.017$) with percentage of best choice as dependent variable.

Figure 4 about here

4. Discussion

We presented a novel computational theory of the pathophysiological underpinnings of motivational deficits in ADHD patients based putative dopamine transmission deficits. This was based on empirical findings of impaired dopaminergic transmission in the mesolimbic pathway, altered RL-related processes and structural and functional ACC alterations in ADHD, and the emerging role of the ACC in RL. Figure 5 summarizes this hypothesis. In the model, the core deficit in ADHD is hypothesized to be a reduced dopamine signal from VTA to ACC (red arrow). This creates deficits in the ACC role of "Critic" in RL computations, bringing about disruptions of environmental outcome predictions, which in turn leads to maladaptive decisions and to disrupted actions. Finally, a reduced feed-forward dopaminergic reinforcement signal from VTA to ACC not only creates deficits in RL-related computations, but also causes disruption of the TD signature of dopamine neurons in VTA, which is in agreement with the predictions of the DTD theory of Tripp and Wickens (2008). It is worth noting that the RVPM computational explanation of DTD is different from the original DTD theory in the identification of the causes of this phenomenon (mesolimbic dopamine deficit), and in also predicting an impaired dopamine response to primary rewards, as actually has been found in humans (Volkow, et al., 2009). Moreover our results could be considered a bridge connecting the DTD and DDT models of ADHD motivational deficits, as we showed in Simulation 1 how a dopaminergic deficit (compatible with DDT) could cause a deficit in dopamine temporal shifting (compatible with DTD) and that both of these conditions are necessary to provide a neurocomputational explanation of the motivational characteristics of ADHD.

Figure 5 about here

The simulation of ADHD in the lesioned RVPM has the potential to provide a unified and computationally explicit account of motivational deficits in patients affected by dopaminergic disruption. The proposal of a pivotal role of ACC in RL-related ADHD pathogenesis provides a testable hypothesis about why ADHD patients are impaired in decision making, why this mainly occurs under partial reinforcement schedules, and why they exhibit temporal discounting (Demurie, et al., 2012)and delay aversion (Marco, et al., 2009). At the same time this theory provides an explanation of several physiological findings, such as the reduction of ERN and Pe in EEG studies in ADHD, their recovery after administration of dopaminergic medication (Groen, et al., 2008; Jonkman, et al., 2007), and the deficit in the mesolimbic pathway driving reward expectations. The RVPM simulations also led to two novel experimental predictions for ADHD: a reduction of ACC activity in correct trials (e.g., the correct-related negativity, CRN) (Roger, Benar, Vidal, Hasbroucq, & Burle, 2010), and a reduced ACC activity linked to reward expectations. The latter neurophysiological predictions are part of the causes, from a behavioral viewpoint, of impaired reward-based decision making showed in Simulation 1. In the remainder of the paper, we discuss related issues and models of ADHD.

4.1 ACC versus striatal dopamine deficit

Dopamine deficit in ADHD patients is typically investigated and documented within mesolimbic and subcortical structures (e.g. VS, hypothalamus) and not yet in cortical structures like the ACC. For example, Volkow et al. (2009) selected specific subcortical regions of interest to investigate dopamine receptor density in ADHD patients. Considering that reward processing signals were documented in the VS (see

paragraph 1.2), the dopaminergic deficit in the VS could be considered a valid anatomical alternative for our RL based perspective of ADHD pathogenesis. In this work we advanced the novel proposal of ACC-based ADHD pathogenesis for two main reasons. The first one comes from theoretical neuroscience. In the *Introduction* we highlighted how ACC has recently begun to be considered as being pivotal in RL, in particular, as a "Critic" deputized to compute expectations about environmental outcomes. Such a pivotal role makes ACC malfunction an interesting candidate for explaining several ADHD characteristics in relation to motivational components, providing also a computational account to the alleged dopamine transfer deficit in ADHD (Simulation 1). The second reason is more empirical and is based on the large amount of data indicating ACC anatomo-functional anomalies in ADHD patients (see Introduction section), which have not been specifically addressed by earlier theories of ADHD neuro-pathogenesis. Finally, although our perspective is based also on the assumption of a reduced dopamine level in the ACC, this should be considered both a plausible hypothesis (as dopamine mesolimbic and mesocortical pathways are tightly coupled) and an explicit prediction for future experimental testing, adding to the experimental predictions from RVPM on ACC functions in ADHD.

4.2 Comparison with other computational models of ADHD

There are four other major computational models of ADHD that require some discussion. The first is Williams and Dayan's (2005) extended TD model to explain delay aversion in ADHD. They showed that a correct balance between different model parameters is necessary to achieve a normal performance in the temporal discounting task. They concluded that ADHD symptoms can be a common phenotypic expression of several different pathophysiological processes caused by

different genetic deficiencies. One difference between the RVPM and the extended TD model is that RVPM simulated several different symptoms and signs of ADHD, not only delay aversion. Moreover, the RVPM specifies more explicitly the role of a dopaminergic deficit in ADHD pathogenesis. Finally, the extended TD model suffers from the drawback, common to all classical TD models in simulating the temporal dopamine shift from reward period to cue period. In particular, this phenomenon is simulated by a continuous gradual temporal shift, while neurophysiological data showed an event-locked pattern of step-like shifting, analogous to the one we showed in Simulation 1 (Figure 2f-g) (Stuber, et al., 2008). The second model (Cockburn & Holroyd, 2010) is also based on the TD algorithm. The authors showed that an asymmetry between prediction error signals, such as $\delta^+ > \delta^-$, can simulate behavioral results from fixed interval/extinction paradigm (Sagvolden, Hendley, & Knardahl, 1992). Although this model has the merit of highlighting the relevance of prediction error in ADHD, it also simulated only one behavioral aspect of ADHD and did not address its underlying pathophysiology. The third model, by Frank et al. (2005) and O'Reilly et al. (2006), consists of a system simulating the circuitry involving basal ganglia, substantia nigra pars compacta, locus coeruleus and frontal premotor cortex.

In the "Actor-Critic" framework, such a system models more closely an Actor rather than a "Critic", and therefore is to some extent complementary to the RVPM. This model had the merit of providing a theoretical framework for understanding fronto-striatal circuitry involved in working memory (WM) functioning (O'Reilly & Frank, 2006), providing a possible theoretical bridge between the striatal dopamine deficit and WM problems in ADHD. To test this model, ADHD computer simulations were provided by hypothesizing a noradrenergic tonic hyperactivity and focused on RT

variability and impaired performance in probability tracking task (Frank, et al., 2007a; Frank, Scheres, & Sherman, 2007b). Interestingly, dopamine agonists improved ADHD performance related to positive outcomes but not to negative outcomes (Frank, et al., 2007a). Frank's model also provided a computational explanation of this asymmetry, highlighting that dopaminergic drugs have excitatory effect on striatal "Go" cells (D1 dependent) and inhibitory effect on striatal "NoGo" cells (D2 dependent). Although this explanation captures the effect of dopamine agonists on ADHD patients, it does not address, in a straightforward way, the effect of the dopamine deficit in ADHD. Indeed both Go and NoGo related behaviors are impaired in ADHD (Frank, et al., 2007a), while the model would seem to predict an enhanced NoGo activity due to the lack of D2 inhibition. As anticipated above, such results are complementary to the RVPM, which implements a "Critic" rather than an "Actor". Furthermore, the authors focused on only a few aspects of ADHD, thus leaving it, as yet, unspecified how they would account for the range of behavioral and neuroscientific data addressed with the current model.. Finally, the fourth model, by Sikstrom et al. (2007), proposes that a tonically reduced DA level underlies ADHD. As a result, phasic DA is more dependent on optimal experimental parameters, leading to performance that is more impaired at extreme interstimulus intervals (ISIs). However, this model was also focused on just one aspect of ADHD. It will be an interesting avenue for further research to see whether the lesioned RVPM model can also account for the effects of ISI observed in ADHD (Metin, Roeyers, Wiersema, van der Meere, & Sonuga-Barke, 2012).

4.3 Relations with the Dual-pathway model

The dual pathway theory (Sonuga-Barke, 2003) postulated two different dopaminergic deficits each affecting different individuals to different degrees: one involving the mesolimbic system, and the other the mesocortical pathway. The impairment of the first would lead to the RL-related symptoms, while the impairment of the second to the executive symptoms (see Sonuga-Barke & Fairchild, 2012 for a recent formulation). In some sense, then, the current model is a computational version of the mesolimbic pathway and its impairment. One issue is whether the two pathways are independently impaired in ADHD or not. Given the different distributions of D1 and D2 receptors in cortical versus subcortical areas (Hall, et al., 1994), differential impairments of receptor types may lead to partial independence of cognitive and motivational characteristics. This relates to a major computational issue, namely, how the mesocortical pathways should be regulated as a function of attentional requests (Braver & Cohen, 2000). A possible answer could be that the mesolimbic pathway is also involved in executive tasks indirectly, by regulating activity in the mesocortical pathway (e.g. via ACC-to-VTA projections) (Devinsky, Morrell, & Vogt, 1995; Geisler, Derst, Veh, & Zahm, 2007). More generally, mesolimbic impairments could disrupt the release of catecholamines (both dopamine and noradrenaline) from different mesencephalic nuclei. For example, ACC is also bidirectionally connected with locus coeruleus (LC) (Aston-Jones & Cohen, 2005), the source of noradrenergic signals. Noradrenaline is involved in both WM modulation (Aston-Jones & Cohen, 2005) and learning (Yu & Dayan, 2005). Hence, one can hypothesize that ACC impairments could be at least partially responsible for the executive dysfunctions in ADHD via LC dysregulation. In addition, it would

explain why also noradrenergic medication can partially relieve the symptoms of ADHD.

4.4 Conclusion

We demonstrated that a computational model, successful in providing a unified view on the many functions attributed to the ACC, was also able to provide a computational account for many motivation-related features of ADHD. This suggests an important role played by ACC in ADHD. More in general, the current work highlights how RL processing may explain many of the psychological processes at the boundary between cognition and motivation.

References

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience, 9,* 357-381.

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience, 14*, 1338-1344.

Amico, F., Stauber, J., Koutsouleris, N., & Frodl, T. (2011). Anterior cingulate cortex gray matter abnormalities in adults with attention deficit hyperactivity disorder: a voxel-based morphometry study. *Psychiatry Res, 191*, 31-35.

Amiez, C., Joseph, J. P., & Procyk, E. (2005). Anterior cingulate error-related activity is modulated by predicted reward. *The European Journal of Neuroscience, 21,* 3447-3452.

Amiez, C., Joseph, J. P., & Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. *Cerebral Cortex, 16*, 1040-1055.

Aston-Jones, G., & Cohen, J. D. (2005). Adaptive gain and the role of the locus coeruleus-norepinephrine system in optimal performance. *Journal of Comparative Neurology, 493*, 99-110.

Barber, M. A., Milich, R., & Welsch, R. (1996). Effects of reinforcement schedule and task difficulty on the performance of attention deficit hyperactivity disordered and control boys. *Journal of Clinical Child Psychology, 25*, 66-76.

Botvinick, M. M. (2007). Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. *Cogn Affect Behav Neurosci, 7,* 356-366.

Braver, T. S., & Cohen, I. (2000). On the control of control: The role of dopamine in regulating prefrontal function and working memory. In S. Monsell & J. Driver (Eds.), *Attention and Performance* (pp. 713-737). Cambridge, MA: MIT Press.

Brookes, K. J., Xu, X., Anney, R., Franke, B., Zhou, K., Chen, W., Banaschewski, T., Buitelaar, J., Ebstein, R., Eisenberg, J., Gill, M., Miranda, A., Oades, R. D., Roeyers, H., Rothenberger, A., Sergeant, J., Sonuga-Barke, E., Steinhausen, H. C., Taylor, E., Faraone, S. V., & Asherson, P. (2008). Association of ADHD with genetic variants in the 5'-region of the dopamine transporter gene: evidence for allelic heterogeneity. *Am J Med Genet B Neuropsychiatr Genet, 147B*, 1519-1523.

Bush, G., Frazier, J. A., Rauch, S. L., Seidman, L. J., Whalen, P. J., Jenike, M. A., Rosen, B. R., & Biederman, J. (1999). Anterior cingulate cortex dysfunction in attention-deficit/hyperactivity disorder revealed by fMRI and the Counting Stroop. *Biol Psychiatry, 45*, 1542-1552.

Carmona, S., Hoekzema, E., Ramos-Quiroga, J. A., Richarte, V., Canals, C., Bosch, R., Rovira, M., Carlos Soliva, J., Bulbena, A., Tobena, A., Casas, M., & Vilarroya, O. (2011). Response inhibition and reward anticipation in medication-naive adults with attention-deficit/hyperactivity disorder: A within-subject case-control neuroimaging study. *Human Brain Mapping*.

Castellanos, F. X., Margulies, D. S., Kelly, C., Uddin, L. Q., Ghaffari, M., Kirsch, A., Shaw, D., Shehzad, Z., Di Martino, A., Biswal, B., Sonuga-Barke, E. J., Rotrosen, J., Adler, L. A., & Milham, M. P. (2008). Cingulate-precuneus interactions: a new locus of dysfunction in adult attention-deficit/hyperactivity disorder. *Biol Psychiatry, 63*, 332-337.

Cockburn, J., & Holroyd, C. B. (2010). Focus on the positive: computational simulations implicate asymmetrical reward prediction error signals in childhood attention-deficit/hyperactivity disorder. *Brain Research, 1365*, 18-34.

Cubillo, A., Halari, R., Smith, A., Taylor, E., & Rubia, K. (2011). A review of fronto-striatal and fronto-cortical brain abnormalities in children and adults with Attention Deficit Hyperactivity Disorder (ADHD) and new evidence for dysfunction in adults with ADHD during motivation and attention. *Cortex*.

Demurie, E., Roeyers, H., Baeyens, D., & Sonuga-Barke, E. (2012). Temporal discounting of monetary rewards in children and adolescents with ADHD and autism spectrum disorders. *Dev Sci, 15*, 791-800.

Devinsky, O., Morrell, M. J., & Vogt, B. A. (1995). Contributions of anterior cingulate cortex to behaviour. *Brain, 118 (Pt 1)*, 279-306.

Douglas, V. I., & Parry, P. A. (1994). Effects of reward and nonreward on frustration and attention in attention deficit disorder. *J Abnorm Child Psychol, 22*, 281-302.

Durston, S., van Belle, J., & de Zeeuw, P. (2011). Differentiating frontostriatal and fronto-cerebellar circuits in attention-deficit/hyperactivity disorder. *Biol Psychiatry, 69*, 1178-1184.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of crossmodal divided attention on late ERP components. II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78*, 447-455.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience, 17*, 51-72.

Frank, M. J., Santamaria, A., O'Reilly, R. C., & Willcutt, E. (2007a). Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology, 32*, 1583-1599.

Frank, M. J., Scheres, A., & Sherman, S. J. (2007b). Understanding decision-making deficits in neurological conditions: insights from models of natural action selection. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences, 362*, 1641-1654.

Freibergs, V., & Douglas, V. I. (1969). Concept learning in hyperactive and normal children. *J Abnorm Psychol, 74*, 388-395.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A Neural System for Error Detection and Compensation. *Psychological Science, 4*, 385-390

Geisler, S., Derst, C., Veh, R. W., & Zahm, D. S. (2007). Glutamatergic afferents of the ventral tegmental area in the rat. *Journal of Neuroscience, 27*, 5730-5743.

Groen, Y., Wijers, A. A., Mulder, L. J., Waggeveld, B., Minderaa, R. B., & Althaus, M. (2008). Error and feedback processing in children with ADHD and children with Autistic Spectrum Disorder: an EEG event-related potential study. *Clinical Neurophysiology, 119*, 2476-2493.

Hall, H., Sedvall, G., Magnusson, O., Kopp, J., Halldin, C., & Farde, L. (1994). Distribution of D1- and D2-dopamine receptors, and dopamine and its metabolites in the human brain. *Neuropsychopharmacology, 11*, 245-256.

Herrmann, M. J., Mader, K., Schreppel, T., Jacob, C., Heine, M., Boreatti-Hummer, A., Ehlis, A. C., Scheuerpflug, P., Pauli, P., & Fallgatter, A. J. (2010). Neural correlates of performance monitoring in adult patients with attention deficit hyperactivity disorder (ADHD). *World J Biol Psychiatry, 11*, 457-464.

Herrmann, M. J., Rommler, J., Ehlis, A. C., Heidrich, A., & Fallgatter, A. J. (2004). Source localization (LORETA) of the error-related-negativity (ERN/Ne) and positivity (Pe). *Brain Res Cogn Brain Res, 20*, 294-299.

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109*, 679-709.

Hoogman, M., Aarts, E., Zwiers, M., Slaats-Willemse, D., Naber, M., Onnink, M., Cools, R., Kan, C., Buitelaar, J., & Franke, B. (2011). Nitric oxide synthase genotype modulation of impulsivity and ventral striatal activity in adult ADHD patients and healthy comparison subjects. *Am J Psychiatry, 168*, 1099-1106.

Jessup, R. K., Busemeyer, J. R., & Brown, J. W. (2010). Error effects in anterior cingulate cortex reverse when error likelihood is high. *Journal of Neuroscience, 30*, 3467-3472.

Jonkman, L. M., van Melis, J. J., Kemner, C., & Markus, C. R. (2007). Methylphenidate improves deficient error evaluation in children with ADHD: an event-related brain potential study. *Biol Psychol, 76*, 217-229.

Kennerley, S. W., Behrens, T. E., & Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience, 14*, 1581-1589.

Ljungberg, T., Apicella, P., & Schultz, W. (1992). Responses of monkey dopamine neurons during learning of behavioral reactions. *Journal of Neurophysiology, 67*, 145-163.

Luman, M., Oosterlaan, J., & Sergeant, J. A. (2005). The impact of reinforcement contingencies on AD/HD: a review and theoretical appraisal. *Clin Psychol Rev, 25*, 183-213.

Luman, M., Tripp, G., & Scheres, A. (2010). Identifying the neurobiology of altered reinforcement sensitivity in ADHD: a review and research agenda. *Neurosci Biobehav Rev, 34*, 744-754.

Luman, M., van Noesel, S. J., Papanikolau, A., Van Oostenbruggen-Scheffer, J., Veugelers, D., Sergeant, J. A., & Oosterlaan, J. (2009). Inhibition, reinforcement sensitivity and temporal information processing in ADHD and ADHD+ODD: evidence of a separate entity? *J Abnorm Child Psychol, 37*, 1123-1135.

Makris, N., Biederman, J., Valera, E. M., Bush, G., Kaiser, J., Kennedy, D. N., Caviness, V. S., Faraone, S. V., & Seidman, L. J. (2007). Cortical thinning of the attention and executive function networks in adults with attention-deficit/hyperactivity disorder. *Cerebral Cortex, 17*, 1364-1375.

Makris, N., Seidman, L. J., Valera, E. M., Biederman, J., Monuteaux, M. C., Kennedy, D. N., Caviness, V. S., Jr., Bush, G., Crum, K., Brown, A. B., & Faraone, S. V. (2010). Anterior cingulate volumetric alterations in treatment-naive adults with ADHD: a pilot study. *J Atten Disord, 13*, 407-413.

Marco, R., Miranda, A., Schlotz, W., Melia, A., Mulligan, A., Muller, U., Andreou, P., Butler, L., Christiansen, H., Gabriels, I., Medad, S., Albrecht, B., Uebel, H., Asherson, P., Banaschewski, T., Gill, M., Kuntsi, J., Mulas, F., Oades, R.,

Roeyers, H., Steinhausen, H. C., Rothenberger, A., Faraone, S. V., & Sonuga-Barke, E. J. (2009). Delay and reward choice in ADHD: an experimental test of the role of delay aversion. *Neuropsychology, 23,* 367-380.

Martel, M. M., Roberts, B., Gremillion, M., von Eye, A., & Nigg, J. T. (2011). External validation of bifactor model of ADHD: explaining heterogeneity in psychiatric comorbidity, cognitive control, and personality trait profiles within DSM-IV ADHD. *J Abnorm Child Psychol, 39,* 1111-1123.

Matsumoto, M., Matsumoto, K., Abe, H., & Tanaka, K. (2007). Medial prefrontal cell activity signaling prediction errors of action values. *Nature Neuroscience, 10,* 647-656.

McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., & Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science, 323,* 800-802.

Metin, B., Roeyers, H., Wiersema, J. R., van der Meere, J., & Sonuga-Barke, E. (2012). A meta-analytic study of event rate effects on Go/No-Go performance in attention-deficit/hyperactivity disorder. *Biol Psychiatry, 72,* 990-996.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience, 16,* 1936-1947.

Nigg, J. T., & Casey, B. J. (2005). An integrative theory of attention-deficit/ hyperactivity disorder based on the cognitive and affective neurosciences. *Dev Psychopathol, 17,* 785-806.

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: a computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation, 18,* 283-328.

Oades, R. D., & Halliday, G. M. (1987). Ventral tegmental (A10) system: neurobiology. 1. Anatomy and connectivity. *Brain Research, 434*, 117-165.

Oliveira, F. T., McDonald, J. J., & Goodman, D. (2007). Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action-outcome associations. *Journal of Cognitive Neuroscience, 19*, 1994-2004.

Parry, P. A., & Douglas, V. I. (1983). Effects of reinforcement on concept identification in hyperactive children. *Journal of Abnormal Child Psychology, 11*, 327-340.

Posner, M. I., & Dehaene, S. (1994). Attentional networks. *Trends in Neurosciences, 17*, 75-79.

Roger, C., Benar, C. G., Vidal, F., Hasbroucq, T., & Burle, B. (2010). Rostral Cingulate Zone and correct response monitoring: ICA and source localization evidences for the unicity of correct- and error-negativities. *Neuroimage, 51*, 391-403.

Rubia, K., Overmeyer, S., Taylor, E., Brammer, M., Williams, S. C., Simmons, A., & Bullmore, E. T. (1999). Hypofrontality in attention deficit hyperactivity disorder during higher-order motor control: a study with functional MRI. *Am J Psychiatry, 156*, 891-896.

Rushworth, M. F. (2008). Intention, choice, and the medial frontal cortex. *Annals of the New York Academy of Sciences, 1124*, 181-207.

Rushworth, M. F., & Behrens, T. E. (2008). Choice, uncertainty and value in prefrontal and cingulate cortex. *Nature Neuroscience, 11*, 389-397.

Sagvolden, T., Aase, H., Zeiner, P., & Berger, D. (1998). Altered reinforcement mechanisms in attention-deficit/hyperactivity disorder. *Behavioural Brain Research, 94*, 61-71.

Sagvolden, T., Hendley, E. D., & Knardahl, S. (1992). Behavior of hypertensive and hyperactive rat strains: hyperactivity is not unitarily determined. *Physiol Behav, 52*, 49-57.

Sagvolden, T., Johansen, E. B., Aase, H., & Russell, V. A. (2005). A dynamic developmental theory of attention-deficit/hyperactivity disorder (ADHD) predominantly hyperactive/impulsive and combined subtypes. *The Behavioral and Brain Sciences, 28*, 397-419; discussion 419-368.

Scheres, A., Dijkstra, M., Ainslie, E., Balkan, J., Reynolds, B., Sonuga-Barke, E., & Castellanos, F. X. (2006). Temporal and probabilistic discounting of rewards in children and adolescents: effects of age and ADHD symptoms. *Neuropsychologia, 44*, 2092-2103.

Scheres, A., Lee, A., & Sumiya, M. (2008). Temporal reward discounting and ADHD: task and symptom specific effects. *J Neural Transm, 115*, 221-226.

Scheres, A., Milham, M. P., Knutson, B., & Castellanos, F. X. (2007). Ventral striatal hyporesponsiveness during reward anticipation in attention-deficit/hyperactivity disorder. *Biol Psychiatry, 61*, 720-724.

Scheres, A., Tontsch, C., Thoeny, A. L., & Kaczkurkin, A. (2010). Temporal reward discounting in attention-deficit/hyperactivity disorder: the contribution of symptom domains, reward magnitude, and session length. *Biol Psychiatry, 67*, 641-648.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1-27.

Seidman, L. J., Valera, E. M., Makris, N., Monuteaux, M. C., Boriel, D. L., Kelkar, K., Kennedy, D. N., Caviness, V. S., Bush, G., Aleardi, M., Faraone, S. V., & Biederman, J. (2006). Dorsolateral prefrontal and anterior cingulate cortex volumetric abnormalities in adults with attention-deficit/hyperactivity disorder identified by magnetic resonance imaging. *Biol Psychiatry, 60*, 1071-1080.

Sikstrom, S., & Soderlund, G. (2007). Stimulus-dependent dopamine release in attention-deficit/hyperactivity disorder. *Psychological Review, 114*, 1047-1075.

Silvetti, M., Seurinck, R., & Verguts, T. (2011). Value and prediction error in the medial frontal cortex: integrating the single-unit and systems levels of analysis. *Frontiers in Human Neuroscience, 5*.

Silvetti, M., Seurinck, R., & Verguts, T. (2012). Value and prediction error estimation account for volatility effects in ACC: a model-based fMRI study. *Cortex, http://dx.doi.org/10.1016/j.cortex.2012.05.008*.

Sonuga-Barke, E. J. (2002). Psychological heterogeneity in AD/HD--a dual pathway model of behaviour and cognition. *Behavioural Brain Research, 130*, 29-36.

Sonuga-Barke, E. J. (2003). The dual pathway model of AD/HD: an elaboration of neuro-developmental characteristics. *Neurosci Biobehav Rev, 27*, 593-604.

Sonuga-Barke, E. J. (2011). Editorial: ADHD as a reinforcement disorder - moving from general effects to identifying (six) specific models to test. *J Child Psychol Psychiatry, 52*, 917-918.

Sonuga-Barke, E. J., Bitsakou, P., & Thompson, M. (2010). Beyond the dual pathway model: evidence for the dissociation of timing, inhibitory, and delay-related impairments in attention-deficit/hyperactivity disorder. *J Am Acad Child Adolesc Psychiatry, 49*, 345-355.

Sonuga-Barke, E. J., & Fairchild, G. (2012). Neuroeconomics of attention-deficit/hyperactivity disorder: differential influences of medial, dorsal, and ventral prefrontal brain networks on suboptimal decision making? *Biol Psychiatry, 72*, 126-133.

Sonuga-Barke, E. J., Sergeant, J. A., Nigg, J., & Willcutt, E. (2008). Executive dysfunction and delay aversion in attention deficit hyperactivity disorder: nosologic and diagnostic implications. *Child Adolesc Psychiatr Clin N Am, 17*, 367-384, ix.

Strohle, A., Stoy, M., Wrase, J., Schwarzer, S., Schlagenhauf, F., Huss, M., Hein, J., Nedderhut, A., Neumann, B., Gregor, A., Juckel, G., Knutson, B., Lehmkuhl, U., Bauer, M., & Heinz, A. (2008). Reward anticipation and outcomes in adult males with attention-deficit/hyperactivity disorder. *Neuroimage, 39*, 966-972.

Stuber, G. D., Klanker, M., de Ridder, B., Bowers, M. S., Joosten, R. N., Feenstra, M. G., & Bonci, A. (2008). Reward-predictive cues enhance excitatory synaptic strength onto midbrain dopamine neurons. *Science, 321*, 1690-1692.

Swanson, J. M., Kinsbourne, M., Nigg, J., Lanphear, B., Stefanatos, G. A., Volkow, N., Taylor, E., Casey, B. J., Castellanos, F. X., & Wadhwa, P. D. (2007). Etiologic subtypes of attention-deficit/hyperactivity disorder: brain imaging, molecular genetic and environmental factors and the dopamine hypothesis. *Neuropsychol Rev, 17*, 39-59.

Takikawa, Y., Kawagoe, R., & Hikosaka, O. (2004). A possible role of midbrain dopamine neurons in short- and long-term adaptation of saccades to position-reward mapping. *Journal of Neurophysiology, 92*, 2520-2529.

Taylor, E., & Sonuga-Barke, E. J. (2008). Disorders of Attention and Activity. In M. Rutter (Ed.), *Rutter's Child and Adolescent Psychiatry*. Malden, MA, USA: Blackwell Publishing.

Tripp, G., & Wickens, J. R. (2008). Research review: dopamine transfer deficit: a neurobiological theory of altered reinforcement mechanisms in ADHD. *J Child Psychol Psychiatry, 49*, 691-704.

van Meel, C. S., Heslenfeld, D. J., Oosterlaan, J., Luman, M., & Sergeant, J. A. (2011). ERPs associated with monitoring and evaluation of monetary reward and punishment in children with ADHD. *J Child Psychol Psychiatry, 52*, 942-953.

van Meel, C. S., Oosterlaan, J., Heslenfeld, D. J., & Sergeant, J. A. (2005). Telling good from bad news: ADHD differentially affects processing of positive and negative feedback during guessing. *Neuropsychologia, 43*, 1946-1954.

Vijayraghavan, S., Wang, M., Birnbaum, S. G., Williams, G. V., & Arnsten, A. F. (2007). Inverted-U dopamine D1 receptor actions on prefrontal neurons engaged in working memory. *Nature Neuroscience, 10*, 376-384.

Volkow, N. D., Wang, G. J., & Baler, R. D. (2011). Reward, dopamine and the control of food intake: implications for obesity. *Trends in Cognitive Sciences, 15*, 37-46.

Volkow, N. D., Wang, G. J., Kollins, S. H., Wigal, T. L., Newcorn, J. H., Telang, F., Fowler, J. S., Zhu, W., Logan, J., Ma, Y., Pradhan, K., Wong, C., & Swanson, J. M. (2009). Evaluating dopamine reward pathway in ADHD: clinical implications. *Jama, 302*, 1084-1091.

Volkow, N. D., Wang, G. J., Newcorn, J., Telang, F., Solanto, M. V., Fowler, J. S., Logan, J., Ma, Y., Schulz, K., Pradhan, K., Wong, C., & Swanson, J. M.

(2007). Depressed dopamine activity in caudate and preliminary evidence of limbic involvement in adults with attention-deficit/hyperactivity disorder. *Arch Gen Psychiatry, 64*, 932-940.

Wiersema, J. R., van der Meere, J. J., & Roeyers, H. (2009). ERP correlates of error monitoring in adult ADHD. *J Neural Transm, 116*, 371-379.

Williams, J., & Dayan, P. (2005). Dopamine, learning, and impulsivity: a biological account of attention-deficit/hyperactivity disorder. *J Child Adolesc Psychopharmacol, 15*, 160-179; discussion 157-169.

Wise, R. A. (2002). Brain reward circuitry: insights from unsensed incentives. *Neuron, 36*, 229-240.

Wise, R. A. (2004). Dopamine, learning and motivation. *Nature Reviews Neuroscience, 5*, 483-494.

Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron, 46*, 681-692.

Figures captions

Figure 1. Overview of the RVPM structure and of its interactions with the SOFTMAX actor module (light blue arrows) and with the external environment (dark blue arrows). Light blue box: summary of environmental parameter manipulations we operated to generate the three simulations (described in detail in the Appendix Methods). TSN: temporal shifting neuron.

Figure 2. Simulation 1. a) Percentage of choices in ADHD and control groups in the test phase. **b)** Percentage of switches between option 1 and option 2 during the early phase of Simulation 1. Error bars represent standard deviation. **c)** Activity of the V unit (reward expectation) for the most rewarded choice (75% reward rate) in both ADHD and control groups. The plots are cue-onset locked **d)** Activity of the whole ACC (sum of all the three units) during the outcome period in unrewarded trials. Timeline outcome onset locked. **e)** Activity of the whole ACC (sum of all the three units) during the outcome period in rewarded trials. Timeline outcome onset locked. All timelines are in milliseconds. Activity magnitudes are in arbitrary units. **f-g)** Activity of the TSN unit. The arrows under the plots indicate the cue and the outcome onsets respectively. *Blue plots*: activity during the early phase of the simulation. *Green plots*: activity during the middle phase. *Red plots*: activity during the late phase. **f)** Dopamine temporal shifting across trials (early, mid, late in training), from feedback onset to cue onset in Controls. **g)** Impaired dopamine temporal shifting from feedback onset to cue onset in ADHD.

Figure 3. Simulation 2. Percentage of choices for the delayed reward as a function of delay duration. Error bars represent 95% of confidence interval. Controls and ADHD simulated patients do not differ in performance when the reward delay is short (1 s),

Figure 4. Simulation 3. Comparison of choice preferences under two different reward scheduling: **a)** partial (i.e., Simulation 1) and **b)** continuous, in which only choice 2 was rewarded.

Figure 5. Schema summarizing the ACC theory of ADHD pathogenesis. The primary cause is due to the impairment of dopaminergic reward signals to the ACC (red arrow). The dashed blue lines represent the chain of consequent impairments due to dopamine reduction, in both the internal processing steps (light blue) and in the behavioral output (blue). The arrangement of the computational modules corresponds to the one we simulated (see Figure 1).

Appendix

Model structure and dynamics

Figure A1a shows the RVPM structure in detail. The reward expectations encoded in the synaptic connections between the C and the V units are updated online by Hebbian learning modulated through the activity of the δ units. During interaction with the environment, the system continuously compares the reward expectations with the real reward outcomes (the dopaminergic RW signals from the VTA module). When the expectation evoked by an event is higher than the actual outcome there is an activation of the δ^- unit, which codes for negative prediction error ("worse than expected" signals). In contrast, when the expectation is lower than the actual outcome there is an activation of the δ^+ unit, coding for what is called positive prediction error ("better than expected"). Negative and positive prediction errors can be intuitively seen as respectively bad and good surprises with respect to reward expectation.

Figure A1b shows the dynamics of each neural unit of the RVPM as cue-locked activation before and after a training session. Each training trial consisted in the presentation of one cue, followed by a reward with high probability (87%). At the end of the training session, the V unit activity coded for the reward expectation linked to the external cue. Figure A1c shows the activity of the TSN neuron, which receives afferents from the ACC module. This neuron simulates the activity of most of the dopaminergic neurons in the VTA and substantia nigra (SN) (Schultz, 1998), showing an activation locked to cue onset and no more to the reward onset by the end of the training. The generation of dopamine shifting by interaction between ACC and VTA is one of the hypotheses embedded in the RVPM and is based on both anatomical

(recurrent connections between ACC and VTA: Devinsky et al., 1995; Geisler et al., 2007) and functional (ACC neurons showing TD signature like in VTA, Quilodran et al., 2008) data. Further discussion about this topic can be found in Silvetti et al., (2011).

Figure A1 about here

Figure A1. RVPM architecture and dynamics. **a)** Model structure as described in the text. **b)** Single unit activity of ACC neurons, before (blue plots) and after (red plots) conditioning in which one cue was rewarded 87% of times. Black bar over the first row indicates the outcome period **c)** Single unit activity for the TSN neuron of the VTA module. The plot colours indicate three different phases of learning (early, middle, and late training). The plot shows the temporal shifting of the dopaminergic activity from the outcome period (black bar) to the cue period. All the plots in b) and c) are cue-onset locked, timescale in milliseconds, activity scale in arbitrary units. (modified from Silvetti et al., 2011)

Although a complete formal description of the RVPM dynamics was already provided in earlier works (Silvetti et al., 2011, 2012), here follows a short summary of it.

V unit dynamics is described by the following differential equation:

$$\frac{dV}{dt} = -\gamma V + \gamma \left[\sum_i C_i w_i \right]^+ \qquad \text{(A1)}$$

where $[x]^+$ indicates the rectification max(0, *x*) and γ is a scaling parameter. C_i is the activity of *i-esim* cue unit, and w_i is the connection from cue *i* to the V unit.

The weights vector *w* between the *C* units and the V unit is updated by Hebbian learning modulated by the activity of the prediction error units (δ). The learning rule is as follows:

$$\frac{dw_i}{dt} = \alpha C_i V(\delta^+ - \delta^-) \qquad \text{(A2)}$$

where α is the learning rate parameter.

The dynamics of the prediction error units is described by the following:

$$\frac{d\delta^+}{dt} = -\gamma\delta^+ + \gamma[RW - \zeta V]^+ \qquad \text{(A3)}$$

$$\frac{d\delta^-}{dt} = -\gamma\delta^- + \gamma\ T[\zeta V - RW]^+ \qquad \text{(A4)}$$

where ζ is a parameter, *RW* is the RW unit activity and *T* is a bell-shaped timing signal peaking at the average reward onsets (Bueti, Bahrami, Walsh, & Rees, 2010; Ivry, 1996; Mauk & Buonomano, 2004; O'Reilly, Frank, Hazy, & Waltz, 2007). The peak of the *T* signal was dynamically adjusted by a running average (including both long and short trial types) on the delay between choice and reward onsets (Silvetti, et al., 2012). The time resolution of the system was 10 ms, i.e. we arbitrarily assigned the value of 10 ms to each cycle of the network state. Equations (A1), (A3) and (A4) were made stochastic by adding at each network cycle a small amount of noise (white

noise with standard deviation (*SD*) = .5). This made the learning process smoother and less dependent on local fluctuations of reward rates.

Parameter settings were taken from the companion works by Silvetti et al. (2011; 2012).

Actor-Critic framework

To implement the Actor module, we used a SOFTMAX function (Equation A5). This system chose between options as a function of the reward expectations computed by the RVPM:

$$p(C_i) = \frac{e^{Vi\,/\,Temp}}{\sum_i e^{Vi\,/\,Temp}} \qquad \text{(A5)}$$

where $p(C_i)$ is the probability of selecting the i^{th} action, *Temp* is the temperature parameter, and V_i is the V response to the last selection of the i^{th} action. The *Temp* parameter indicates the greediness of the SOFTMAX actor. The lower that the *Temp* value is, the higher is the probability that the system will select the action with the highest reward expectation (exploitation). In contrast, high values of *Temp* lead to more exploration, with frequent selection of the less valued action. The setting we chose for this parameter (see Specific Methods section for Simulation 3) guaranteed exploitative behavior with a small percentage of exploration. Although the modulation of exploration versus exploitation could be interesting for modeling ADHD, its analysis is beyond the aims of this paper.

Simulations

Each trial of each simulation, consisted of a choice between two options. The model was required to learn by trial and error which of the two led to the highest long-term reward. In each trial, one of the two options (encoded by cue units C1 or C2) was chosen by the SOFTMAX, based on their respective values (expected rewards, Equation A5). The chosen C unit generated a square wave of unit amplitude. After a delay period (specific for each simulation) the RW unit generated a reward signal with a certain probability (specific for each simulation). The RW signal consisted of a square wave with a duration equal to 400 ms. Its amplitude was different for control versus ADHD model versions (4 and 2, respectively). The lower amplitude in the ADHD model simulates defective dopamine transmission in the mesolimbic system (Volkow, et al., 2009). The numerical value of dopamine signal for control was set to simulate neurophysiological results in Silvetti et al., (2011) and it was suitable to simulate metabolic data in humans (Silvetti, et al., 2012). The exact value of impaired dopaminergic signal was set arbitrarily to 50% of the normal signal. At the beginning of the simulation, the synaptic connections between the C units and the V unit were randomly set to small values (close to 0.01).

Simulation 1: Specific Methods

We administered a choice task in which the two options were rewarded with different probabilities. The activated C unit generated a square wave of unit amplitude and duration equal to 2000 ms. 1600 ms after C activity onset, the RW unit generated a reward signal with probability 0.75 for C1 and 0.25 for C2. The *Temp* parameter was set to 1. We ran 40 simulation sessions (runs) in total, each representing one subject.

Each run consisted of 60 trials. In half of the runs the RVPM had normal dopamine signals (control group), while in the other half had reduced dopaminergic signals (ADHD group). The measure of the neural activity of the different modules of the RVPM during each trial is the grand average of the last 5 trials (at the end of learning) of all the runs for each experimental group. The statistical comparison of the amplitude of signals in the RVPM was based on their respective signal power within the time bin of interest (cue period and outcome periods). In order to test the reduction of dopamine transfer from primary rewards to predictive cues (TD effect), we computed the 2-way interaction *group* (ADHD, Controls) $\times$ *epoch (cue period, outcome period),* on TSN signal at the end of learning (red plots in Fig. 2f-g).

The behavioral analyses were conducted on the choice percentages. In order to test whether the group factor (ADHD, Controls) influences performance, we computed a t-test between the percentage of choices toward the most rewarding option for both the groups, in the last 30 trials of the task (after the learning phase). The comparison of the percentage of choice switches during the learning phase was computed on the first 30 trials.

Simulation 2: Specific Methods

The system chose between two actions differing in reward delay. In both cases, the reward was certain. We simulated 20 normal and 20 ADHD subjects. For each subject we started by setting the reward delay at 1s for both actions. Every 72 trials we increased the reward delay in steps of 1s for one of the two actions (up to 9s of delay). For each delay step the system learned the new contingencies and we computed the percentage of choices for both actions in the last 42 trials. Besides the reward contingencies, the RVPM learned online also the global average delay (of both long and short delay trials) of reward onset after each choice (Silvetti, et al., 2012), so that

long delays exceeded the expected reward onset, while short delays came earlier than expected. This procedure allowed a discounting curve to be plotted in which each data point represents the preference of the system toward the delayed reward (Figure 3, main text). The dynamics of each single trial was the same as in the previous simulation. All model settings were unchanged from Simulation 1. Statistical comparison between the discounting curves was performed by computing the 95% confidence interval of each data point (Figure 3, main text).

Simulation 3: Specific Methods

We administered a continuous reinforcement schedule, in which one choice was rewarded 100% and the other 0%. Whereas in probabilistic tasks subjects explored the possible choices in order to maximize the reward in an uncertain environment, the optimal behavior in continuous reinforcement schedules consists simply in finding the rewarded action and sticking to that choice. For this reason we modeled both the control and the ADHD groups by reducing the *Temp* parameter of the SOFTMAX from 1 to 0.5, evoking more exploitation for both. The other parameters remained unchanged, like in Simulation 1. To test whether the continuous reward schedule improves ADHD performance relative to control, we computed 2-way interaction *group x reward schedule*.

Appendix References

Bueti D, Bahrami B, Walsh V, and Rees G. (2010) Encoding of temporal probabilities in the human brain. *Journal of Neuroscience*, 30, 4343-4352.

Ivry RB. (1996) The representation of temporal information in perception and motor control. *Current Opinion in Neurobiology*, 6, 851-857.

Mauk MD and Buonomano DV. (2004) The neural basis of temporal processing. *Annual Review of Neuroscience*, 27, 307-340.

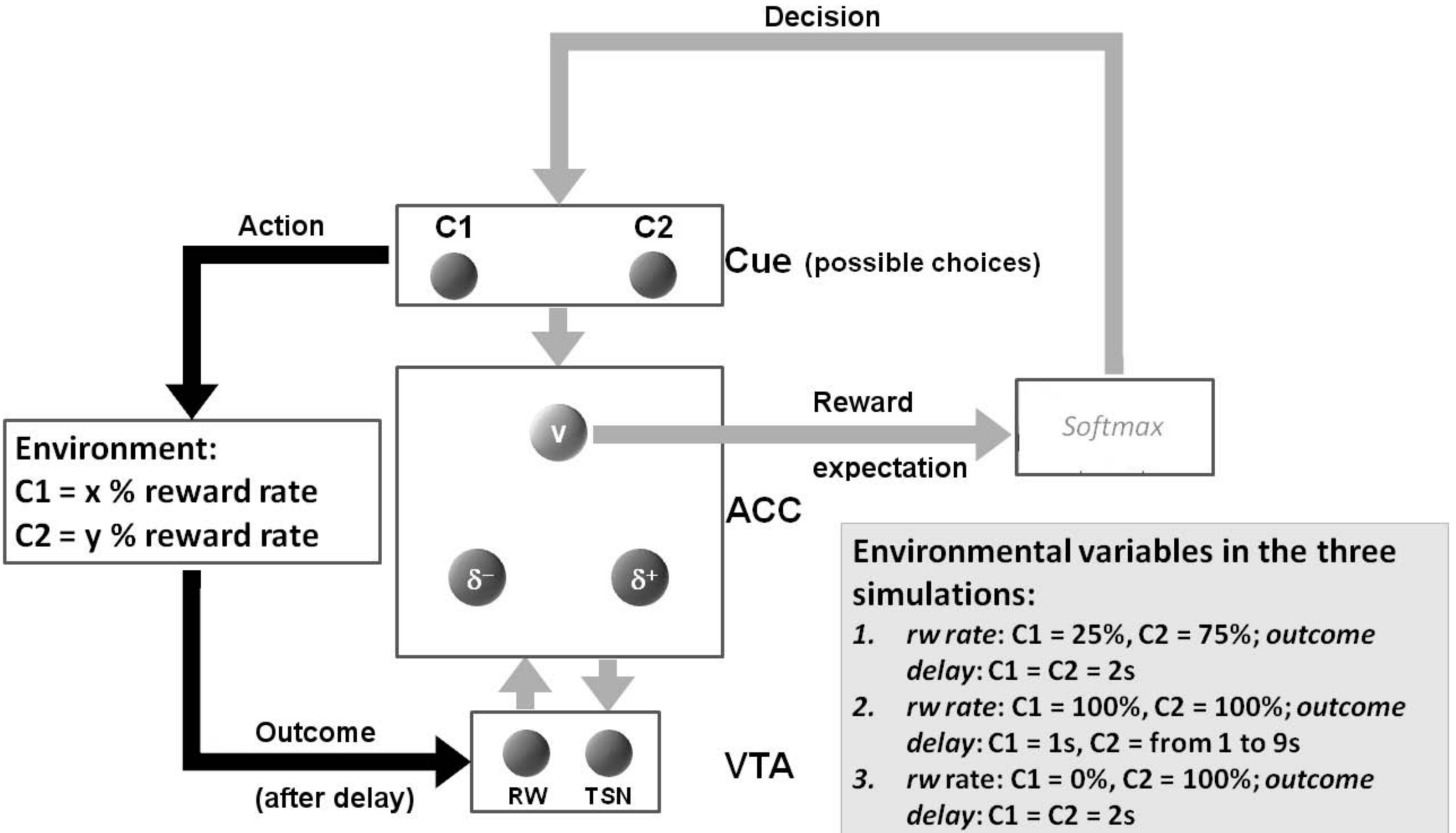
O'Reilly RC, Frank MJ, Hazy TE, and Watz B. (2007) PVLV: the primary value and learned value Pavlovian learning algorithm. *Behavioral Neuroscience*, 121, 31-49.

Schultz W (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, 80 1-27.

Silvetti M, Seurinck R, Verguts T (2011). Value and prediction error in the medial frontal cortex: integrating the single-unit and systems levels of analysis. *Frontiers in Human Neuroscience* **5**(75).

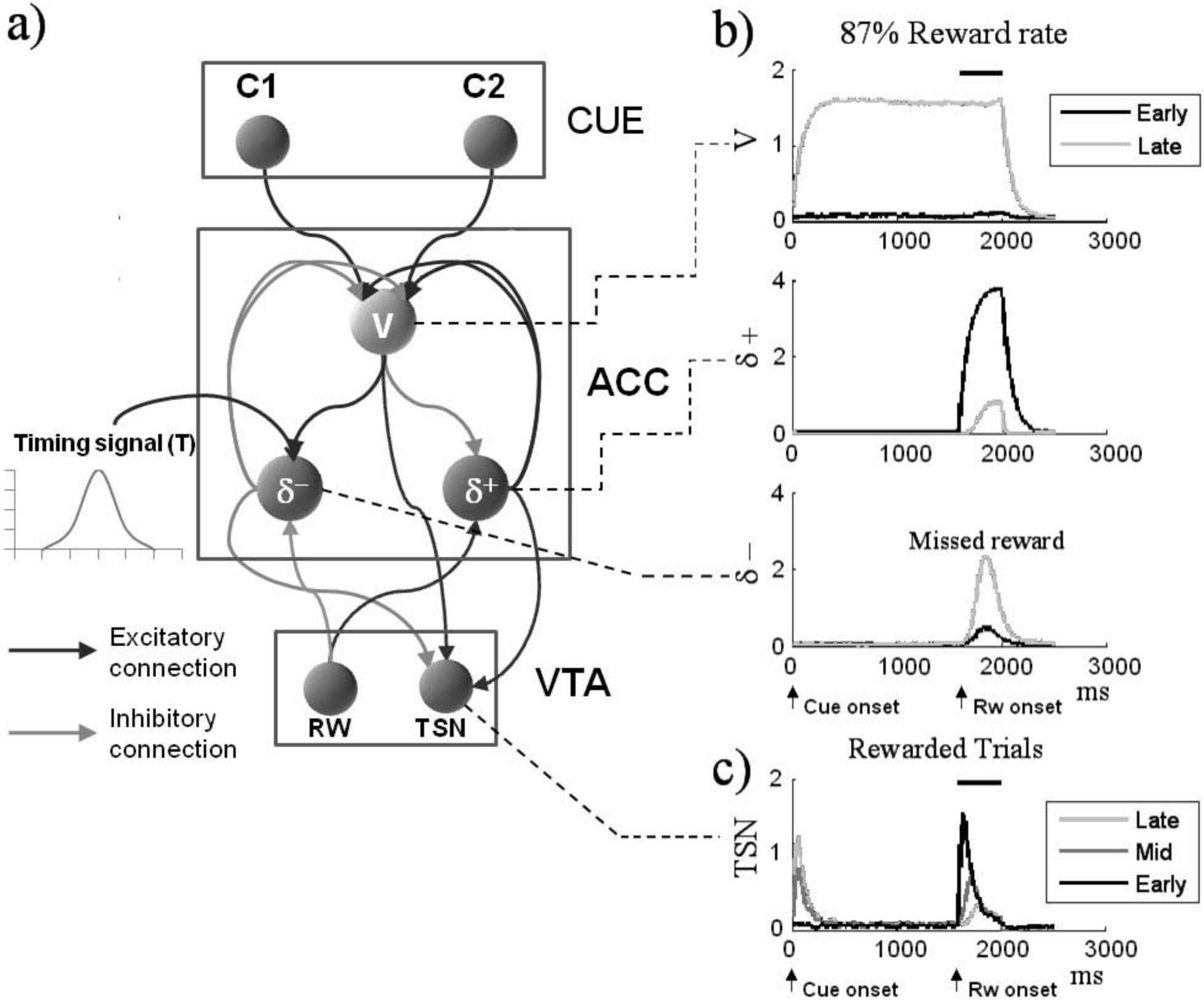
Silvetti, M., Seurinck, R., & Verguts, T. (2012). Value and prediction error estimation account for volatility effects in ACC: a model-based fMRI study. *Cortex, http://dx.doi.org/10.1016/j.cortex.2012.05.008*.

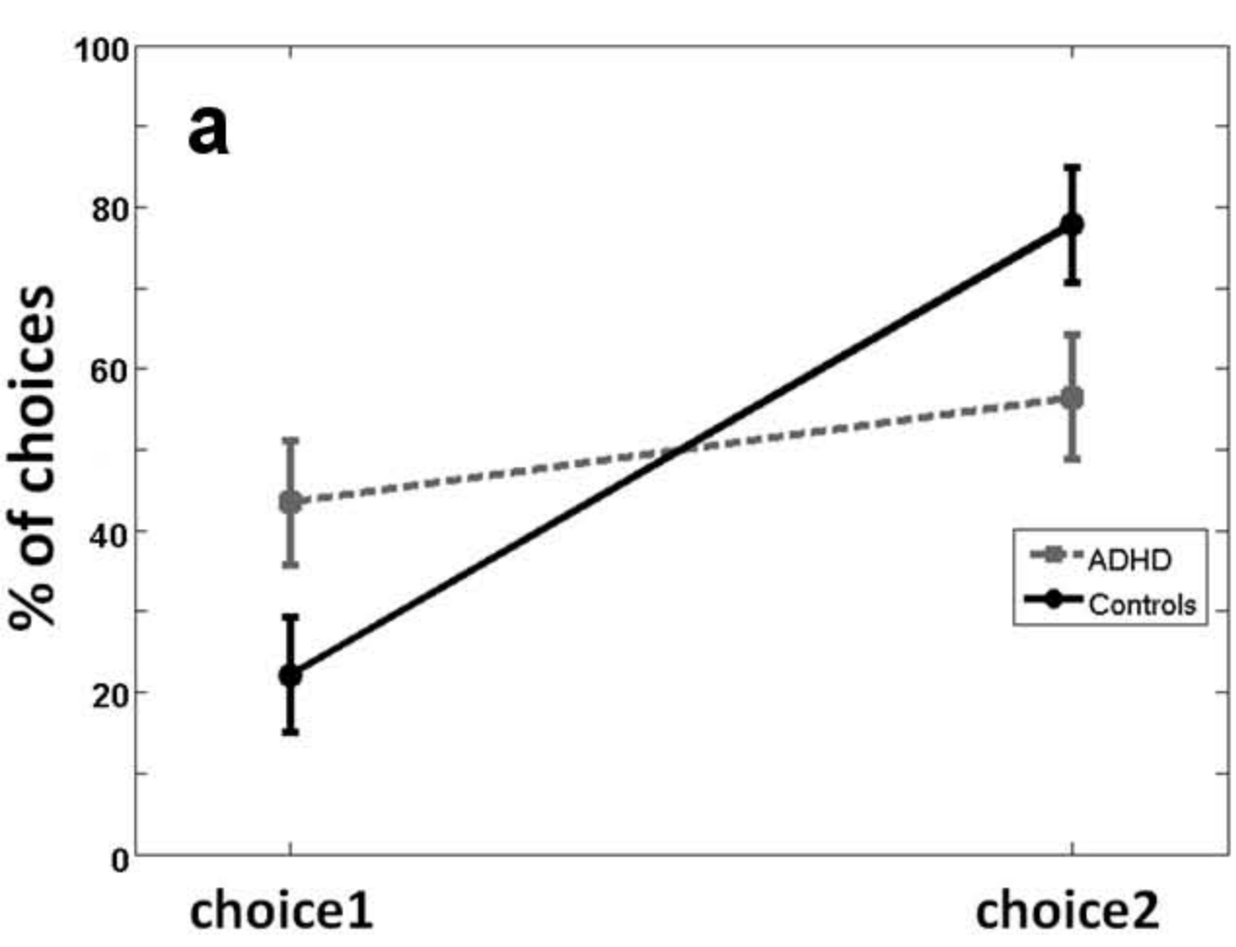
Volkow ND, Wang GJ, Kollins SH, Wigal TL, Newcorn JH, Telang F, *et al* (2009). Evaluating dopamine reward pathway in ADHD: clinical implications. *Jama* 302, 1084-1091.

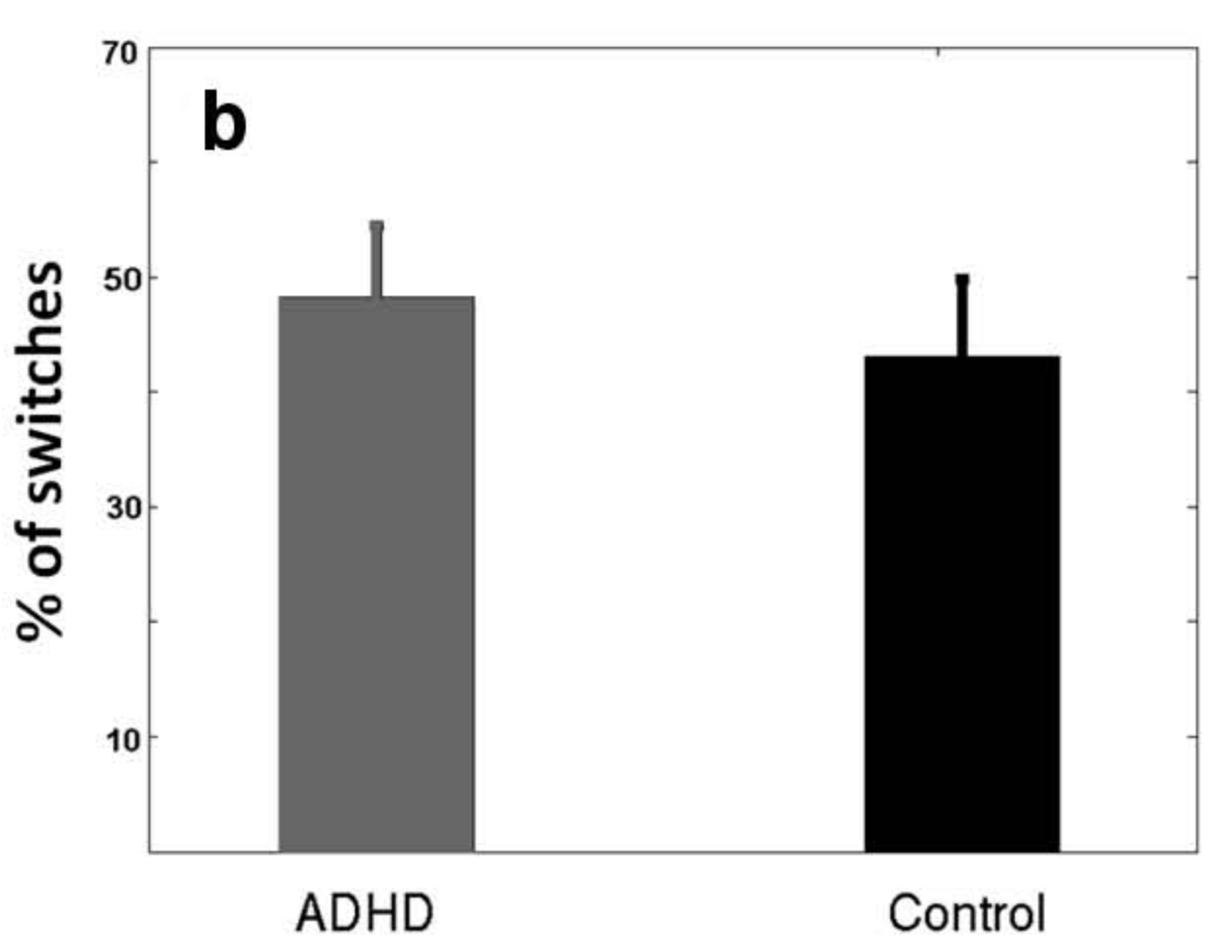


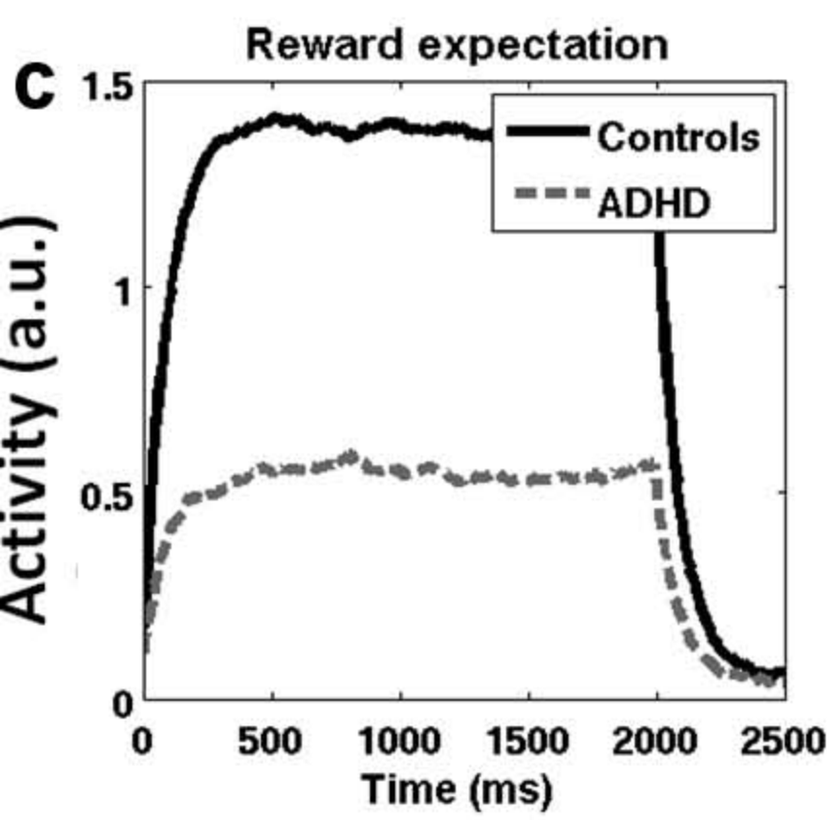
Environmental variables in the three simulations:

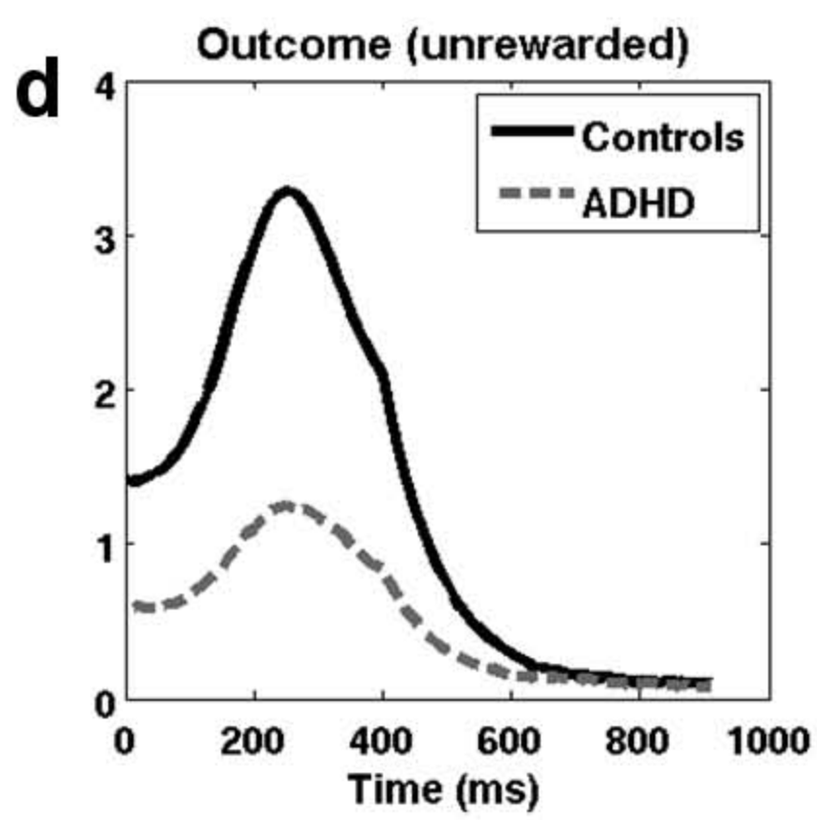
1. *rw rate*: C1 = 25%, C2 = 75%; *outcome delay*: C1 = C2 = 2s
2. *rw rate*: C1 = 100%, C2 = 100%; *outcome delay*: C1 = 1s, C2 = from 1 to 9s
3. *rw* rate: C1 = 0%, C2 = 100%; *outcome delay*: C1 = C2 = 2s

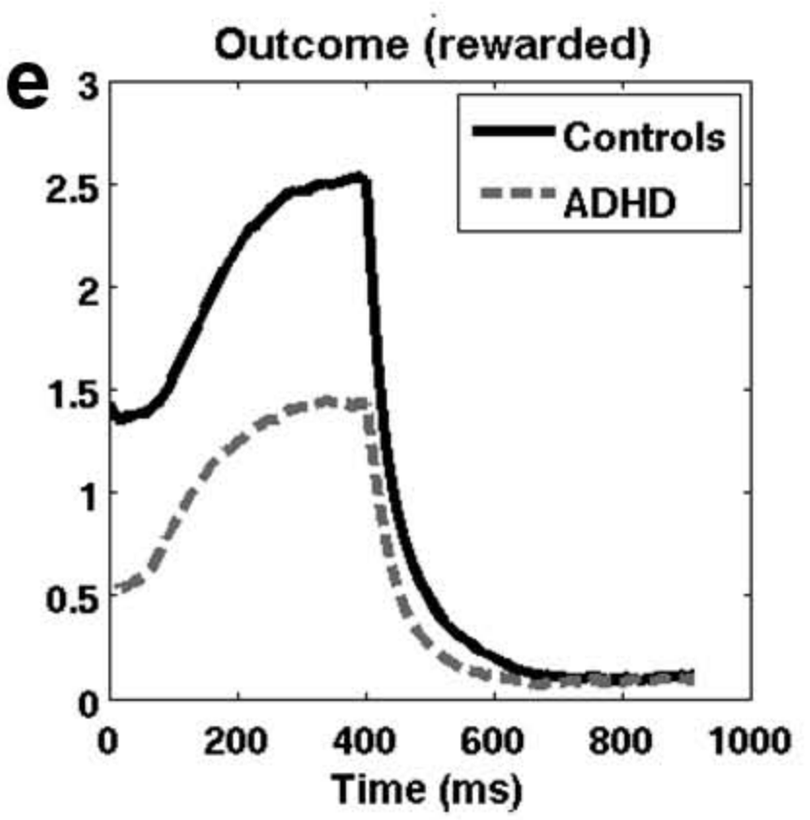

a)
C1
C2
CUE
V
ACC
Timing signal (T)
δ−
δ+
Excitatory connection
Inhibitory connection
RW
TSN
VTA
b)
87% Reward rate
V
Early
Late
δ+
δ−
Missed reward
Cue onset
Rw onset
ms
c)
Rewarded Trials
TSN
Late
Mid
Early
Cue onset
Rw onset
ms

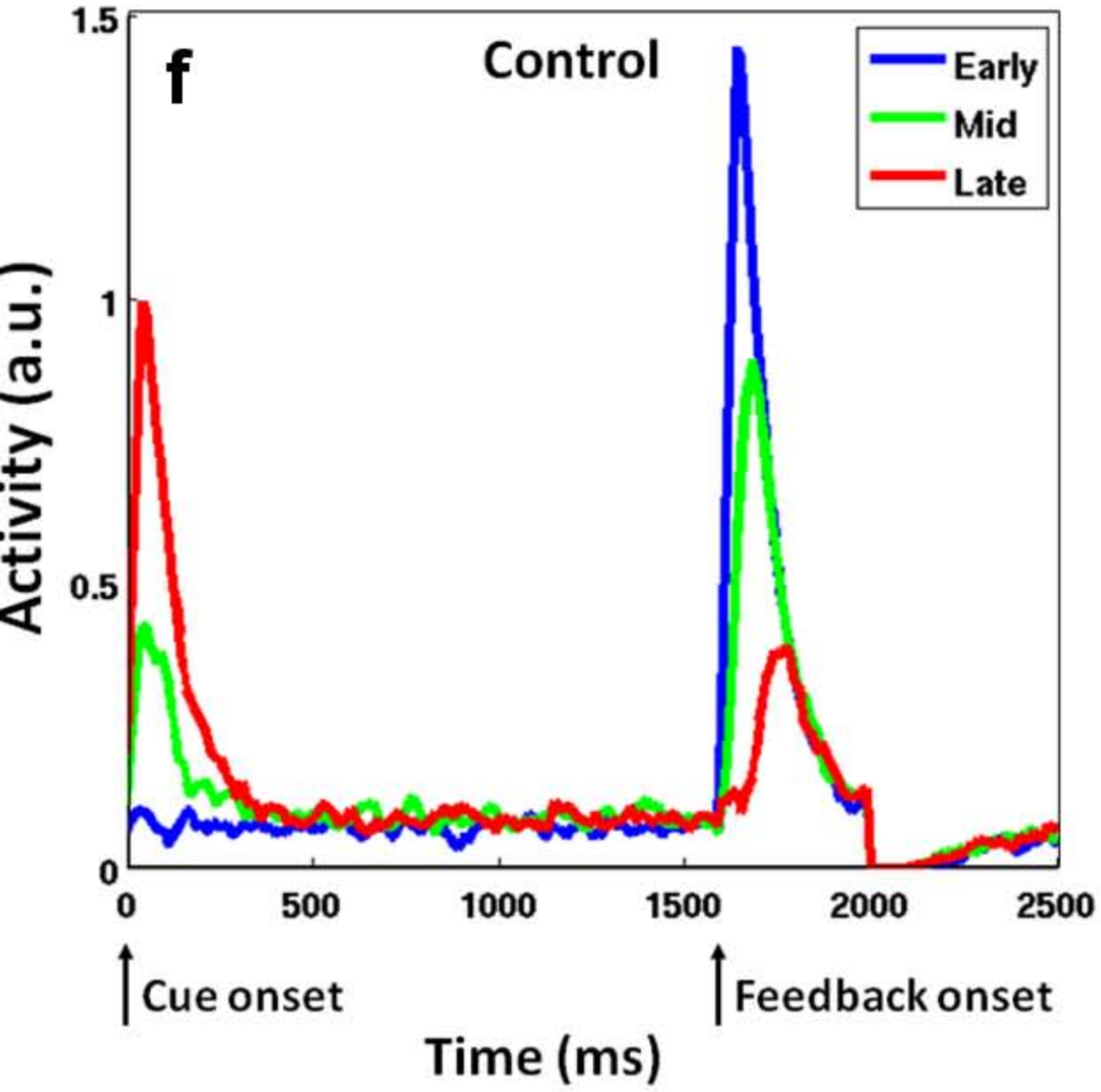

a
% of choices
100
80
60
40
20
0
choice1
choice2
ADHD
Controls

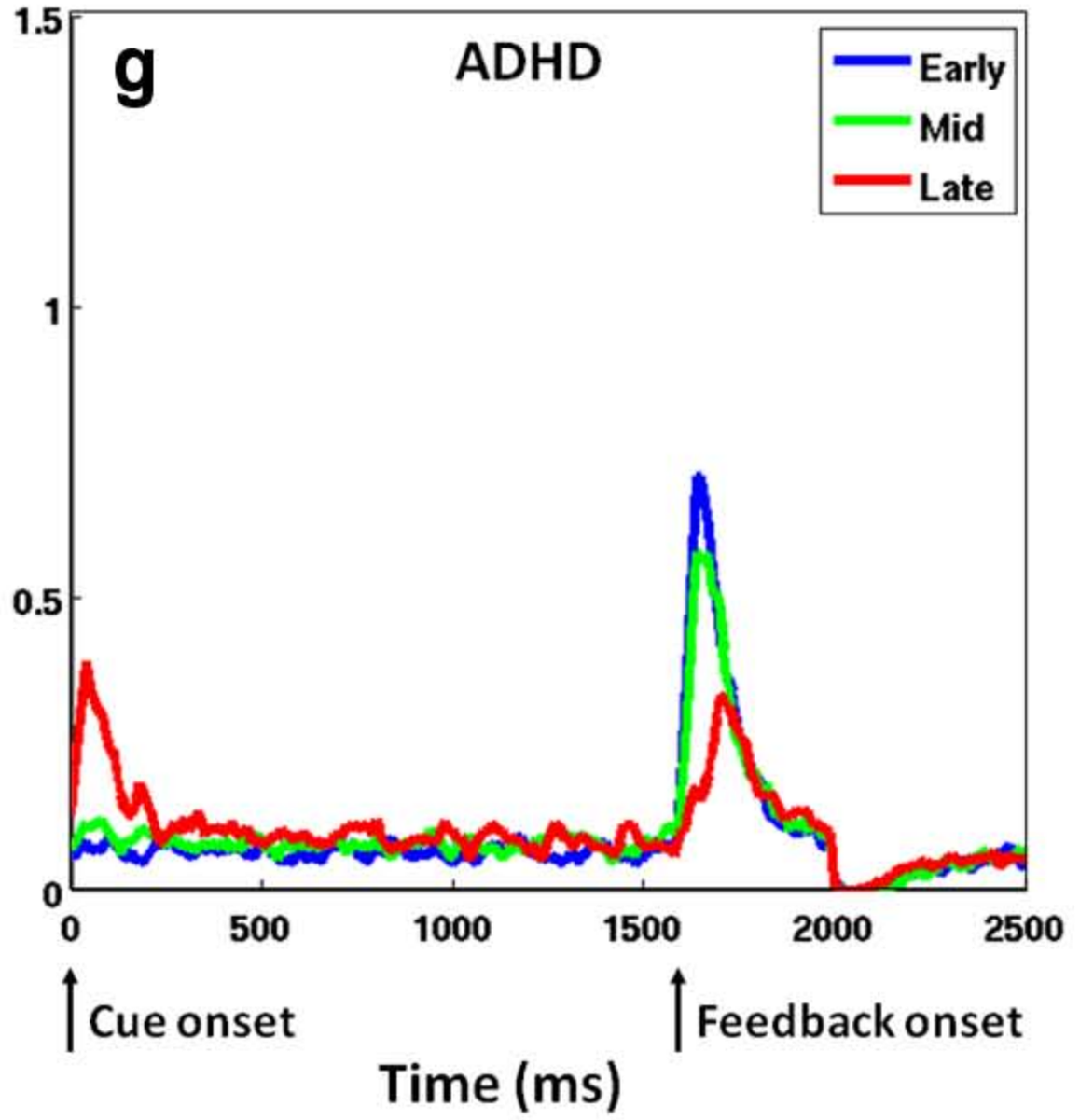

b
% of switches
70
50
30
10
ADHD
Control

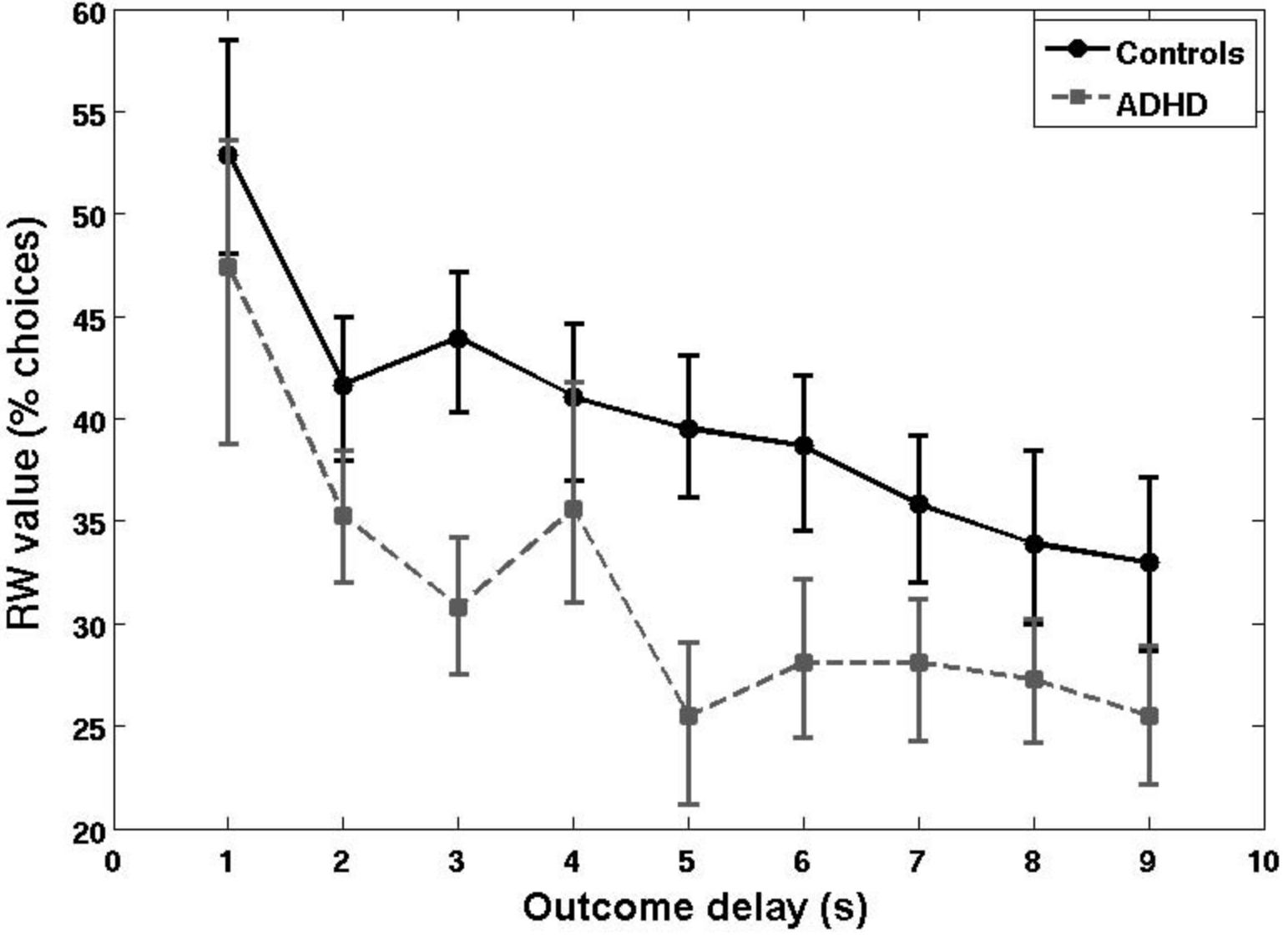

c
Reward expectation
Activity (a.u.)
1.5
1
0.5
0
0
500
1000
1500
2000
2500
Time (ms)
Controls
ADHD

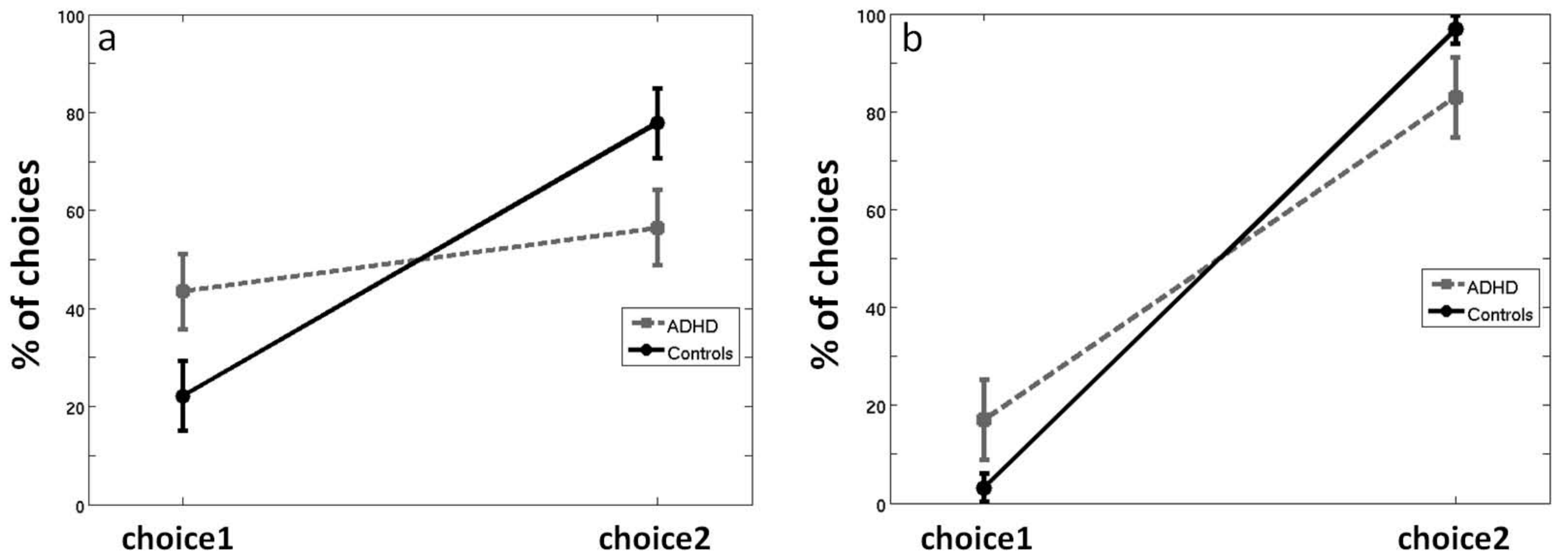

d
Outcome (unrewarded)
4
3
2
1
0
0
200
400
600
800
1000
Time (ms)
Controls
ADHD

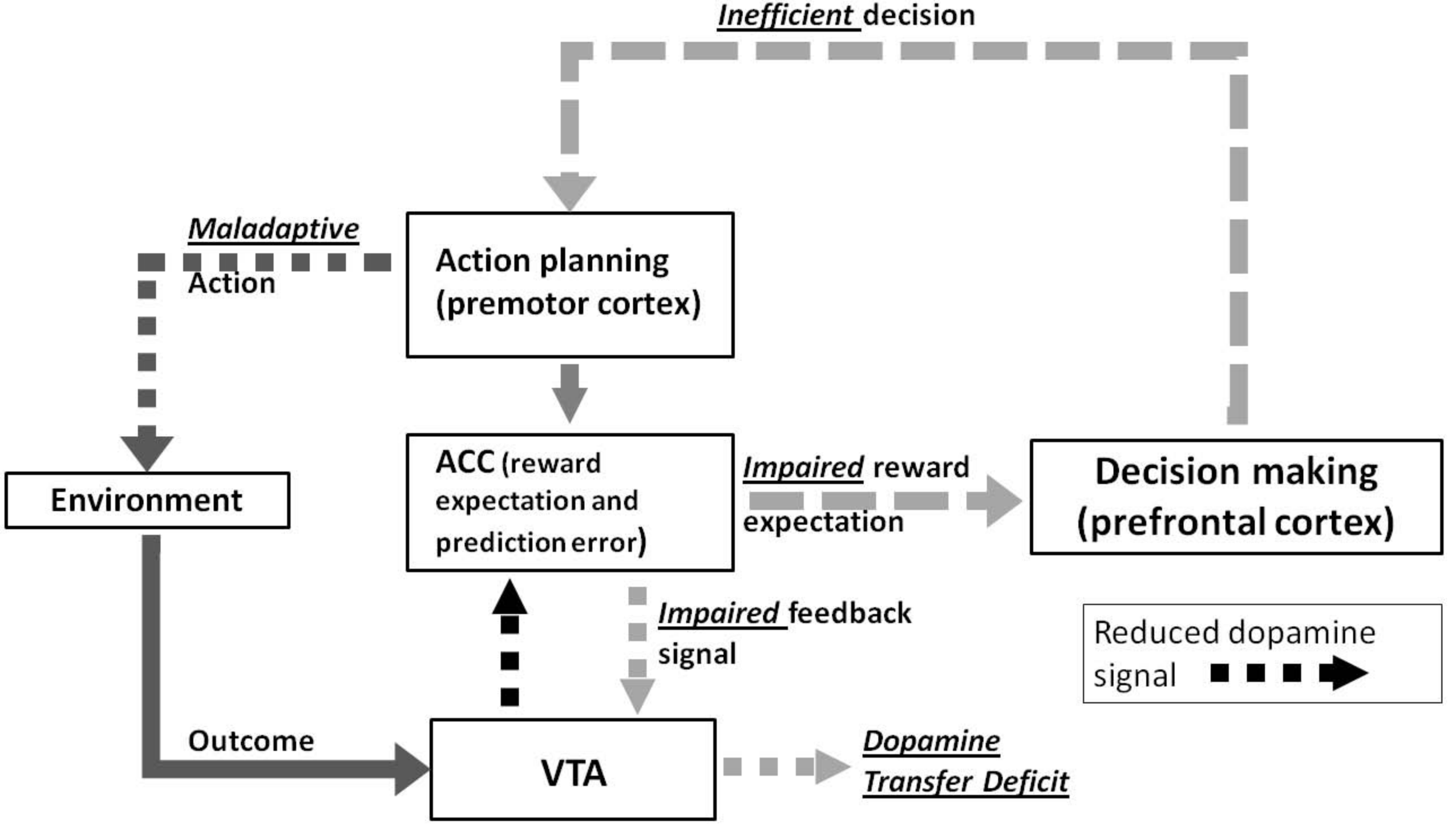

e
Outcome (rewarded)
3
2.5
2
1.5
1
0.5
0
0
200
400
600
800
1000
Time (ms)
Controls
ADHD


f
Control
Activity (a.u.)
1.5
1
0.5
0
0
500
1000
1500
2000
2500
Cue onset
Feedback onset
Time (ms)
Early
Mid
Late


g
ADHD
1.5
1
0.5
0
0
500
1000
1500
2000
2500
Cue onset
Feedback onset
Time (ms)
Early
Mid
Late


Controls
ADHD
RW value (% choices)
Outcome delay (s)
60
55
50
45
40
35
30
25
20
0
1
2
3
4
5
6
7
8
9
10


a
b
% of choices
100
80
60
40
20
0
choice1
choice2
ADHD
Controls


Inefficient decision
Maladaptive
Action
Action planning
(premotor cortex)
Environment
ACC (reward
expectation and
prediction error)
Impaired reward
expectation
Decision making
(prefrontal cortex)
Impaired feedback
signal
Reduced dopamine
signal
Outcome
VTA
Dopamine
Transfer Deficit