Available online at www.sciencedirect.com
ScienceDirect
Neural Networks 20 (2007) 676–689
Neural Networks
www.elsevier.com/locate/neunet

Using ϵ-greedy reinforcement learning methods to further understand ventromedial prefrontal patients' deficits on the Iowa Gambling Task

Kiran Kalidindi*, Howard Bowman

Centre for Cognitive Neuroscience and Cognitive Systems, Computing Laboratory, University of Kent, Canterbury, Kent, CT2 7TR, England, United Kingdom

Received 8 November 2005; received in revised form 6 April 2007; accepted 6 April 2007

Abstract

An important component of decision making is evaluating the expected result of a choice, using past experience. The way past experience is used to predict future rewards and punishments can have profound effects on decision making. The aim of this study is to further understand the possible role played by the ventromedial prefrontal cortex in decision making, using results from the Iowa Gambling Task (IGT). A number of theories in the literature offer potential explanations for the underlying cause of the deficit(s) found in bilateral ventromedial prefrontal lesion (VMF) patients on the IGT. An *error-driven* ϵ-greedy reinforcement learning method was found to produce a good match to both human normative and VMF patient groups from a number of studies. The model supports the theory that the VMF patients are less strategic (more explorative), which could be due to a working memory deficit, and are more reactive than healthy controls. This last aspect seems consistent with a 'myopia' for future consequences.

Keywords: Iowa Gambling Task; Ventromedial prefrontal cortex; Orbitofrontal cortex; Reinforcement learning; Decision making

1. Introduction

In the decision making literature of the past decade, a popular paradigm has been the Iowa Gambling Task (IGT) (Bechara, Tranel, & Damasio, 2000). The IGT was originally designed to elucidate some of the particular deficits found in patients with bilateral ventromedial prefrontal cortex lesions (VMF). The IGT is a reinforcement learning problem, in that participants must learn from rewards and punishments to evaluate the most appropriate action. Our aim is to find valuation functions, which describe the average behaviour found in VMF patients and normative human groups on the IGT, using models based on ϵ-greedy methods (Sutton & Barto, 1998). The ϵ-greedy framework was used because of the simplicity and flexibility it offered in testing a variety of theories. Our work has a related motivation to other modelling work on the IGT (Busemeyer & Stout, 2002; Yechiam, Busemeyer, Stout, & Bechara, 2005). However, our work differs in two important respects, (1) it attempts to clarify and extend current theories; and (2) it tests the models against data from four versions of the IGT across a number of studies, rather than the frequently used ABCD version alone. Additionally, other modelling work has not simulated the time-course of selections across the task (Wagar & Thagard, 2004), or has modelled a reduced choice variant of the IGT (Frank & Claus, 2006) that has not been tested on VMF patients.

The IGT attempts to mimic real world decision making, where the outcome of choices and strategies have an element of immediate and, particularly, long-term uncertain consequence. All four versions of the task we consider contain four decks of cards, two of which are advantageous (decks C and D in the original ABCD version) and two of which are disadvantageous (decks A and B in the original ABCD version). Through selection, players need to learn which decks are best. Initially, the bad decks seem the best, as they offer higher immediate reward. However, they also offer higher uncertain losses, which only becomes evident after a number of selections. Importantly though, as the task progresses, normal healthy humans learn that the best decks are those that offer smaller immediate rewards, but also lower uncertain punishments, whereas VMF patients seem unable to fully use this distinction. Overall, the IGT tests a number of aspects of decision making including,

* Corresponding author. Tel.: +44 0 1227 827553; fax: +44 0 1227 762811.
E-mail addresses: kk49@kent.ac.uk (K. Kalidindi), H.Bowman@kent.ac.uk (H. Bowman).

0893-6080/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neunet.2007.04.026

within task learning, management of reversals in contingencies and evaluation of regular rewards and punishments over uncertain ones. It should be noted that in all versions of the IGT, the disadvantageous decks provide the best regular returns (outcomes that occur on every selection of a particular deck).

The IGT paradigm has been used as a method for distinguishing decision making deficits in bilateral VMF patients compared to normal healthy controls (Bechara, Damasio, Damasio, & Anderson, 1994; Bechara, Damasio, Damasio, & Lee, 1999; Bechara, Damasio, Tranel, & Damasio, 1997; Bechara et al., 2000; Fellows & Farah, 2005), and with various other frontal lesion patient groups (Bechara, Damasio, Tranel, & Anderson, 1998; Clark, Manes, Antoun, Sahakian, & Robbins, 2003; Fellows & Farah, 2005; Manes et al., 2002), including patients with unilateral VMF lesions (Tranel, Bechara, & Denburg, 2002). For a review of a number of other studies, with a variety of subject groups see Dunn, Dalgleish, and Lawrence (2006).

This paper continues by summarizing four different versions of the IGT, and goes on to consider five theories in the literature, which each attempt to define the underlying cause of deficits found in VMF patients' performance on the IGT. With the aid of simulations of human normal healthy controls' (NHCs)' and VMF patients' IGT profiles, these theories are considered in greater depth, and conclusions are drawn about the most suitable theory. This has allowed the authors to suggest that VMF patients are less strategic (more explorative), which could be due to a working memory deficit, and are more reactive (more influenced by recent results) than healthy controls.

2. Versions of the Iowa Gambling Task (IGT)

We consider four versions of the IGT, ABCD, A′B′C′D′, EFGH and E′F′G′H′. (For further details of the task see Bechara et al. (2000).) It is important to note that the bad decks, A(′) and B(′), in the A(′)B(′)C(′)D(′) versions have the largest variance in potential wins and losses per card, making them more 'risky'. Whereas, in the EFGH and E′F′G′H′ versions the good decks, E(′) and G(′), have the highest variance.

In the A(′)B(′)C(′)D(′) versions, the good decks, C(′) and D(′), provide regular wins of \$50 and average losses of \$25 per selection, giving a mean return of \$25. In deck C(′) the losses are smaller and more frequent than in deck D(′). For bad decks, A(′) and B(′), regular wins are equal to \$100 and uncertain losses average \$125 per selection, giving an average loss of \$25 per selection. In deck B(′) there are occasional large losses, whereas in deck A(′) there are more frequent smaller losses. In the A′B′C′D′ and E′F′G′H′ variants, the good decks become better and the bad decks become worse over the course of the task. More precisely, in deck A′, the frequency of punishment is increased by 10% every 10 cards. Whereas in deck B′, the magnitude rather than the frequency of each uncertain punishment is increased by 10% every 10 cards. Parallel decreases in the punishments were applied to decks, C′ and D′, C′ had a 10% decrease in the frequency of punishment after every 10 cards, and in deck D′ there was a matching decrease in the magnitude of loss.

A fairly similar adjustment is made to the score card for EFGH to produce the E′F′G′H′ variant of the task. Here, in deck F′, there is a 6% decrease in the frequency of delayed/uncertain reward after every 10 selected cards from that deck. With a corresponding decrease, in only magnitude, rather than frequency, in uncertain reward for deck H′. In deck E′, there is a 6% increase in the magnitude of reward for each win and a matching increase in total wins in deck G′, but in G′ it is generated by increasing the frequency of wins.

The measure used to compare participants' choices throughout the task are their net scores, which is calculated by adding up the number of advantageous choices, selections from decks C and D for the ABCD version, and subtracting the number of disadvantageous choices, from the other two decks, A and B (i.e. net score = (C+D)−(A+B) or (E+G)−(F+H) for the EFGH version). Participants' net scores are often broken down into 5 blocks of 20 selections, to show how performance evolves over the course of the task (Bechara et al., 2000) (i.e. block 1, selections 1–20; etc.). This is the format that the human and simulation data will be presented in in the results and analysis segment (see Section 5) of this paper.

3. Competing theories of VMF patient deficits

The current work considers five theories present in the literature, that each offer possible underlying causes for the decision making deficits found in bilateral VMF lesion patients tested on the Iowa Gambling Task (IGT). We review these five hypotheses here.

3.1. 'No preferences' — A. Sloman

In Sloman (2004), Aaron Sloman suggests that Bechara and Damasio et al.'s VMF patients have lost their preferences and emotions. Sloman points out that it is not necessarily true that because VMF patients have reduced emotions and decision making deficits that emotions are causally required for 'rational' decision making. He suggests that the VMF patients have lost their preferences and therefore, have both reduced emotions and decision making deficits. This view of 'no preferences' is supported by anecdotal evidence (Sacks, 1998), where a patient with large bilateral orbitofrontal cortex lesions professed to no longer having preferences.

One would expect 'no preferences' to lead to random selection on the IGT, and may reflect an inability to retain information on the results of past selections, potentially a working memory deficit. The questions remains open whether an intact working-memory is required for this type of decision-making (Bechara et al., 1998; Hinson, Jameson, & Whitney, 2002; Maia & McClelland, 2004). In general, working memory, particularly the retention of information over short delays, is considered to require dorsolateral prefontal (DLF) regions (Goldman-Rakic & Leung, 2002), rather than orbital frontal regions (Stone, Baron-Cohen, & Knight, 1998). Bechara et al. (1998) found that patients with right dorsolateral/medial prefrontal lesions had working memory deficits, but not decision making deficits on the IGT. However, Fellows and

Farah (2005) and Manes et al. (2002) have found that other DLF patients were poor at the IGT.

3.2. 'Risk-seeking' — Sanfey et al.

In the original ABCD version of the IGT, the riskiest decks (i.e. with highest variance) were also the disadvantageous decks. Therefore, Sanfey, Hastie, Colvin, and Grafman (2003) have suggested that VMF patient behaviour might be caused by them being overtly attracted to risk. Sanfey et al. (2003) found, in their own gambling task, that they had risk-VMF and safe-VMF sub-groups. The two VMF sub-groups had different behaviour profiles, but there were not obvious lesion location or volume differences in their VMF damage. In Sanfey et al. (2003)'s gambling task, there was no pre-set reversal, but on-line learning was required in the task, each trial was independent, and the main aim of the task was to test attitudes towards risk.

In another study, Rogers et al. (1999) introduced the Cambridge Gamble Task (which requires no online learning, uses single-shot probabilistic judgements and tests impulsivity), where they found that orbitofrontal cortex (OFC) lesion patients did not, as often as normals and dorsolateral prefrontal cortex lesion (DLF) patients, make the probabilistically most rational gamble. In addition, OFC patients bet at a reduced rate compared to normal controls and DLF patients. This suggests that patients with OFC lesions were risk-seeking in terms of not picking the most rational gamble, but were more cautious than normals when it came to the amount they risked and so were not impulsive. Therefore, these results are somewhat inconclusive. Although in another study, patients with aneurysms of the anterior communicating artery (who often exhibit similar emotional and judgemental deficits to VMF patients) were found to take greater risks than NHCs on the Cambridge Gamble Task (Mavaddat, Kirkpatrick, Rogers, & Sahakin, 2000).

3.3. Frequency vs. magnitude — Frank and Claus

Frank and Claus (2006) have suggested that the OFC is important in the representation of magnitude and that the basal ganglia have a greater role in representing the frequency of rewards and punishments resulting from a choice/stimulus (with perhaps a limited representation of magnitude). Frank and Claus (2006) have suggested that complete OFC lesions would result in a participant relying on the basal ganglia and therefore, mainly basing their decisions on the frequency of losses and wins. Bechara et al.'s VMF patients' IGT net scores simulated in our present study are from patients with substantial, but not complete OFC lesions. Frank and Claus (2006) use their model to predict the behaviour of patients with complete OFC lesions on a modified version of the IGT (Peters & Slovic, 2000), which has only been tested on normal healthy controls (NHCs).

3.4. 'Myopia' for future consequences — Bechara et al.

Bechara et al. (2000) have described the decision making behaviour of their VMF patients as displaying a 'myopia' for future consequences (insensitivity to future consequences). This theory was chosen over two other theories: (i) hypersensitivity to reward, and (ii) insensitivity to punishment, in Bechara et al. (2000). This choice was made for two reasons. Firstly, both theories (i) and (ii) would have resulted in disadvantageous behaviour in the A(′)B(′)C(′)D(′) versions of the task and advantageous behaviour in the E(′)F(′)G(′)H(′) versions. This was not found since VMF patients acted disadvantageously in all versions. Secondly, they found that when they increased the relative delayed punishment in the bad decks (A′B′C′D′ version) or increased the relative delayed reward in the good decks (E′F′G′H′ version) the VMF patients performance did not improve. This provides support for the insensitivity to future consequences ('myopia') theory. Anecdotally, the group have, additionally, found that their VMF patients act in the short term and often make decisions which are detrimental in the long run to their social and professional lives (Damasio, 1994). In addition, this group have suggested that an explanation for this 'myopia' is a lack of 'somatic-markers', which are bodily representations of past experience.

3.5. Reversal learning — Rolls, Clark et al. and Fellows and Farah

In both monkeys and humans, frontal lobe lesions have been found to cause participants to perseverate in reversal learning tasks (Clark, Cools, & Robbins, 2004; Fellows & Farah, 2003; Hornak et al., 2004; Rolls, 2000, 2004; Rolls, Hornak, Wade, & McGrath, 1994). Although the patients in the Rolls et al. (1994) study had diffuse ventral prefrontal cortex lesions, Fellows and Farah (2003, 2005)'s patients had more localised damage, with a focus in the ventromedial prefrontal cortex, confirmed by MRI. Therefore, the studies by Fellows and Farah allow for a better comparison, in terms of lesion location, to the VMF patients described in Bechara et al.'s studies. Fellows and Farah (2005) found that when they removed the reversals in the IGT by re-arranging the score-card, developing the 'shuffled' ABCD variant, their VMF patients acted similarly to their NHCs. However, it should be noted that Bechara et al.'s VMF patients were originally selected on the basis of their behavioural deficits as well as their lesion locations, whereas Fellows and Farah's VMF patients were selected on lesion location alone.

4. ϵ-greedy action-value method

In this paper, we have tried to simulate the human NHCs (Normal Healthy Controls) and human VMF patients' data from Bechara et al. (1999, 2000), Bechara and Damasio (2002), Bechara, Dolan, and Damasio (2002) and Clark et al. (2003), using basic reinforcement learning algorithms, based on ϵ-greedy action-value methods (see p. 27, Sutton and Barto (1998)). The ϵ in the ϵ-greedy method signifies the probability of exploration on each trial, where ϵ can take values from 0 to 1. If $\epsilon = 0$, then the algorithm is purely exploitative, and will always select the choice with the highest estimated value (the greedy action). If $\epsilon = 1$, then every trial is explorative, where, in this set-up, each action is equiprobable.

On exploitation trials, the greedy action is taken. This selection depends on the valuation function used, which dictates how reward and punishment information is evaluated. A number of valuation functions have been used to try to recreate the human data, test the five theories under consideration and to compare our models to others in the literature. The choice of valuation function depends on the version of the ϵ-greedy model being used, and only affects the greedy action. However, in all versions of the model, to decide whether to select a greedy or explorative action, Eq. (1) is used on each trial.

$$\begin{aligned}&\text{if } (X(t) > \epsilon), \{\text{Exploitation}\}\\ &\quad \forall i (1 \le i \le N).\\ &\qquad \text{if } (i \in Z), \quad \Pr[a_i] = \frac{1}{|Z|}\\ &\qquad \text{otherwise}, \quad \Pr[a_i] = 0\\ &\text{otherwise}, \{\text{Exploration}\} \quad \forall i (1 \le i \le N). \quad \Pr[a_i] = \frac{1}{N}\end{aligned} \tag{1}$$

where

$$Z = \{j \mid 1 \le j \le N \wedge \forall k (1 \le k \le N), ((k \ne j) \Rightarrow Q_t(a_k) \le Q_t(a_j))\}$$

$X(t)$ is the value of a random variable at time t, that ranges from 0 to 1. The probability of an action a_i during a trial is $\Pr[a_i]$. N denotes the number of available actions. $Q_t(a_k)$ denotes the estimated value of taking action a_k at time t. Z is the set of maximal actions, with $|Z|$ denoting the set's cardinality (i.e. the number of elements). Therefore, if a greedy action is to be taken, the action with the maximum estimated return is selected. If a number of actions have the joint highest estimated value, then one of those actions is selected at random with equal probability (i.e. a member of set Z is selected with probability $\frac{1}{|Z|}$). The estimated reward or punishment is dictated by the reward schedule experienced so far and the valuation function used. In an exploration trial, the probability of selecting a particular choice, $\Pr[a_i]$, is only dependent on the number of available choices, N. Eq. (1) represents the general case; therefore, in the case of the IGT (which has 4 choices/decks), $\Pr[a_i] = 0.25$ during an exploration trial.

In the simulations considered here, one of a number of simple valuation functions are used. If we use the original ABCD variant of the IGT as an example, then each action would be a selection of a card from one of the four decks, A, B, C or D. In addition, in all nine models, the estimations at the beginning of the task are set to zero for each possible action as the participants have no prior knowledge of the task (see IGT participant instructions in the methods section of Bechara et al. (1999)). Unless otherwise specified, in all the following valuation functions, if an action is not selected, then $Q_t(a) = Q_{t-1}(a)$.

Sample-averaging. This method simply uses the mean of the rewards and punishments received, so far. Therefore, through sufficient exploration, and the law of large numbers, this method would become asymptotic to the optimal policy, minus the cost of exploration.

If, on the Tth play, action a has been chosen k_a times prior to T, yielding returns $r_t(a)$, then its value is estimated using Eq. (2) (with $r_t(a) = 0$ when a is not selected on the tth trial). However, the policy can never be quite optimal if a constant exploration is used. For example, if an exploration rate of $\epsilon = 0.2$ is used, then on 0.2 of the trials a potentially non-optimal action may be selected. Therefore, on the IGT, where two of the four (50%) choices are advantageous, on 0.1 of the trials, this parameter setting would select disadvantageously, even if the correct value of each action was known.

$$Q_T(a) = \frac{1}{k_a} \sum_{t=1}^{T} r_t(a). \tag{2}$$

The *sample-averaging* version of the model has one free parameter, ϵ, which ranges from 0 to 1. The full parameter space was searched for the best matches to the human data sets, in 0.01 steps. This model provides a good test of the 'no preferences' theory, i.e. an exploration rate close to 1 suggests a tendency towards 'no preferences'. In addition, a mean of 1000 simulations are used to represent the average result, for each parameter setting. This is the same for all the models. An average across many simulations is used since there is inherent variance in each run of a model due to the randomness of exploration, even for the same parameter settings and valuation function.

Variance-driven learning. This valuation function calculates the variance in the results from each deck experienced by the simulation. A standard equation for the variance is used (see Eq. (3)).

$$Q_T(a) = \frac{1}{k_a} \sum_{t=1}^{T} (r_t(a) - \bar{r}_T(a))^2. \tag{3}$$

The number of times an action has been taken is again k_a, while $\bar{r}_T(a)$ is the mean return for a particular action a at time T. This model was theoretically inspired by the 'risk-seeking' theory. The use of variance by VMF patients is considered an alternative to the strategies used by NHCs. Again, this model only has one free parameter, the exploration rate ϵ, which was examined using 0.01 steps across the 0 to 1 range of the parameter.

Frequency-driven learning. This model is theoretically inspired by the recent work of Frank and Claus (2006). Frank and Claus suggest that following OFC lesions, since decision making becomes more dependent on the basal ganglia, assessment of choices becomes more *frequency driven*, with respect to the frequency of positive and negative outcomes from a particular action. Therefore, we suggest that this theory might, in its strongest version, generate a valuation function similar to that found in Eq. (4).

$$\begin{aligned}&\text{if } (r_t(a) > 0), \quad Q_t(a) = Q_{t-1}(a) + 1\\ &\text{elseif } (r_t(a) < 0), \quad Q_t(a) = Q_{t-1}(a) - 1\\ &\text{otherwise}, \quad Q_t(a) = Q_{t-1}(a).\end{aligned} \tag{4}$$

This approach is targeted at modelling the VMF patient data, and only has one free parameter, ϵ, which is investigated across the range 0 to 1, in 0.01 steps.

Error-driven learning. This valuation function is loosely inspired by the delta-rule (O'Reilly & Munakata, 2000) and is set out in Eq. (5).

$$Q_t(a) = Q_{t-1}(a) + \gamma(r_t(a) - Q_{t-1}(a)). \tag{5}$$

In the *error-driven* valuation function (Eq. (5)), γ denotes the learning rate. The *error-driven* version of the model has two free parameters, γ and ϵ, both of which were explored exhaustively using a range from 0 to 1 with 0.01 steps, producing 10,000 different parameter settings. The 'myopia' for future consequences theory will be captured within the parameter space of this model.

Error-frequency learning. A potentially less extreme instantiation of the Frank and Claus (2006) frequency-driven theory might include an extra learning rate parameter. This would allow the inclusion of a factor for how responsiveness to recent and past results affects current behaviour. This variation of the *frequency-driven* model, the *error-frequency* model is presented in Eq. (6).

$$\begin{aligned}
&\text{if } (r_t(a) > 0), \quad Q_t(a) = Q_{t-1}(a) + \gamma(1 - Q_{t-1}(a)) \\
&\text{elseif } (r_t(a) < 0), \quad Q_t(a) = Q_{t-1}(a) - \gamma(1 + Q_{t-1}(a)) \\
&\text{otherwise}, \quad Q_t(a) = (1 - \gamma) \cdot Q_{t-1}(a).
\end{aligned} \tag{6}$$

Similarly to the other two parameter model, the parameters γ and ϵ were explored exhaustively using a range from 0 to 1 with 0.01 steps, producing 10,000 different parameter settings.

Reversal learning. This approach takes the *error-driven* model, and adds our interpretation of a reversal learning deficit. The interpretation being that, when a choice/stimulus is currently associated with a positive estimation of reward ($Q_{t-1}(a) > 0$), but the next result shifts the estimation to predict a negative outcome ($Q_t(a) < 0$), or vice versa, then learning is slowed. To implement this, an additional parameter λ (reversal-deficit rate) was added to Eq. (5); this only has an impact when there is a reversal in expectations/estimations (see Eq. (7)).

$$\begin{aligned}
&\text{if } (\text{sign}(Q_{t-1}(a)) = \text{sign}(Z) \vee Q_{t-1}(a) = 0), \\
&\quad Q_t(a) = Z \\
&\text{otherwise}, \\
&\quad Q_t(a) = Q_{t-1}(a) + \lambda \cdot \gamma(r_t(a) - Q_{t-1}(a)) \\
&\text{where} \quad Z = Q_{t-1}(a) + \gamma(r_t(a) - Q_{t-1}(a)).
\end{aligned} \tag{7}$$

The function sign(x) denotes the sign of x. The addition of λ gives this model 3 free parameters, which substantially increases the parameter space. Therefore, to reduce the number of simulations, 0.05 steps were taken to investigate the parameter space, which has a range from 0 to 1 for each parameter. This condenses the number of investigations from the 100,000 required when using 0.01 steps to 8,000. If λ is small then there is a severe deficit in reversal learning that would result in perseveration errors following a reversal. Perseveration errors represent the main form of reversal learning deficits found in VMF patients on simple reversal learning tasks (Fellows & Farah, 2003) (i.e. not a probabilistic reversal learning task).

Error-variance learning. To provide a less extreme test of the *variance-driven* hypothesis (Sanfey et al., 2003), an *error-variance* model is also considered. This model (see Eq. (8)) weights an estimation of the mean result, based on the *error-driven* model, against the standard deviation for that choice using a free parameter θ (the variance-weight) that ranges from 0 to 1.

$$\begin{aligned}
Q_T(a) = {} & (1 - \theta) \cdot (Q_{T-1}(a) + \gamma(r_T(a) - Q_{T-1}(a))) \\
& + \theta \cdot \sqrt{\text{Eq. (3)}}.
\end{aligned} \tag{8}$$

The 3 free parameters are investigated in combination, using 0.05 steps for each parameter.

Error-valence learning. This version examines whether there is a difference in how participants value losses vs. gains. This approach is again based on the *error-driven* model, but in this case the return $r_t(a)$ is passed through a valence function (see Eq. (9)). The *error-valence* model was used to test whether NHCs and VMF patients value losses and gains equally. A similar valence parameter has been used in other modelling work on the IGT (Busemeyer & Stout, 2002).

$$r_t(a) = w \cdot R_t(a) + (1 - w) \cdot L_t(a). \tag{9}$$

The parameter, $0 < w < 1$, is comparable to the attention weight parameter in Busemeyer and Stout (2002). However, here it is called the valence weight. $R_t(a)$ denotes the positive reward resulting from a card selection, while $L_t(a)$ symbolizes the loss element of the result. This model has 3 free parameters, which are investigated in combination, with 0.05 steps for each parameter. As is the case with the *reversal learning* model, this requires simulations at each of the 8000 possible parameter settings.

Working-memory model. This model examines whether memory can be further dissected beyond the learning rate parameter. This is achieved with a parameter that allows for decay in retained information. The working-memory valuation function is presented in Eq. (10).

$$\begin{aligned}
&\forall i (1 \le i \le N) \cdot \quad \text{if } (a_c = a_i), \\
&Q_t(a_i) = Q_{t-1}(a_i) + \gamma(r_t(a_i) - Q_{t-1}(a_i)) \\
&\text{otherwise}, \\
&Q_t(a_i) = \rho \cdot Q_{t-1}(a_i).
\end{aligned} \tag{10}$$

The parameter, $0 < \rho < 1$, influences the amount reward information decays when that specific choice is not selected and is called the working memory parameter. The current chosen action is denoted by a_c. This model has 3 free parameters, which are investigated in combination, with 0.05 steps for each parameter.

The human profiles used to compare the simulated data with come from an amalgamation of data from relevant studies, comprising, Bechara et al. (1999, 2000), Clark et al. (2003), Bechara and Damasio (2002) and Bechara et al. (2002). The net scores can be found in Table 1. In addition, this table shows a comparison between the human data and a random strategy to test whether the human data are significantly different from chance. We describe a random strategy as a binomial

Table 1
An amalgamation of net score data from relevant studies, comprising, Bechara et al. (1999, 2000), Clark et al. (2003), Bechara and Damasio (2002) and Bechara et al. (2002)

Net score blocks Task variant	1–20	21–40	41–60	61–80	81–100	Total	No. of participants
ABCD–N	−5.9	2.3	7.1	8.6	9.7	21.8	54
Sig($P < 0.05$)	Yes	Yes	Yes	Yes	Yes	Yes	
Sig($P < 0.001$)	Yes	Yes	Yes	Yes	Yes	Yes	
ABCD–V	−4.7	0.7	2.0	−3.0	−4.2	−9.2	15
Sig($P < 0.05$)	Yes	No	No	Yes	Yes	Yes	
Sig($P < 0.001$)	No	No	No	No	No	No	
EFGH–N	1.0	6.5	10.0	9.0	8.5	35	20
Sig($P < 0.05$)	No	Yes	Yes	Yes	Yes	Yes	
Sig($P < 0.001$)	No	Yes	Yes	Yes	Yes	Yes	
EFGH–V	−2.0	0.0	0.0	1.0	3.0	2	10
Sig($P < 0.05$)	No	No	No	No	No	No	
Sig($P < 0.001$)	No	No	No	No	No	No	
A′B′C′D′–N	−4.7	3.7	7.0	7.4	7.7	21.1	66
Sig($P < 0.05$)	Yes	Yes	Yes	Yes	Yes	Yes	
Sig($P < 0.001$)	Yes	Yes	Yes	Yes	Yes	Yes	
A′B′C′D′–V	−6.1	0.0	−3.4	−3.1	−4.7	−17.3	18
Sig($P < 0.05$)	Yes	No	Yes	Yes	Yes	Yes	
Sig($P < 0.001$)	Yes	No	No	No	No	Yes	
E′F′G′H′–N	1.4	6.4	11.4	10.1	9.6	34.8	36
Sig($P < 0.05$)	No	Yes	Yes	Yes	Yes	Yes	
Sig($P < 0.001$)	No	Yes	Yes	Yes	Yes	Yes	
E′F′G′H′–V	−1.6	−3.9	−3.1	0.0	−5.5	−14.1	16
Sig($P < 0.05$)	No	No	No	No	Yes	Yes	
Sig($P < 0.001$)	No	No	No	No	No	No	

The nomenclature used to describe the task variants is their version names with—N for normal healthy controls and—V for VMF net scores. In addition, the table includes the results of significance testing (using Student's t-test) with a null hypothesis (NH) that the net scores are random (i.e. a comparison of the human data to a binomial distribution with $\Pr = 0.5$). The alternate hypothesis (AH) is that they have a non-random strategy. The answer of yes or no signifies whether the NH can be rejected at $P < 0.05$ and $P < 0.001$.

distribution of net scores. It is possible to use a Bernoulli process ($\Pr = 0.5$), because, even though there are usually four choices, net scores only differentiate between advantageous and disadvantageous choices. (Further details of these analyses can be found in Kalidindi, Bowman, and Wyble (2005).)

4.1. Model selection

When considering a number of models and trying to decide which is most likely to provide a good description of the underlying cognitive processes it is important to consider a number of factors. These have been summarized by Jacobs and Grainger (1994) and set out in Pitt, Myung, and Zhang (2002) to include (a) how plausible are the model's assumptions (biologically and psychologically); (b) is the theoretical explanation reasonable and consistent with current knowledge; (c) can the parameters or parts of the model be interpreted in a sensible way; (d) how good is the fit to the observed data; (e) how generalizable is the model (i.e. does it predict future/new data); and (f) how complex is it?

The discussion section of this paper examines the models' relationships to the first three factors. The two factors, (d) and (f), goodness of fit and complexity are weighed against one-another using the adjusted root mean square deviation (RMSD) (Pitt et al., 2002) found in Eq. (11). It is appropriate to use RMSD rather than a maximum likelihood method as we are working from means of the data from the literature, and not the original distribution of data across individual participants. Factor (e) can be used to test competing models when they are similar in terms of RMSD.

$$\mathrm{RMSD} = \sqrt{\mathrm{SSE}/(N-k)}. \tag{11}$$

SSE denotes the sum of the squared deviations between the observed and predicted data. N represents the number of data points fitted and k denotes the number of parameters in the model. In the current study, N is set to 5 to reflect the five net scores (1 for each block) per version of the IGT. A total is then presented, generated from the sum of RMSDs calculated for each version of the task. These can be found in Table 2.

5. Results and analysis

In terms of RMSD, the *error-driven* model with $\epsilon = 0.70$ and $\gamma = 0.90$ (RMSD = 10.8) provides the best match to human VMF patients, see Table 2. This is further supported by three other more complex models (*reversal learning*, *error-variance* and *error-valence*) collapsing to this nested version.

Table 2
The 'best' parameter settings for all models across the ABCD, A′B′C′D′, EFGH and E′F′G′H′ versions of the Iowa Gambling Task (IGT)

Parameters Models	ϵ	γ	λ	w	θ	ρ	RMSD	'Shuff' RMSD
Sample-averaging								
Normals	0.60	*	*	*	*	*	8.0	(1.9)
VMF	0.98	*	*	*	*	*	14.6	
Variance-driven								
Normals	0.03	*	*	*	*	*	29.8	
VMF	0.98	*	*	*	*	*	13.7	
Frequency-driven								
Normals	0.55	*	*	*	*	*	16.7	
VMF	0.98	*	*	*	*	*	14.5	
Error-driven								
Normals	0.57	0.16	*	*	*	*	10.0	(1.6)
VMF	0.70	0.90	*	*	*	*	10.8	
Error-frequency								
Normals	0.54	0.08	*	*	*	*	21.3	
VMF	0.90	0.38	*	*	*	*	15.5	
Reversal learning								
Normals	0.49	0.40	0.41	*	*	*	9.8	(2.6)
VMF	0.71	0.92	0.96	*	*	*	13.7	
Error-variance								
Normals	0.40	0.49	*	*	0.15	*	9.7	(7.3)
VMF	0.70	0.94	*	*	0.05	*	13.6	
Error-valence								
Normals	0.41	0.40	*	0.67	*	*	9.7	(5.9)
VMF	0.68	0.92	*	0.51	*	*	13.6	
Working-memory								
Normals	0.59	0.18	*	*	*	1	11.8	
VMF	0.58	0.93	*	*	*	0.88	13.2	

RMSD denotes the sum of parameter adjusted root mean square deviations for the 4 mentioned versions of the IGT, averaged across the top five parameter settings. The final column ('Shuff' RMSD) scores in brackets are for the 'shuffled' ABCD version, used to test the generalizability of those models.

For example, for the *error-valence* model, if the valence weight is $w = 0.5$ then it becomes the *error-driven* model as losses and gains are equally weighted. This is almost the case ($w = 0.51$). Therefore, since the values for ϵ and γ (see VMF models in Table 2) are almost the same for both models, we can reasonably treat them as the same model. Similarly, the extra parameter in each of the *reversal learning* and *error-variance* models is set to levels close to them being an *error-driven* model, with ϵ and γ again being close to $\epsilon = 0.70$ and $\gamma = 0.90$.

The only model that offers a potentially different interpretation of the results is the *working-memory* model ($\epsilon = 0.58$, $\gamma = 0.93$ and $\rho = 0.88$). It is similar in terms of learning rate to all the models mentioned above, however, unlike those models it does not require an increase in exploration rate to reproduce the VMF patient data compared to NHCs. But, the *working-memory* model's RMSD is bigger (13.2) than that for the *error-driven* model (10.8). Again, suggesting that the *error-driven* model is the best match.

In the case of NHCs, the *sample-averaging* model ($\epsilon = 0.60$) has the lowest RMSD = 8.0. However, within its parameter space it does not contain a good model of VMF patients. The closest the *sample-averaging* model can get to a VMF patient profile is by becoming random, $\epsilon = 0.98$. This is the maximum value for the exploration rate because we are averaging over the best five parameter settings and this model only has one parameter. However, clearly from Table 1, there are a number of blocks where net scores for VMF patients are significantly different from a random profile. The next best NHC models are the *error-valence* and *error-variance* models, both with an RMSD = 9.7. Therefore, it is difficult to decide which model to select. One method to resolve this impasse is to see whether the models can generalize and replicate well a new data set. To test this property of generalization, the top five NHC models, with their current parameter setting where used to simulate NHCs on the 'shuffled' ABCD variant of the IGT (Fellows & Farah, 2005). The final column of Table 2 presents the RMSDs for the 'shuffled' ABCD variant ('Shuff' RMSD) in brackets for the relevant models. From this additional analysis, it is evident that the *error-valence* and *error-variance* models do not generalize well to this new data set, unlike the *sample-averaging*, *error-driven* and *reversal learning* models. The best model in terms of generalization is the *error-driven* model RMSD = 1.6. Therefore, the *error-valence* and *error-variance* models seem to have over fit the current data. Furthermore, when using RMSD to weigh goodness of fit against complexity, the *error-driven* model ($\epsilon = 0.57$ and $\gamma = 0.16$) only performs slightly worse than the *error-valence* and *error-variance* models with a RMSD = 10.0. In addition, were the RMSD scores for all four IGT versions added to the 'shuff'

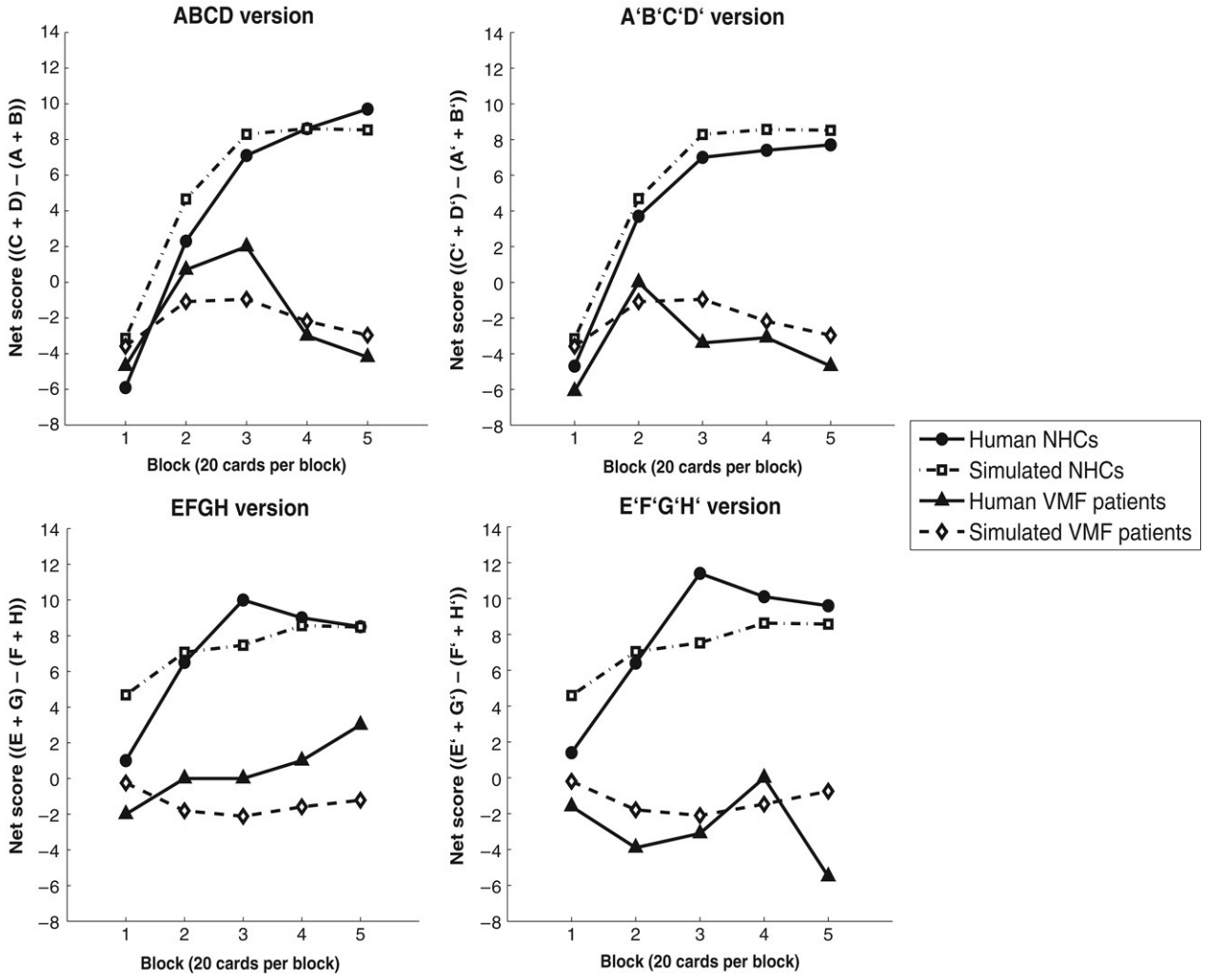

Fig. 1. Net score comparisons between the simulations and the human data for the four different versions of the IGT, using an *error-driven* valuation function and parameters $\epsilon = 0.57$ and $\gamma = 0.16$ for NHCs, and $\epsilon = 0.72$ and $\gamma = 0.90$ for VMF patients. 10,000 runs of the simulation were averaged to gain the simulated curves shown.

RMSD scores from simulations of the 'shuffled' ABCD variant then the NHC *error-driven* provides the lowest total RMSD. For these reasons, the *error-driven* models for NHCs and VMF patients are presented in Fig. 1 as the best models among the nine designs in the current study.

The *error-driven* ϵ-greedy model produces a good match to the human NHCs' profiles on all four versions of the IGT, apart from the initial data points on the E(′)F(′)G(′)H(′) versions. In addition, the *error-driven* model produces a good match to VMF patients on all versions of the task except the EFGH variant (see Fig. 1). Therefore, to further investigate this discrepancy between the simulations and the human VMF patient profile on the EFGH version the best parameter settings, in terms of RMSD and goodness of fit (summed square difference), for each model were found for this version and profile alone. The results of this analysis suggested that the best model was the *error-driven* model ($\epsilon = 0.27$ and $\gamma = 0.41$) with RMSD = 1.6. The next best model was a random model.

Neither the *variance-driven, error-frequency* or *frequency-driven* models produced a good match to the human NHC or VMF patient data. In respect to NHC profiles all three models performed poorly. The closest the *variance-driven* and *frequency-driven* models could get to a match for VMF patients was when they collapsed into a random model ($\epsilon = 0.98$). The *error-frequency* was simply a worse fit to the VMF patient data than the *error-driven* model. This is evidence against our interpretations of the two theories represented in these three models.

6. Discussion

In the following five sections of the discussion we consider each of the five theories in light of the results from the simulations and the current literature.

6.1. VMF patients have 'no preferences'

If this theory were implemented literally, then one would expect VMF patient behaviour on the IGT to be random (similar to a Bernoulli process, with Pr = 0.5). This is not the case as can be seen from Table 1, where blocks 1, 4 and 5 are found to be non-random ($P < 0.05$) for the ABCD version of the

IGT. Furthermore, there are non-random net scores in blocks 1, 3, 4 and 5 in the A′B′C′D′ version ($P < 0.05$) and in block 5 on the E′F′G′H′ version ($P < 0.05$). In addition, our *sample-averaging* ϵ-greedy models with exploration rates that tend towards $\epsilon = 1$ act randomly, but they do not replicate Bechara et al.'s VMF patients' profiles, only getting close to their behaviour on the EFGH version, which is the only profile that has no non-random blocks. In addition, the *sample-averaging* model does not produce a negative net score once the reversals have occurred. This is evident as the model produces a match to VMF patients, which is slightly worse than a random model.

The original 'no preferences' hypothesis presented in Sloman (2004), probably comes from anecdotal evidence, where a bilateral orbitofrontal lesion patient stated that she had "no preferences" (Sacks, 1998). Without testing this patient on the IGT it is impossible to know whether they would actually act randomly; that is, whether their subjective sense of not possessing preferences can be confirmed by objective evaluation. In addition, as our study is particularly focused on Bechara et al.'s VMF patients' behaviour on the IGT, we suggest that in this strong form, the 'no preferences' theory does not predict these patients' choices (apart from perhaps on the EFGH variant). However, our *error-driven* model (which does have a higher ϵ) does suggest that this theory, in a weaker form, might reflect an aspect of VMF patients' deficits on the IGT.

Another potential reason for seeming to have 'no preferences' was investigated using the *working-memory* model. A value of ρ near to zero would suggest a severe working memory deficit, as this would provide a time dependent loss of information. A high learning rate (γ), near to one, alone is not the same as a time only dependent working memory deficit, because with a high learning rate information is retained about the last selection of a choice, even after a number of other choice selections have been made.

For the best VMF patient *working-memory* model ($\epsilon = 0.58$, $\gamma = 0.93$ and $\rho = 0.88$), the learning rate is similar to that found in the *error-driven, reversal learning, error-variance* and *error-valence* VMF patient models. However, as mentioned, unlike those models the *working-memory* model does not have an increased exploration rate, but has one similar to the *error-driven* NHC model. Instead, the VMF model has a reduced value for ρ compared to its NHC parameter setting of 1, i.e. no time dependent working-memory deficit. Randomness could, we suggest, be thought of as either (1) an inherent property of the system; or (2) a way of describing the properties of a system for which the underlying mechanisms are still unknown. Therefore, it could be argued that the *working-memory* model offers an explanation/mechanism for the increase in exploration rate through its working-memory parameter ρ. In that, VMF patients might seem to act more randomly than NHCs because information retained about choices decays slightly, whenever that item is not selected. Therefore, although the *working-memory* model is more complex than the *error-driven* model, as suggested by the model's RMSD score (note, they each have a similar goodness of fit to the data), the *working-memory* model potentially elaborates on the theoretical story developed with the *error-driven* model. A small working-memory deficit in these VMF patients is consistent with the finding that some of them perform abnormally on delayed response tasks (Bechara et al., 1998).

6.2. VMF patients are 'risk-seeking'

If VMF patients' main deficit were risk-seeking, one would expect them to select more from the advantageous decks in the E(′)F(′)G(′)H(′) variants than normals. This is not supported by the human VMF patient data or our *variance-driven* model, which requires an exploration rate near to 1, to 'best' replicate the human VMF patients. Furthermore, the less extreme version of a variance driven model, the *error-variance* model collapses to a version of the *error-driven* model when generating its best match to the VMF data. The *error-variance* model spans the range of models from the *error-driven* model to the more extreme *variance-driven* model.

This provides support for the view that variance information is not used in these VMF patients' decision making on the IGT. Furthermore, although the *error-variance* model produces a good fit to the NHC data presented for the A(′)B(′)C(′)D(′) and E(′)F(′)G(′)H(′) versions it seems to have over fit the data and does not generalize well to additional NHC data taken from Fellows and Farah (2005)'s 'shuffled' ABCD variant.

Sanfey et al. (2003) found that there were two subgroups within their VMF population, one risk-VMF and the other safe-VMF. This could also be the case in Bechara et al.'s VMF population, but we are not aware of such an analysis having been performed on Bechara et al.'s data sets. Therefore, with the present data, the 'risk-seeking' interpretation does not seem to be the main cause of Bechara et al.'s VMF patients' deficits on the IGT.

6.3. VMF patients are only influenced by frequency

A *frequency-driven* model does not replicate the human VMF patients' data or the human NHCs' data. This is because the simulations for NHCs select more frequently from the decks with the lowest frequency of losses, decks B(′) and D(′) in the A(′)B(′)C(′)D(′) versions and deck G(′) in the E(′)F(′)G(′)H(′) versions. However, a model with a penchant for these decks does not recreate the required NHC profile. Interestingly, schizophrenia patients were found to have a preference for decks B and D in the ABCD version of the task (Shurman, Horan, & Nuechterlein, 2005).

The main motivation for this model was to discover whether a *frequency-driven* approach would replicate the human VMF patients' data. The *frequency-driven* model did not achieve this aim, with the 'best' match from this model requiring an exploration rate $\epsilon = 0.98$. Therefore, effectively, the simulation was acting randomly and thus, was not using the frequency information.

In the case of the *error-frequency* model, the best parameter settings for NHCs produce a poor match to the human data. For the best match to VMF patients, this model is not as effective

as the *error-driven* model, but does produce an improved match compared to that of the NHC model to the human NHC profile. Therefore, it raises the possibility that further investigation of these kinds of models might produce a reasonable match to VMF patients. A possible model to be investigated in future, might contain a magnitude compressing function to test a model that ranges from the current *error-driven* model to the *error-frequency* model. This would more exhaustively test the Frank and Claus theory that the basal ganglia might hold a frequency-based representation of results.

6.4. VMF patients have a 'myopia' for future consequences

This hypothesis, suggested by Bechara et al. (2000), is based directly on the VMF patient group who are the main focus of this paper. In addition, this is the theory that most closely reflects what can be gleaned from the NHC ($\epsilon = 0.57$ and $\gamma = 0.16$) and VMF patient ($\epsilon = 0.70$ and $\gamma = 0.90$) *error-driven* ϵ-greedy models. Furthermore, strong support for the *error-driven* model is provided by the fact that the more complex *error-valence*, *error-variance* and *reversal learning* models provide their best replication of the human VMF data when they approximate the *error-driven* model ($\epsilon = 0.70$ and $\gamma = 0.90$).

However, an acceptance of the 'myopia' for future consequences hypothesis is dependent on one's interpretation of the theory. The theory is closely related to the 'somatic-marker' hypothesis which suggests that bodily signals representing past experiences are not available to guide decision making in VMF patients (Damasio, 1994). Therefore, a possible operational definition of the theory for the IGT could be that subjects with this deficit are driven more by the immediate value of a choice rather than the mean value of repeating that choice: the future consequences. Again, here there is ambiguity, what is meant by immediate value? This could be the regular value, e.g. in the ABCD version, the guaranteed \$100 wins in decks A and B. Alternatively, more closely related to the current models, the immediate value could reflect the current or most recent return from a choice. However, in the IGT, these concepts overlap. Therefore, for greater clarity, we suggest our models offer a specific form of 'myopia' for future consequences based on our operational definition of this complex concept. VMF patients are driven more by the recent value of stimuli/choices than NHCs (higher γ) and that they ignore this information more often than NHCs (higher ϵ), that is VMF patients are more reactive (higher γ) and less strategic (higher ϵ) than NHCs.

That VMF patients are less strategic than NHCs, might provide a possible reason for their quicker return to a deck after receiving a loss from that deck (Bechara, Damasio, Tranel, & Damasio, 2005). However, as mentioned, this could be due to a working memory deficit (Bechara et al., 1998), which might be a causal reason for VMF patients' less strategic behaviour. A potential reason for VMF patients being more reactive could relate to the theory that frontal patients are often unable to transcend the default mode (Mesulam, 2002), which is evident in utilization behaviour. (Lhermitte, 1983, 1986; Shallice, Burgess, Schon, & Baxter, 1989). They are only able to see the immediate/recent value of a stimulus or choice.

The VMF patient model ($\epsilon = 0.70$ and $\gamma = 0.90$) further suggests that VMF patients do not average over the results from a number of selections. Therefore, they are more frequently influenced by regular wins or losses, than uncertain events, as these are quickly overridden by new results. This could be consistent with rodent OFC lesion studies, where lesions cause increased preference for small certain reinforcers over larger uncertain ones (Mobini et al., 2002). The *error-driven* model suggests that certain rewards would keep their value, whereas uncertain rewards might only retain their value for a short time after they are received, as when they are withheld their positive stimulus–reinforcement association would quite quickly be devalued. In addition, Doya (2002) has suggested that acetylcholine might represent the learning rate in a reinforcement learning model similar to the current error-learning ϵ-greedy models and this neuromodulator is known to have significant associations with OFC and medial prefrontal function (Arsten & Robbins, 2002). Our result is further corroborated by Yechiam et al. (2005), who used a model comparable to the *error-valence* model, which they tested with data from the same VMF patients on the ABCD version of the task. They found the similar result of increased randomness and a higher learning rate compared to NHCs, with only a very small increased responsiveness to losses over gains.

6.5. VMF patients have a reversal learning deficit

The 'best' VMF patient *reversal learning* model ($\epsilon = 0.70$, $\gamma = 0.96$, $\lambda = 0.98$) suggests that VMF patients have less of a reversal learning deficit on the IGT than NHCs ($\epsilon = 0.49$, $\gamma = 0.40$, $\lambda = 0.41$). In addition, a superior ability to learn reversals has some support from three other *error-based* models, the *error-variance*, *error-driven* and *error-valence* models, in that these models all show a higher learning rate in the VMF patient versions than in their NHC versions. One consequence (amongst a number of others) of a smaller learning rate would be a tendency to perseverate. This is in conflict with the simple and probabilistic reversal learning literature, where NHCs are faster than VMFs at learning reversals (Dias, Robbins, & Roberts, 1997; Fellows & Farah, 2003; Hornak et al., 2004; Rolls et al., 1994).

In addition, there is direct evidence for a reversal learning deficit in VMF patients on the IGT, which arises from a recent study. Fellows and Farah (2005) proposed that reversal learning deficits found in their VMF patients were the cause of such patients' poor performance on the IGT. To test their proposal, they created a 'shuffled' version of the ABCD IGT, where the score card was rearranged to remove the reversal. A reversal occurs when the mean return for advantageous decks becomes greater than that for the disadvantageous decks. In the standard ABCD version, this occurs at card 9 for deck B and card 5 for deck A (see the score-card in Bechara et al. (2000)). Fellows and Farah found that their VMF patients did not perform significantly differently from their NHCs on the 'shuffled' ABCD version. In addition, they found a positive correlation

between improvement on the 'shuffled' ABCD version and the number of reversal errors in a simple reversal learning task (Fellows & Farah, 2003, 2005).

However, it is unclear whether the VMF patients in Fellows and Farah (2005) are directly comparable to those from Bechara et al.'s studies. This is because Fellows and Farah's VMF patients continue to improve their net scores over the five blocks of the standard ABCD IGT (the profile is similar to that of Bechara et al.'s VMF patients on the EFGH IGT (see Fig. 1)). This is unlike Bechara et al.'s VMF patients, who improve their net scores and then actively behave disadvantageously in blocks 4 and 5. This difference in behaviour on the ABCD IGT could be due to Fellows and Farah's VMF patients being selected just on lesion location and not on both everyday life decision making deficits and lesion location, as Bechara et al.'s VMF patients were.

Furthermore, Fellows and Farah's VMF patients' with the worst reversal learning deficits have a focus of lesions in the left posteromedial orbitofrontal cortex, and the net score profile for their VMF patients on the standard ABCD IGT is similar to that of patients with unilateral left-VMF damage (Tranel et al., 2002) and those with left unilateral frontal lesions (Clark et al., 2003). Importantly, it should also be noted that the improvement in net scores for Fellows and Farah's VMF patients on the 'shuffled' ABCD IGT were produced almost entirely by 3 out of the 9 VMF patients, who all nearly doubled their net scores, and each selected over 90 advantageous cards. This raises the possibility that these 3 high performers were able to simplify the task, by completely ignoring the disadvantageous decks after receiving initial losses from these decks. They were able to apply a fixed rule, and did not continue to explore. The three high scoring VMF patients had larger net scores than the average NHC on the 'shuffled' ABCD IGT, suggesting on average that NHCs did continue to explore the disadvantageous decks, even after initial losses on those decks.

Moreover, neither Bechara et al. nor Fellows and Farah's VMF patient groups suffered from perseverative reversal learning deficits on the standard ABCD IGT. For example, Fellows and Farah found that after the big loss ($1250) on deck B (card 9), 8 out of their 9 VMF patients stopped selecting from deck B (L.K. Fellows, personal communication). A similar normative move away from losses has been observed by Bechara et al. (2005). The difference between NHCs and VMF patients in both sets of studies is that the VMF patients return to deck B quicker than NHCs. This nuance is modelled in our simulations by the increased exploration rate in the VMF patient versions of the *error-based* models compared to the NHC versions, or by a decreased working-memory parameter in the *working-memory* model. In future work, we intend to model the net scores of the VMF patients from Fellows and Farah (2005) to further understand the differences between these two VMF patient groups and the role of the 'shuffled' ABCD IGT in developing current theory.

Other modelling work supports the use of an increased learning rate for modelling Huntington's disease (HD) patients who have similar deficits on the IGT to Bechara et al.'s VMF patients, and produce perseveration deficits on an extra-dimensional shift task and a probabilistic reversal task (Lawrence, Sahakian, Rogers, Hodges, & Robbins, 1999). HD patients, who were tested on the ABCD IGT (Stout, Rodawalt, & Siemers, 2001), were modelled with an increased learning rate compared to NHCs (Busemeyer & Stout, 2002). HD is a progressive neurodegenerative syndrome, which causes particular damage to the caudate nucleus and putamen structures within the striatum. This suggests that differential damage to cortico-striatal loops may cause overlapping deficits. Furthermore, Busemeyer and Stout (2002) do not specifically address the issue of reversal learning deficits as a possible underlying cause for HD patients' IGT deficits.

Overall, the *error-based* simulations suggest that for Bechara et al.'s VMF patients, a reversal learning deficit is not the cause of their poor performance on the IGT. If they had a 'typical' reversal learning deficit, one would expect them to perseverate at the reversal, which they do not. However, evidence from Fellows and Farah (2005) does suggest that for a subset of their VMF population, removal of the reversal does allow them to improve their net scores, but that their deficit on the standard ABCD IGT is not due to perseveration errors (in the sense of moving from deck B). Therefore, these results still require further investigation, and possibly modelling. In the following section we consider some issues pertaining to the differences between simple reversal learning tasks and the IGT.

6.6. The IGT and simple reversal learning tasks: How similar are they?

There are important differences between the IGT and simple reversal learning tasks (SRLT)s (e.g. Fellows and Farah (2003)). Firstly, in an SRLT, performance is improved when recent results are given most importance. "If a choice is positive, then pick that stimulus again and, if it is negative then pick the other stimulus." This parsimonious rule allows the participant to be more or less optimal in response to change at the reversals. This type of strategy could plausibly have been used by NHCs in Fellows and Farah (2003), because the number of errors they generated on the SRLT seem to average an error per reversal. In addition, the task instructions for the SRLT state that only one of the choices represents a win at one time, while the other results in a loss (L.K. Fellows, personal communication). Therefore, in contrast, if participants were to take into account the *cumulative* amount of wins received from a stimulus and evaluate this against the loss at the reversal, one might, expect them to perseverate.

In addition, in the SRLT, change in one stimulus reflects the automatic change in the other choice, not in magnitude but in direction, from positive to negative or vice versa. However, in the IGT, the initial rational behaviour is to select from the disadvantageous packs (Maia & McClelland, 2004), then after the reversal (card 9, deck B and card 5, for deck A, (see the score-card of Bechara et al. (2000))), the participant should begin to select more from the advantageous packs. However, once the participant begins selecting more from the advantageous packs, the 'rational' behaviour is to continue

selecting from the advantageous decks even though there might be an occasional loss. In the IGT, in this sense, there is only one reversal, which can only be correctly evaluated by comparing many or all results from the different decks. Therefore, successful IGT performance requires a longer memory of magnitude and direction of results, rather than simply which choice last provided a positive or negative result, as in the SRLT.

Thus, the 'rational' behaviour for these two tasks, in terms of learning rates, would be a (relatively) low learning rate for the IGT and a high one for the SRLT. This represents an inherent difference in the two tasks and may offer a reason for the seemingly problematic high learning rates presented in our models for VMF patients on the IGT. As mentioned earlier, one consequence of a low learning rate would be increased perseveration errors on an SRLT.

This raises at least three possibilities for the current VMF patient model.

(1) Bechara et al.'s VMF patients have a high learning rate in both tasks and therefore, do not have a perseverative reversal learning deficit on the SRLT. This is still an empirical question as they have not been, to our knowledge, tested on such a task.

(2) At least two studies have suggested that frontal regions are particularly important for reversal learning rather than initial stimulus–reinforcement acquisition (Daum, Schugens, Channon, Polkey, & Gray, 1991; Fellows & Farah, 2003). Therefore, without inhibition from OFC regions, representations in posterior cortex might result in prepotent stimulus-associations driving behaviour (Knight & Stuss, 2002). In the SRLT, one might speculate that, the initial rewarding stimulus becomes prepotent due to repeated positive reinforcement (arising from a well-established temporal cortex representation (Daum et al., 1991) and selection (motor or cognitive impulsivity)). This repetition of selection is necessary if a VMF patient is to reach a reversal in the first place. For example, in Rolls et al. (1994), the criterion before a reversal is 9 correct responses out of the last 10 trials and in Fellows and Farah (2003) it is 8 correct selections in a row. Therefore, in VMF patients on an SRLT this prepotent behaviour might act like a low learning rate. However, in NHCs intact OFC regions provide the task appropriate cognitive and emotional flexibility to inhibit this prepotent action and produce single trial reversals by applying the rule in working memory to swap the stimulus-associations for both the positive and negative stimuli (Deco & Rolls, 2005), generating an effect similar to a high learning rate.

In the IGT, the requirement is to utilize information from a number of selections to gain a fairly accurate expectation of the result of a choice. VMF patients have limited access to at least one such behaviourally important and flexible representation found in intact OFC regions (Roesch & Schoenbaum, 2006; Rolls, 2004). We speculate that VMF patients are, again, strongly influenced by stimulus–reinforcement associations in posterior cortex (Knight & Stuss, 2002). However, we speculate, due to the frequent changes in returns on the IGT, VMF patients, more or less, produce a new stimulus–reinforcement association for each new result in posterior cortex. This would act like the high learning rate found in the current simulations, therefore, suggesting that without intact OFC regions, posterior areas are less able to reflect 'averaged', and therefore more accurate, expectancies. This process could be consistent with the disinhibited ERPs (i.e. P3 amplitudes) found in OFC patients compared to controls for novel and emotional stimuli (Knight & Scabini, 1998; Rule, Shimamura, & Knight, 2002). We would suggest that the rewards and punishments in the IGT are emotional (Rolls, 2005) and that the changing results provide frequent novelty. Furthermore, as selections are not as frequently repeated on the IGT compared to the SRLT there would be less of a requirement to inhibit motor or cognitive impulsivity.

(3) Not all difficulties on a simple reversal learning task are due to perseveration errors. In some cases, in monkeys, particularly those with more medial OFC lesions, there seems to be an inability to retain which item is currently rewarded. In these cases, the deficit is only evident after both stimuli have been associated with reward and non-reward. It seems the animal is unable to hold the required information about the current rewarding stimulus on-line in working-memory (Zald, 2006). A slight working-memory deficit in VMF patients has been suggested by our *working-memory* model and shown by Bechara et al. (1998).

Therefore, we suggest that it is still an open question whether these VMF patients have reversal learning deficits on an SRLT and the exact form they may take. However, in addition, we have offered a more detailed (2), but speculative, mechanism for possible differences in learning rates in the two tasks for both NHCs and VMF patients. Overall, our suggestion is that in the IGT, the role of VMF regions is in 'averaging' over results for a selection and to inhibit over-reaction to individual results. However, in relation to perseveration errors, in an SRLT the VMF is important in applying a rule from working memory to rapidly change expectancies to override a prepotent response. Therefore, the function of the VMF is context specific and so, when damaged, this is likely to have specific consequences for behaviour, but in all these contexts the VMF is important in flexible representations of expectancies (Roesch & Schoenbaum, 2006; Rolls, 2000; Tremblay & Schultz, 2000).

7. Conclusions

The models and simulations presented in this paper suggest that, among the theories considered, a 'myopia' for future consequences provides the best description for deficits found in Bechara at al.'s VMF patients on the IGT. However, it might be more appropriate and complete to suggest that VMF patients are less strategic (higher ϵ), possibly due to working-memory deficits, and more reactive (higher γ) than NHCs. Both these aspects are evident in VMF patients' real life behaviour (Damasio, 1994). Our work is supported by results from Yechiam et al. (2005) which, similarly to (Busemeyer & Stout, 2002), offers a relatively accessible method for explaining behaviour in complex tasks. Furthermore, in the current study cognitive modelling has been used to explicitly instantiate aspects of verbal theories to test their viability against human data.

Acknowledgements

This research has been supported by the Computing Laboratory at the University of Kent, UK. In addition, we appreciated the insightful comments and suggestions provided by the reviewers and Brad Wyble.

References

Arsten, A. F. T., & Robbins, T. W. (2002). Neurochemical modulation of prefrontal cortical function in humans and animals. In D. T. Stuss, & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 51–84). Oxford University Press.

Bechara, A., Damasio, A. R., Damasio, H., & Anderson, S. W. (1994). Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition*, *50*, 7–15.

Bechara, A., & Damasio, H. (2002). Decision-making and addiction (part I): Imparied activation of somatic states in substance dependent individuals when pondering decisions with negative consequences. *Neuropsychologia*, *40*, 1675–1689.

Bechara, A., Damasio, H., Damasio, A. R., & Lee, G. D. (1999). Different contributions of the human amygdala and ventromedial prefrontal cortex to decision-making. *The Journal of Neuroscience*, *19*(13), 5473–5481.

Bechara, A., Damasio, H., Tranel, D., & Anderson, S. W. (1998). Dissociation of working memory from decision making within the human prefrontal cortex. *Journal of Neuroscience*, *18*(1), 428–437.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (1997). Deciding advantageously before knowing the advantageous strategy. *Science*, *275*(5304), 1293–1295.

Bechara, A., Damasio, H., Tranel, D., & Damasio, A. R. (2005). The Iowa Gambling Task and the somatic marker hypothesis: Some questions and answers. *Trends in Cognitive Science*, *9*(4), 159–162.

Bechara, A., Dolan, S., & Damasio, H. (2002). Decision-making and addiction (part II): Myopia for future or hypersensitivity to reward?. *Neuropsychologia*, *40*, 1690–1705.

Bechara, A., Tranel, D., & Damasio, H. (2000). Characterization of the decision-making deficit of patients with ventromedial prefrontal cortex lesions. *Brain*, (123), 2189–2202.

Busemeyer, J. R., & Stout, J. C. (2002). A contribution of cognitive decision models to clinical assessment: Decomposing performance on the bechara gambling task. *Psychological Assessment*, *14*(3), 253–262.

Clark, L., Cools, R., & Robbins, T. W. (2004). The neuropsychology of ventral prefrontal cortex: Decision-making and reversal learning. *Brain and Cognition*, *55*, 41–53.

Clark, L., Manes, F., Antoun, N., Sahakian, B. J., & Robbins, T. W. (2003). The contributions of lesion laterality and lesion volume to decision-making impairment following frontal lobe damage. *Neuropsychologia*, *41*, 1474–1483.

Damasio, A. R. (1994). *Descartes error: Emotion, reason and the human brain*. New York: Grosset/Putnam.

Daum, I., Schugens, M. M., Channon, S., Polkey, C. E., & Gray, J. A. (1991). T-maze discrimination and reversal learning after unilateral temporal or frontal lobe lesions in man. *Cortex*, *27*(4), 613–622.

Deco, G., & Rolls, E. T. (2005). Synaptic and spiking dynamics underlying reward reversal in the orbitofrontal cortex. *Cerebral Cortex*, *15*, 15–30.

Dias, R., Robbins, T. W., & Roberts, A. C. (1997). Dissociable forms of inhibitory control within prefrontal cortex with an analog of wisconsin card sort test : Restriction to novel situations and independence from on-line processing. *The Journal of Neuroscience*, *17*(23), 9285–9297.

Doya, K. (2002). Metalearning and neuromodulation. *Neural Networks*, *15*, 495–506.

Dunn, B. D., Dalgleish, T., & Lawrence, A. (2006). The somatic marker hypothesis: A critical evaluation. *Neuroscience and Biobehavioural Reviews*, *30*(2), 239–271.

Fellows, L. K., & Farah, M. J. (2003). Ventromedial frontal cortex mediates affective shifting in humans: Evidence from a reversal learning paradigm. *Brain*, *126*, 1830–1837.

Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral Cortex*, *15*(1), 58–63.

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making and reversal. *Psychological Review*, *113*, 300–326.

Goldman-Rakic, P. S., & Leung, H. (2002). Functional architecture of the dorsolateral prefrontal cortex in monkeys and humans. In *Principles of frontal lobe function* (pp. 85–95). Oxford University Press.

Hinson, J. M., Jameson, T. L., & Whitney, P. (2002). Somatic markers, working memory, and decision making. *Cognitive, Affective, and Behavioral Neuroscience*, *2*(4), 341–353.

Hornak, J., O'Doherty, J., Bramham, J., Rolls, E. T., Morris, R. G., Bullock, P. R., et al. (2004). Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans. *Journal of Cognitive Neuroscience*, *16*(3), 463–478.

Jacobs, A. M., & Grainger, J. (1994). Models of visual word recognition - sampling the state of the art. *Journal of Experimental Psychology: Human Perception and Performance*, *29*, 1311–1334.

Kalidindi, K., Bowman, H., & Wyble, B. (2005). Using ϵ-greedy reinforcement methods to further understand ventromedial prefrontal patients' deficits on the Iowa gambling task. *Technical report 12-05*. Computing Laboratory, University of Kent. http://www.cs.kent.ac.uk/people/rpg/kk49/.

Knight, R. T., & Scabini, D. (1998). Anatomic bases of event-related potentials and their relationship to novelty detection in humans. *Journal of Clinical Neurophysiology*, *15*, 3–13.

Knight, R. T., & Stuss, D. (2002). Prefrontal cortex: The present and the future. In D. T. Stuss, & R. T. Knight (Eds.), *Principles of frontal lobe function* (pp. 573–597). Oxford University Press.

Lawrence, A. D., Sahakian, B. J., Rogers, R. D., Hodges, J. R., & Robbins, T. W. (1999). Discrimination, reversal, and shift learning in huntington's disease: Mechanisms of impaired response selection. *Neuropsychologia*, *37*, 1359–1374.

Lhermitte, F. (1983). 'Utilization behaviour' and its relation to lesions of the frontal lobes. *Brain*, *106*, 237–255.

Lhermitte, F. (1986). Human autonomy and the frontal lobes. Part II: Patient behaviour in complex and social situations: The "environmental dependency syndrome". *Annals of Neurology*, *19*(4), 335–343.

Maia, T. V., & McClelland, J. L. (2004). A reexamination of the evidence for the somatic marker hypothesis: What participants really know in Iowa Gambling Task. *Proceedings of the National Academy of Sciences*, *101*, 16075–16080.

Manes, F., Sahakian, B. J., Clark, L., Rogers, R. D., Antoun, N., Aitken, M., et al. (2002). Decision-making processes following damage to the prefrontal cortex. *Brain*, *125*, 624–639.

Mavaddat, N., Kirkpatrick, P., Rogers, R. D., & Sahakin, B. J. (2000). Deficits in decision-making with aneurysms of the anterior communicating artery. *Brain*, *123*(10), 2109–2117.

Mesulam, M. (2002). The human frontal lobes: Transcending the default mode through contingent encoding. In *Principles of frontal lobe function* (pp. 8–30). Oxford University Press.

Mobini, S., Body, S., Ho, M. Y., Bradshaw, C. M., Szabadi, E., Deakin, J. F., et al. (2002). Effects of lesions of the orbitofrontal cortex on sensitivity to delayed and probablisitic reinforcement. *Psychopharmacology*, *160*, 290–298.

O'Reilly, R., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience*. MIT Press.

Peters, E., & Slovic, P. (2000). The springs of action: Affective and analytical information processing in choice. *Personality and Social Psychology Bulletin*, *26*(12), 1465–1475.

Pitt, M. A., Myung, I. J., & Zhang, S. (2002). Toward a method of selecting among computational models of cognition. *Psychological Review*, *109*(3), 472–491.

Roesch, M., & Schoenbaum, G. (2006). From associations to expectancies: Orbitofrontal cortex as gateway between the limbic system and representational memory. In *The orbitofrontal cortex* (pp. 199–236). Oxford University Press.

Rogers, R. D., Everitt, B. J., Baldacchino, A., Blackshaw, A. J., Swainson, R., Wynne, K., et al. (1999). Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: Evidence for monoaminergic mechanisms. *Neuopsychopharmachology*, *20*, 322–339.

Rolls, E. T. (2000). The orbitofrontal cortex and reward. *Cerebral Cortex*, *10*, 284–294.

Rolls, E. T. (2004). The function of the orbitofrontal cortex. *Brain and Cognition*, *55*, 11–29.

Rolls, E. T. (2005). *Emotion explained*. Oxford University Press.

Rolls, E. T., Hornak, J., Wade, D., & McGrath, J. (1994). Emotion related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery and Psychiatry*, *57*, 1518–1524.

Rule, R. R., Shimamura, A. P., & Knight, R. T. (2002). Orbitofrontal cortex and dynamic filtering of emotional stimuli. *Cognitive, Affective and Behavioural Neuroscience*, *2*(3), 264–270.

Sacks, O. (1998). The man who mistook his wife for a hat. *Touchstone*.

Sanfey, A. G., Hastie, R., Colvin, M. K., & Grafman, J. (2003). Phineas gauged: Decision-making and the frontal lobes. *Neuropsychologia*, *41*, 1218–1229.

Shallice, T., Burgess, P. W., Schon, F., & Baxter, D. M. (1989). The origins of utiliztion behaviour. *Brain*, *112*, 1587–1598.

Shurman, B., Horan, W. P., & Nuechterlein, K. H. (2005). Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment in the Iowa Gambling Task. *Schizophrenia Research*, *72*, 215–224.

Sloman, A. (2004). What are emotion theories about? In *Invited talk at cross-disciplinary workshop on architectures for modeling emotion at the AAAI spring symposium at Stanford University in March 2004*. See: http://www.cs.bham.ac.uk/research/cogaff/sloman-aaai04-emotions.pdf.

Stone, V. E., Baron-Cohen, S., & Knight, R. T. (1998). Frontal lobe contributions to theory of mind. *Journal of Cognitive Neuroscience*, *10*, 640–656.

Stout, J. C., Rodawalt, W. C., & Siemers, E. R. (2001). Risky decision making in huntington's diesease. *Journal of the International Neuropsychological Society*, *7*, 92–101.

Sutton, R. S., & Barto, A. G. (1998). A bradford book. In *Reinforcement learning: An introduction*. The MIT Press.

Tranel, D., Bechara, A., & Denburg, N. L. (2002). Asymmetric functional roles of right and left ventromedial prefrontal corticies in social conduction, decision-making, and emotional processing. *Cortex*, *38*, 589–612.

Tremblay, L., & Schultz, W. (2000). Modifications of reward expectation-related neuronal activity during learning in primate orbitofrontal cortex. *The Journal of Neurophysiology*, *83*(4), 1877–1885.

Wagar, B. M., & Thagard, P. (2004). Spiking phineas gage: A neurocomputational theory of cognitive affective integration in decision making. *Psychological Review*, *111*(1), 67–79.

Yechiam, E., Busemeyer, J. R., Stout, J. C., & Bechara, A. (2005). Using cognitive models to map relations between neuropsychological disorders and human decision-making deficits. *Psychological Science*, *16*(12), 973–978.

Zald, D. H. (2006). Neuropsychological assessment of the orbitofrontal cortex. In *The orbitofrontal cortex* (pp. 449–480). Oxford University Press.