Centre for Cognitive Neuroscience and Cognitive Systems, Computer Science at Kent

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

Kiran Kalidindi, Howard Bowman and Brad Wyble

Technical Report No. 12-05

October 2005

Published by the Computing Laboratory,

University of Kent, Canterbury, Kent, CT2 7NF, UK

1 Introduction

The Iowa Gambling Task (IGT) (Bechara et al., 2000a) is an experimental paradigm developed by Antoine Bechara et al. in the Department of Neurology at the University of Iowa College of Medicine. The task 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 choice of action. Our aim is to find valuation functions, which describe the normative behaviour found in human groups on the IGT, using models based on ϵ-greedy methods (Sutton and Barto, 1998). 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 consequences. The task contains four decks, two of which are good (advantageous) and two of which are bad (disadvantageous). Players need to, through selection, learn which decks are best. Initially, the bad decks are seemingly the best, as they offer higher immediate reward but higher uncertain losses, which only becomes evident after some selections. But as the task progresses, normal healthy humans learn that the best decks are those which offer smaller immediate rewards, but also lower uncertain punishments, whereas VMF patients seem unable to fully use this distinction.

The paradigm has been used as a method for distinguishing aspects of decision making deficits in VMF patients compared to normal healthy controls (Bechara et al., 1994, 1997, 1999, 2000a; Fellows and Farah, 2005), and with various other frontal lesion patient groups (Manes et al., 2002; Fellows and Farah, 2005; Clark et al., 2003), including patients with unilateral VMF lesions (Tranel et al., 2002). The IGT has additionally been used to examine differences in performance between healthy controls and behaviour in substance abusers (Bechara and Damasio, 2002a; Bechara et al., 2002b; Grant et al., 2000), psychopaths (Schmitt et al., 1999), schizophrenics (Shurman et al., 2005), and patients with other brain diseases, including Parkinson's and Huntington's (Stout et al., 2001).

VMF patients and healthy controls act quite differently on the task, with normal controls developing an advantageous strategy. Whereas, VMF patients act disadvantageously, and fail to sustain an advantageous strategy. An overall advantageous strategy has been described as having a net score greater than 0 across all 100 trials, or disadvantageous when less than 0 (Bechara et al., 1998). We add the criterion that the mean score should be significantly different from random. This allows a simple definition of strategies: advantageous, random and disadvantageous. (A disadvantageous strategy is one with a negative net score (i.e. below zero), and significantly different from the actions of a random agent.) To prevent confusion, we describe a random strategy as a binomial distribution of net scores. This is possible, because even though there are usually four choices, net scores only differentiate between advantageous and disadvantageous choices. The net score is calculated by adding up the number of advantageous choices, selections from decks C and D, and subtracting the number of disadvantageous choices, from the other two decks, A and B (net score = (C+D) - (A+B)). Participants' net scores are often broken down into 5 blocks of 20 selections, to show how they evolve over the course of the task (Bechara et al., 2000a). This is the format that will be used in this paper. (Further details of the IGT can be found in Section 2).

Additionally, skin conductance responses (SCRs) have been recorded during the task and these have been considered strong evidence for the somatic marker theory proposed by Damasio (1994). It has been found that, as the task progresses, during deliberation of the next choice of deck, normal human controls begin to generate anticipatory SCRs (SCRs taken for the 5 second period before a choice is actioned) and that they are higher when that deliberation ends with selection of a disadvantageous choice (decks, A or B), compared to an advantageous selection (decks, C and D). This is considered further in section 2.1.

This is seemingly a classic reinforcement learning problem, which we are examining, using the ϵ-greedy action-value method (Sutton and Barto, 1998), to see if we can recreate the two data sets, i.e. how do the parameter settings vary according to the data set we are trying to reproduce? This could help us infer further possible mechanisms found in normal controls, and how these mechanisms can be altered to create VMF patient profiles. It should also help us better understand the structure of the underlying IGT paradigm.

2 Versions of the Iowa Gambling Task (IGT)

In the IGT, a player is initially restricted to four choices, where each choice represents a pick of the top card from one of the four decks, this card is then discarded. In the original ABCD version of the task, there are 40 cards per deck, and, because the total number of trials in the task is 100, it is possible for a deck to run out. When this occurs, selections must be made from the remaining decks. This is said to rarely occur with real subjects (Damasio, 1994). In the original ABCD version, decks A and B are

Deck	A	B	C	D
Mean ($)	-25	-25	25	25
Stand. Dev.	134.6	375.0	28.0	75

Table 1: **The mean amount won or lost per card selection and the standard deviation of these amounts for each of the four decks in the original ABCD version of the IGT.**

Deck	E	F	G	H
Mean ($)	25	-25	25	-25
Stand. Dev.	375.0	29.0	135.6	75

Table 2: **The mean amount won or lost per card selection and the standard deviation of these amounts for each of the four decks in the variant EFGH version of the IGT.**

considered disadvantageous, because, although they give larger regular rewards of $100 per card, the uncertain losses accrued make the decks lose $250 per 10 cards selected. However, on decks C and D, the regular wins are just $50, but the uncertain losses are also smaller, giving an average win of $250 per 10 cards selected from either deck. Therefore, decks A and B initially seem to be the best, as the decks are set up so the first losses in decks A and B do not occur until a few cards have been selected. Decks A and B then become detrimental to the accruement of wealth, with a mean loss of $ -25. (For the exact contingencies of the task see Bechara et al. (2000a).) It is important to note that decks A and B also have the largest variance in potential wins and losses per card, making them more 'risky' (see table 1).

A more or less reverse version of the ABCD task has also been developed: the EFGH version (Bechara et al., 2000a). Here, the advantageous decks are E and G, where there is a regular loss of $-100 per card and the uncertain chance of a high reward. This makes, these decks average a $250 win per 10 cards. Whereas, on decks F and H there is only a loss of $-50 per card with smaller average uncertain wins, providing an overall average loss of $-250 for every ten cards selected in those decks. Therefore, the EFGH version is different, not only because regular results are losses with infrequent wins, but also as the best decks, E and G, have the greatest variance from their relevant means. This is compared to the ABCD version, where the best decks, C and D, have the lowest variance, compared to the bad decks, A and B. The means and standard deviations for these versions can be found in tables 1 and 2 which clearly illustrates this difference.

The additional differences in standard deviations between decks, C and F and A and G, are due to variations in the distributions of the amounts of potential losses and wins. But, it should be noted that the frequency of uncertain losses and wins were consistent across corresponding decks, where the corresponding decks are: A and G, B and E, C and F, and D and H.

The details of the computerised variants of the IGT considered, can also be found in Bechara et al. (2000a). The first is the A’B’C’D’ variant, based on the original ABCD task, where decks A’ and B’ progressively get worse, causing increasing losses when repeatedly selected. Additionally, in this version C’ and D’ get better, by giving decreasing uncertain losses. 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 was 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 was 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 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.

These versions were also run on a computer and so wins and losses were more abstract than in the original ABCD and EFGH versions, where actual cards and facsimile money were used. The effect of this on participant behaviour is difficult to assess, as no comparisons between the physical card version and a matching computerized version, with the same contingencies and the same participant groups have been found in the literature (the original card-based ABCD task vs. computer-based ABCD variant of the IGT). But the healthy control profiles for the ABCD version verses the A’B’C’D’ version are similar. Normal controls’ profiles are equally similar for the two EFGH variants, which suggests the effects of the changing contingencies is ameliorated. A real test of the affect of the computerized abstraction is actually difficult because normal subjects seem to learn the principles of the paradigm and so cannot be tested on it more than once. Therefore, it would have to be tested normatively, with large groups. Another difference in versions A’B’C’D’ and E’F’G’H’ is that there are 60 cards per deck instead of the 40 cards per deck in the ABCD and EFGH versions. This reduces the chance of decks being exhausted during the task, which could change the number of options in later trials.

2.1 Interpretations of Variance and SCRs

Interestingly, differences in the variance of the good decks compared to the bad decks in the A'B'C'D and E'F'G'H' versions of the task, provides evidence against a prominent theory in the literature. An important part of somatic-marker theory claims that anticipatory SCRs generated by normal subjects when they are about to select a card, are the result of somatic-markers. Bechara et al. found that when normals start selecting advantageously, their anticipatory SCRs become higher prior to picking from bad decks than good decks. Additionally, Bechara et al. claimed that these SCRs drove decision making, even when normal subjects did not have conscious knowledge of the contingencies of the task (Bechara et al., 1997). This has been countered in a recent paper (Maia and McClelland, 2004), where a more comprehensive questionnaire was used to measure the conscious knowledge of the participants. Maia and McClelland (2004) claim that normal subjects have a fairly good knowledge of the task's contingencies quite early on in the game, and that it is unnecessary to claim a non-conscious biasing of decision making to understand behaviour in the IGT. Actually, an interesting finding is that normal subjects seemingly have a better knowledge of the task than is reflected in their behaviour: If participants picked only from the decks they thought were best they would get higher net scores than they do. A possible interpretation is that they always include some degree of exploration in their actions. It should be noted that the Maia and McClelland (2004) study was with young college students rather than VMF patient age matched controls (most VMF lesion patients would be in their 40s or older).

With this explanation, a return can be made to the discussion on differences in variance between the good and bad decks in the two ABCD and the two EFGH versions. It has been suggested by some researchers (Maia and McClelland, 2004; Tomb et al., 2002), that the higher anticipatory SCRs seen in normal human subjects, when they are picking from disadvantageous decks compared to advantageous decks, is a reflection of the higher variance in the bad decks, A and B, on the ABCD version. However, results from Bechara et al. (2002b); Bechara and Damasio (2002a) and our explanation of variance differences across versions, suggest that SCRs are likely to reflect more than variance alone. In the E'F'G'H' version (Bechara et al., 2002b), after about card 20, normal controls get higher anticipatory SCRs for the bad decks, F and H, than for good decks, E and G. This is even though the good decks, E and G, have the higher variance. However, the profiles for anticipatory SCRs for the good and bad decks, in the A'B'C'D' and the E'F'G'H' versions, are different over time. Bechara and Damasio (2002a) and

Bechara et al. (2002b) show that for decks A' and B' there is a tripling in the anticipatory SCRs, after the first 10 cards, where it remains for the rest of the task. The anticipatory SCRs for decks C' and D' are initially similar to those for decks A' and B', but remain fairly static across trials. However, in the E'F'G'H' version of the task, the bad decks, F' and G', have a smaller increase (they nearly double) in their anticipatory SCRs, compared with those for decks A' and B'. The increase is also much more gradual and continues throughout the task. The profile of anticipatory SCRs over trials, for good decks, E' and G', is also different from the good decks, C' and D', in the A'B'C'D' variant. For the good decks, E' and G', the anticipatory SCRs increase (by 50%) by the 11-20th cards, but then drops off quite substantially (to about half their initial value), making the SCRs less for the good decks than for the bad decks, F' and H'. Therefore, overall, even when the good decks have the highest variance, SCRs better reflect a valuation of which decks are good or bad. This is shown by the higher net scores during the periods of the task when they have higher SCRs for the bad decks than the good decks. Therefore, even though the anticipatory SCR profiles between the A'B'C'D' and the E'F'G'H' versions are different, and may contain an element of evaluation of risk, the strong claim that the SCRs only represent relative riskiness of the good and bad decks is not supported by this evidence.

3 Competing theories about VMF patient deficits

We now move to present four theories from the current literature, as the possible underlying causes for decision making deficits of bilateral VMF lesion patients in the Iowa Gambling Task (IGT).

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. 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, post injury. If this argument were taken literally then one would expect VMF patient behaviour on the IGT, to be random (using a binomial series, with P = 0.5). This is not the case as can be seen from table 3, where blocks 1, 4 and 5 are found to be non-random ($P < 0.05$) for the ABCD version of the IGT.

3.2 Reversal learning - Rolls, Cools and Fellows

In both monkeys and humans, frontal lobe lesions have been found to cause participants to perseverate in reversal learning tasks (Rolls et al., 1994; Rolls, 2000, 2004) and (Fellows and Farah, 2003, 2005; Clark et al., 2004; Hornak et al., 2004). Although the patients in the Rolls et al. (1994) study had diffuse frontal lesions, Fellows and Farah (2003, 2005)'s patients have more localised damage, with a focus of damage 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. 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.

The ventromedial region of the prefrontal cortex includes the medial sub-region of the orbitofrontal cortex (OFC), where the OFC also includes more lateral areas. The orbitofrontal cortex is strongly connected with reward and representations of stimulus-reinforcement associations. In monkeys, the orbitofrontal is important as a secondary olfactory association cortex, which represents the reward value of stimuli (Rolls, 2004).

3.2.1 Edmund T. Rolls - 'Emotions', reversal learning and the OFC

Rolls has related emotions and reversal learning. "For emotional behaviour, rapid learning is clearly important. For some types of emotional behaviour, rapid relearning of the reinforcement value of a stimulus may be useful." (Rolls, 1999).

From his monkey studies, Rolls has learnt from single cell recordings that neurons in inferior temporal cortex only recognise objects, but not there reward value. This he learnt by putting monkeys through a reversal task, where he found the response in the single neurons did not change with stimulus-reinforcement associations being reversed. Additionally, neurons in inferior temporal cortex are not affected by satiety.

An important discovery by Rolls has been in differentiating between 'wanting a food' and 'what a food is', which are represented in monkeys' secondary and primary taste cortices, respectively. The differences between these taste cortices is that satiety affects the responses of taste neurons in secondary taste cortex, but not in primary taste cortex. In that, hunger modulates the reward value of taste. This is similar to reported reductions in pleasantness of odour produced by feeding to satiety, in humans. Secondary taste cortex is also found in the orbitofrontal cortex.

Rolls has made single cell recordings in the OFC of monkeys, and found that areas associated with visual inputs of 'what', from the ventral stream, change their spiking rate during reversal of a visual image and a primary reinforcer, a sweet or aversive (saline) taste/food. In taste regions, when S^+ changes to S^-, we get a decrease in spiking rate, and, an opposing increase, when a stimulus S^- becomes S^+. This creates a positive correlation between firing rates in OFC visual neurons, and behaviour. In contrast, in an olfactory to taste reversal, when S^+ stimulus goes to S^-, the reversal causes an increased spiking rate. There is a negative correlation between spiking rate in olfactory OFC neurons and behaviour. But the reversal is slower. Rolls has found that in orbitofrontal olfactory regions 35% of neurons, in non-human primates, represent the reward value of olfactory stimuli (wanting). The other 65% were not influenced by whether an odour was associated with a glucose or saline taste (what). "An important principle in the representation of reward in the orbitofrontal cortex is that not only is there a detailed representation of different rewarding stimuli (including olfactory, taste, visual and textural stimuli), but also the reward value of each stimulus is updated continually" Rolls (2000). It has also being found that in the OFC, visual and taste pathways converge on the same neurons. The amygdala shares some of these functions, but seems to be slower at responding than the OFC, and is involved in learning associations between neutral stimuli and primary reinforcers (but not secondary reinforcers, like the OFC).

3.2.2 Lesley Fellows

In some ways Rolls' work sets the scene for the recent work of Lesley Fellows (Fellows and Farah, 2003, 2005), which provides a good case for reversal learning being the cause of VMF patients IGT deficit, and their wider real life behavioural problems. Fellows and Farah (2003) tested VMF patients on a simple reversal learning task, where participants could select from one of two decks. One deck provided the positive stimulus (S^+), which gives \$50, the other was the negative stimulus (S^-), which when selected, lost the partcipant \$-50. Initially, S^+ is allocated randomly to a deck, and the other deck becomes S^-. Learning of the stimuli is considered to have occurred when 8 consecutive selections (learning criterion) of the S^+ have been made. The number of mistakes/errors (an error is when S^- is selected) made before this initial learning are called the 'learning errors'. After this initial learning has taken place then a reversal of the stimuli-reinforcement association occurs, in that, the deck which was S^+ becomes S^-, and vice versa for the other deck. The contingencies remain the same until the learning criterion is achieved again (S^+ has been selected 8 times in a row, without mistake), causing another reversal. The task continues until 50 card selections have been made. This allows for a maximum of 5 reversals.

The results of the study (Fellows and Farah, 2003), found neither patients with dorsolateral prefrontal (DLPFC) or VMF lesions were significantly affected in their initial stimulus-reinforcement learning. However, ventromedial prefrontal damage selectively impaired reversal learning. VMF patients made twice as many errors as normal healthy controls (NHCs) during the reversal phase of the task.

Additionally, Fellows and Farah found reversal learning deficits were negatively correlated with a measure of day-to-day behavioural deficits, described using the Instrumental Activities of Daily Living (IADL) scale (Gallo et al., 2000), where scores range from 7 (dependent in all domains) to 21 (fully independent). However, her study shows that this reversal learning deficit may be restricted to lesions in the medial area of the orbitofrontal cortex. The patients with the severest reversal learning deficits have a mean focus of lesions in the ventromedial prefrontal area of their left hemispheres. Whereas, Bechara et al.'s VMF patients had a slightly greater focus of damage in the right hemisphere; both groups' VMF lesions are fairly bilateral, but the damage is probably more extensive in Bechara et al.'s VMF patients. It must be noted that this is only a qualitative analysis of the lesion overlaps found in the literature. As a general note, patients with VMF lesions often have damage that may also involve the basal forebrain and the genu (knee-like bend or curved part) of the corpus callosum, Fellows (2005).

3.2.3 Comparing Fellows and Farah (2005)'s VMF patients and Bechara et al.'s VMF patients on the IGT

To further elucidate whether VMF patients' deficits in the IGT are caused by a more fundamental reversal learning problem, Fellows and Farah (2005) tested 9 VMF patients on the original ABCD version of the IGT and on a 'shuffled' version of the task. The 'shuffled' version of the task removes the reversal in contingencies found in the original 'unshuffled' ABCD variant of the task, by rearranging the original score card (see Bechara et al. (2000a)). In the 'shuffled' variant (Fellows and Farah, 2005), cards 1-8 from all decks (A, B, C and D) are placed at the bottom of their respective decks. Then, additionally, in deck B, cards originally numbered 11 and 14 are swapped. From figure 1, it can be seen that this creates two big losses ($-1250), within the first three cards of deck B. This compares to there originally being one of these large losses, the largest possible in the original ABCD version of the IGT, every 10 cards. This makes both deck A and especially deck B seem disadvantageous from the begining of the 'shuffled' variant of the task. Therefore, there is no reversal, but the patients would still have to manage uncertain results against regular results. It also allowed the re-testing of the dissociation of ability to learn the initial stimulus-reinforcement, compared with reversal learning, in VMF patients (Fellows and Farah, 2003)

Before we can fully consider Fellows and Farah (2005)'s data, it is important to note some differences in performance and lesion location between her patients, and those found in the Bechara et al. studies (Bechara, 2004; Bechara et al., 2000b, 1999). When comparing the VMF profiles in Fellows and Farah (2005) vs. the amalgamation of data found in table 3, the data points are found to be significantly different in blocks 1 and 5 (block 1, t = 2.27, $P < 0.05$; block 5, t =2.36 P = 0.02), on the 'unshuffled' original ABCD version. A test of overall significance, comparing total net scores between the two sets of data, shows they are significantly different (t = 2.08, $P < 0.05$). Also, possibly most importantly, Fellows' VMF patients seem to improve at the task over time, although this is mainly a qualitative assessment, as only block 1 is significantly different from a random binomial series, and the overall increase in net score is small (4 points in 5 blocks). When Fellows and Farah's 'unshuffled' ABCD IGT VMF profiles are compared to those of Bechara et al.'s normals and VMF patients, then we notice that Fellows and Farah's VMF patients do not learn the initial stimulus-reinforcement associations, as well. This is not what one would expect from the results of the simple reversal learning task (Fellows and Farah, 2003). This is also the case with Fellows and Farah (2005)'s normals, who pick more from the advantageous decks in block 1 of the task, for some of which decks, A and B, are actually the best decks (Maia and McClelland, 2004). In Bechara et al.'s studies, normals (also normals in Clark et al. (2003), net score = -6, in block 1) and VMF patients, select more from decks A and B, than C and D in block 1 (see table 3, VMF net score = -4.7, in block 1), which is actually the best strategy for the early selections, until the reversals. Therefore, it seems that the initial stimulus-reinforcement learning ability shown in the simple reversal learning task (Fellows and Farah, 2003), does not fully follow into the IGT for Fellows and Farah's VMF patients, but is present in Bechara et al.'s VMF patients. Although, if all these results are based on normative comparisons, then Fellows and Farah's VMF patients do act more similarly to their normative group of NHCs (Normal Healthy Controls), than to the normative groups found in Bechara (2004); Bechara et al. (2000b, 1999); Clark et al. (2003). But in block 1, Fellows and Farah's normals are significantly different from the normals taken in the Clark et al. (2003) and Bechara et al.'s studies ($P < 0.001$, $t = 3.92$). Fellows and Farah's VMF patients act randomly, like they have 'no preferences', a deficit present in Sloman (2004)'s description of VMF patient's problems, and similar to Sacks (1998)'s case study description of a patient with severe bilateral damage to the orbitofrontal cortex, who had 'no preferences'.

When examining the MRI data, differences lie in the main overlap of lesions across patients in each group. Bechara et al.'s VMF patients seem to have a greater overlap in the right VMF compared to a greater focus in the left VMF for Fellows and Farah's patients, although this cannot be stated conclusively from the literature, as Bechara et al.'s papers only distinguish a maximum overlap as 4 or more patients. This is unlike Fellows who distinguishes between a complete scale of 1 - 10 possible patient overlaps, which covers the 7 VMF patients with the worst reversal learning deficits (errors ≥ 9).

In imaging studies (Kringelbach et al., 2003; O'Doherty et al., 2000), it has been found that activity in regions of the caudal OFC reflect current reward value of either a taste or odour, rather than the unvarying features of the stimulus. Therefore, if this region is damaged one might expect a reduction in preference.

Performance of Fellows and Farah's VMF patients in the 'unshuffled' (original) ABCD IGT, could be seen as random ('no preference'), or slow improvement. This differs from their performance on the 'shuffled' ABCD IGT, as their performance is not significantly different from normal healthy controls (Fellows and Farah, 2005). Again, as in the 'unshuffled' ABCD version, Fellows' VMF patients only slightly improve their net scores over the five blocks in the 'shuffled' version. Although an improvement in net score is less warranted in the 'shuffled version', as the contingencies change less. Therefore, a possible description of Fellows and Farah's VMF patients' performance could be that they, more or less, stick to their initial stimulus-reinforcement assessment and change it very little, rather than having 'no preferences'. Overall, Fellows and Farah (2005) do seem to provide strong evidence for reversal learning being the underlying deficit in VMF patients poor 'unshuffled' ABCD IGT performance. Fellows and Farah's VMF patients have a mean selection of advantageous cards in the high 60s and the normal controls select on average in the low 70s from decks C and D in the ‘shuffled’ version of the IGT. However, as mentioned before, performance in the IGT differs both qualitatively and statistically, between Bechara et al.’s VMF patients and those in Fellows and Farah’s studies. This could be due to fundamental differences in the patients’ deficits or simply a sampling problem. It has been noted that researchers often believe, that the probability of successfully repeating experimental results, with different samples, especially small ones, is relatively high (Kahneman et al., 1982). This has been described as the ‘belief in the law of small numbers’ (Kahneman et al., 1982), and is the strongly held belief found among both lay people and experienced researchers, that even small samples will be representative of the populations from which they come. This is a statistically weak belief, but is sometimes unavoidable in specific patient groups, where injuries or diseases are rare.

3.2.4 Lateral differences in lesion location as a cause of differences in VMF behaviour between study groups (Fellows and Farah vs. Bechara et al.) on the IGT

The lateral effects of damage to the ventromedial prefrontal cortex have been investigated. In a small study (n = 7) (Tranel et al., 2002), patients with unilateral VMF lesions (left n = 3, right n = 4) were tested on the IGT. It was found that those with predominantly right hemisphere VMF damage performed more like those with bilateral VMF damage. Whereas, the patients with predominantly left hemisphere VMF damage, performed more like normal healthy controls. Deficits in IGT performance, laterality of VMF damage, emotional processing and personality, and social conduct were all correlated. The patients with right hemisphere VMF lesions were not gainfully employed and scored the maximum score 3 (worst) in most dimensions of social conduct and met the criteria for ‘acquired sociopathy’, using the ‘Iowa Scales’ (Barrash et al., 2000). (A measure of changes to emotional functioning and personality that occur following brain injury, with specificity to VMF lesions.) The ‘Iowa Scales’ “... identify a core set of disturbances: (1) general dampening of emotional experience (impoverished emotional expressiveness, apathy, and inappropriate affect); (2) poorly modulated emotional reactions (poor frustration tolerance, irritability, lability); (3) disturbances in decision making, especially in the social realm (indecisiveness, poor judgement, inflexibility, social inappropriateness and insensitivity or lack of empathy); (4) disturbances in goal-directed behaviour (problems in planning, initiation, persistence, and behavioural rigidity or non-spontaneity); (5) a striking lack of insight into these personality changes” (Tranel et al., 2002).

These changes are judged by people who have known the patients for some time prior and post brain injury.

The patients with predominantly left sided damage scored 1 (scores run from 1 to 3, 1 = no change or impairment, 2 = moderate change or impairment, 3 = severe change or impairment, since brain injury) for each of the aspects of social conduct tested, all were in gainful employment, were normal in emotional processing and personality as tested by the 'Iowa Scales' and only had a small deficit on the IGT compared to normals in blocks 4 and 5. Social conduct was tested in the following four domains (Tranel et al., 2002): social status, which was defined as an alteration in the patient's financial security and/or peers' judgement of social attainment; employment status, i.e. alteration in patients' occupation; interpersonal functioning i.e. normal social relationships with significant others, such as friends and family and clinical rating, which were from multiple sources, e.g. interview data, neuropsychological test performance, etc. In contrast, those with predominantly right hemisphere damage averaged 3 for all of the social conduct domains, performed disadvantageously on the IGT, had 'acquired sociopathy' and, as previously mentioned, were not gainfully employed.

Although this is a very small study, it is somewhat supported by a much larger study, which was less lesion specific, with these frontal patients only showing limited damage to the medial prefrontal and orbitofrontal cortex (Clark et al., 2003). Clark et al. (2003) found that, across all blocks, frontal patients with right frontal lesions chose more from the disadvantageous decks, A and B, than decks C and D. However, left frontal patients selected slightly more from the advantageous decks, C and D. Both profiles are similar to results for left and right VMF lesions in the Tranel et al. (2002) study. However, lateral effects were not present when the Clark et al. (2003) patients were tested on the Cambridge Gamble Task and the Risk Task which are tests that breakdown some of the multiple aspects of the IGT.

Interestingly, the Tranel et al. (2002) data for patients with left hemisphere VMF lesions on the IGT is somewhat similar to those for bilateral VMF patients in Fellows and Farah (2005) and Clark et al. (2003)'s left frontal patients. Initial stimulus-reinforcement learning in the IGT is well preserved in Tranel et al. (2002) study (net score = -5, in block 1), but less so in the Fellows and Farah (2005) (net score = -2, in block 1) and Clark et al. (2003) (net score = -2, in block 1) studies. In that, as suggested by Maia and McClelland (2004) and Fellows and Farah (2005), in the initial part of the original ABCD IGT, from what is known, the best cards to select are A and B and not C and D. This can be seen in figure 5. When only a few cards have been selected from each deck (probably block 1) then decks, A and B, are the most rational choice. It only remains advantageous (in terms of mean result) to select from deck A for the first 3 cards, then becomes disadvantageous for the 4th selection and advantageous again for selection 5 in that deck, but after that, it is no longer better than either decks C or D. However, for deck B, being the best deck (in terms of mean result) continues until the 9th selection from that deck, after which it is never on average best to select from deck B again (see figure 5). Therefore, both VMF patients in Fellows and Farah (2005) (Block 1 net score = -2) and those with left hemispheric VMF damage in Tranel et al. (2002) (block 1 net score = -5), both begin (block 1) with a negative net score, which is initially rational, but steadily improve their performance across the proceeding blocks, ending up with a positive net score (Tranel et al. study, block 5 net score = 4, and in Fellows and Farah study, block 5 net score = 2). Both left unilateral VMF and bilateral VMF patients, although much slower than normal controls, improve their performance.

This is unlike the profiles of the bilateral VMF patients shown in Bechara et al.'s papers, who start by learning the initial stimulus-reinforcement normatively (net score block 1 = -4.7), then improve until block 3 (net score = 2), but then decline in performance, ending block 5 with a negative net score of -4.2 (see table 5). This is one of the most interesting points in Bechara et al.'s VMF data, after seemingly having learned that decks A and B are not the best decks (block 3), bilateral VMF patients (with most lesion overlap in the right hemisphere) reverse their behaviour, without another change in the contingencies. (The reversal in contingencies is very likely to have occurred by block 3, card selection 40-60.) Although, they do not seem to have an obvious short-term problem with rapid reversal, which has been described as an important function of the OFC by Edmund T. Rolls, in that, Bechara et al.'s VMF patients rapidly stop selecting from a deck if they get a large loss, but return to these decks more rapidly than normal healthy controls (Bechara et al., 2005).

This post block 3 reversal in net scores (net scores become negative) does not occur in the VMF patients in Fellows and Farah (2005) (greater overlap in the left hemisphere) or for the left VMF patients in Tranel et al. (2002). They slowly improve, or at least have a tendency to follow their initial stimulus-reinforcement associations throughout the task, rather than improve and then get worse, as Bechara et al.'s VMF patients do.

Therefore, the left hemisphere VMF lesion patients from Tranel et al. (2002) and Fellows and Farah

(2005)'s bilateral VMF patients seem to be most similar in overlapping focus of lesion location and IGT performance. However, it is not so clear how personal and social behaviour correspond across these groups. None of Tranel et al. (2002)'s left VMF patients had registered social or emotional deficits based on the measures used in that paper, and neither did Fellows and Farah (2005)'s VMF patients as a whole group. There is a strong correlation though with Instrumental Activities of Daily Living (IADL) and deficits in reversal learning in Fellows and Farah (2003)'s VMF group (Spearman $\rho = 0.84$, $P = 0.0003$). Additionally, there is a correlation between improvement in performance on the 'shuffled' ABCD IGT compared to the original 'unshuffled' ABCD variant, and impairment in reversal learning ($r = 0.53$, $P < 0.05$).

This is good evidence that the removal of the reversal in the IGT improves performance mostly for people with a reversal deficit, and that the reversal in the IGT is possibly their major cause of detrimental decision making on the original ABCD IGT. The next empirical test to perform would be to test Bechara et al.'s population on Fellows and Farah (2003)'s reversal learning task and the 'shuffled' version of the IGT, to compare results. Fellows and Farah (2005) do acknowledge differences between their VMF study group and those studied by Bechara et al.. Firstly, Fellows and Farah's VMF patients have less extensive lesions and have only been selected on lesion location, compared to Bechara's VMF patients, who's selection was based on lesion location and impaired real life decision making. Reversal learning in a simplistic sense, as mentioned above, is not Bechara et al.'s VMF patients' problem with the IGT, as they avoid punishment in the short term and do not perseverate more than normals, but return more rapidly than normals to those punishing choices (Bechara et al., 2005). This is also the case with Fellows and Farah's VMF patients who also do not perseverate on the large loss in deck B, as 8 out of 9 patients stop selecting from that deck after the $1250 loss. They also return to deck B faster than NHCs (Normal Healthy Controls), where the mean number of selections before returning to deck B after the card 9 loss are 10 and 4 respectively. It should also be noted, overall, Fellows and Farah's VMF patients' profiles on the 'unshuffled' ABCD IGT only just fulfil the measure of disadvantageous behaviour with an overall net score of -1. However, Bechara et al.'s VMF patients' average an overall net score of -9.2.

In Fellows and Farah (2003)'s study, the VMF patients had reversal learning deficits, but not an initial stimulus-reinforcement learning impairment. This is in line with other results which have shown a dissociation between reversal learning and initial acquisition of a visual discrimination task. Patients with temporal lesions have impairments in initial learning and perform as well as controls on reversals, whereas frontal patients suffer from a reversal, and not an initial association, learning deficit (Daum et al., 1991). The lesion location of this reversal deficit has been further linked to the ventral prefrontal cortex. Rolls et al. (1994) found that patients with frontal damage including ventral areas were more impaired on reversals and extinction than those frontal patients without damage in the ventral prefrontal cortex. In the same paper, Rolls et al. found a correlation between socially inappropriate behaviour and reversal learning deficits.

It is of interest to note that a double dissociation between reversal learning and higher-level attentional set shifting has been shown in marmosets using the Intra-dimensional/Extra-dimensional (ID/ED) shift test (Dias et al., 1996), a task designed to breakdown the component elements of the Wisconsin Card Sorting Task (WCST) (Downes et al., 1989). This double dissociation is also supported by evidence from withdrawal of L-DOPA medication from Parkinson's suffers. It has been found that withdrawal of medication significantly impaired performance on a task of attentional set-shifting, but ameliorated deficits in a probabilistic reversal learning task (Cools et al., 2001). Therefore, the finding that Bechara et al.'s VMF patients are not significantly worse than healthy controls on the WCST, does not prove they do not have a more basic reversal learning deficit.

Evidence from single cell recordings in animals and functional neuro-imaging in humans implicates ventral fronto-striatal circuitry in reversal learning (Clark et al., 2004). In a PET study on an analogue of the ID/ED shift task, with separate blocks for reversal learning and attentional set-shifts, significant changes were found in DLPFC during attentional set-shifting, and in the ventral caudate nucleus during reversal learning (Rogers et al., 2000)

To further investigate the role of the VMF in reversal learning, there have been a couple of studies examining probabilistic reversal learning paradigms, an event-related fMRI study on normal healthy individuals (Cools et al., 2002) and an orbitofrontal and DLPFC lesion study (Hornak et al., 2004). In Cools et al. (2002)'s study, significant activation changes in the right ventrolateral prefrontal cortex (VLPFC) and the ventral striatum (VS) were found, when comparing the final reversal error, the incorrect choice before a participant adjusts to the reversal and selects correctly again, and correct responses. There was also significant signal change in the right VLPFC between the final reversal error, preceding errors and probabilistic errors (those errors when a participant is told they have responded incorrectly, when they have selected correctly, i.e. it is not a real reversal, but an event used to increase the task complexity and introduce errors which are to be ignored). From these results it has also been suggested that signal change in right VLPFC on the final reversal error is not dependent on the number of preceding errors. This suggests that the signal change is due to a realization that the errors signify a reversal, not just a probabilistic error.

citetCools2002's suggests that "... shifting of lower-level stimulus-reward associations, as opposed to shifting of a higher-level attentional set, is sufficient to activate the VLPFC [ventrolateral prefrontal cortex]." A further probabilistic reversal learning study, using a paradigm which had indicated significant use of the OFC (O'Doherty et al., 2001) has been tested on patients with a variety of surgically circumscribed prefrontal lesions. The study by Hornak et al. (2004) found patients with surgically circumscribed bilateral OFC lesions were severely impaired on the task, in that they accumulated less money than controls. The bilateral OFC patients also perseverated (they often failed to switch choices after a large loss), and, in contrast, often switched choices although they had just received a reward. The study found that unilateral OFC (or medial prefrontal cortex) lesions did not cause impairments, neither did DLPFC patients when they paid attention to all the available feedback provided in this reversal learning paradigm. Again, Hornak et al. (2004) found that initial stimulus-association was not affected by any of the lesions presented in this study, providing more evidence for the dissociation between initial stimulus-association learning and reversal learning.

3.3 Risk-seeking (Sanfey et al., 2003), Cambridge Gamble Task (Rogers et al., 1999)

It has been commented that the IGT has too many components to easily understand the deficit(s) present in poor compared to good performance, where good performance is defined by normative results. Therefore, at least two groups have attempted to further break down the IGT into simpler tests, which test more definable abilities. The two considered here are Rogers et al. (1999) and Sanfey et al. (2003).

In Rogers et al. (1999)'s study, which introduces the Cambridge Gamble Task (which requires no online learning, just probabilistic judgements and tests impulsivity), OFC lesion patients did not, as often as normals and DLPFC patients, make the probabilistically most rational gamble. This suggests risk-seeking in patients with OFC lesions, but in contrast, OFCs and DLPFCs bet at a reduced rate compared to normal controls. Also, unlike VMF performance on the IGT, Rogers et al. (1999)'s OFC patients improved their decision quality on their second attempt of the Cambridge Gamble Task, in that, they picked more often from the most likely gamble when they attempted the task a second time (81% first attempt vs. 89% second attempt).

Sanfey et al. (2003) devised a test in between the Cambridge Gamble Task and the IGT. In this test there is no pre-set reversal, but on-line learning is required in the task, each trial is independent, and the main aim of the task is to test attitudes towards risk (Sanfey et al. define risk as variance in returns over time). In the Sanfey et al. (2003) task, there were five deck distributions (deck 1 = lowest risk, to deck 5 = highest risk), but on each turn only two decks were presented and picked from. This simplifies the task in terms of less multi-tasking, a deficit evident in frontal patients (Shallice and Burgess, 1991) on the Multiple Errands Test (Alderman et al., 2003). (Multi-tasking in this context would be to compare and update more competing valuation functions, see models later.) The participant only had to select from the pair on display, and then a result was shown from the 25 possible outcomes for each deck, taken from their distributions (Sanfey et al., 2003). The 25 possible results were randomly cycled through, so that all other possible outcomes were shown before a value could be repeated.

The two decks only differed in the variance and probability of potential results, ranging from the NoLoss deck, where participants could not lose a point (but only won a maximum of 20 points), to the Long-Shot deck, where there was a small probability of winning 150 points, but also a reasonable chance of losing (up to 50 points). Overall, each deck had the same pay off over many trials. All decks had identical expected values, in that they all on average gave 10 points per card. An example of this kind of equivalence is: win 60 points (with 0.25 chance) or lose 20 (with 0.75 chance), giving an average pay off of (60 x 0.25) + (0.75 x -20) = 15 - 15 = 0; or, win 150 points (with 0.2 chance) or lose 37.5 (with 0.8 chance), giving an average pay off of (150 x 0.2) + (0.8 x -37.5) = 30 - 30 = 0. However, in a single trial, the second bet would be considered to be the most ‘risky’ in terms of potential variance in results. In their initial analysis, Sanfey et al. (2003) found VMF patients (n=9) had no overall preference for any deck and selected from each deck 0.5 of the time. This would support Aaron Sloman’s ‘no preference’ theory, and is supported if the Fellows and Farah’s VMF patients’ ‘unshuffled’ decks profiles are considered alone. Sanfey et al. (2003)’s VMF patients seemed to act randomly, whereas normals showed a preference for lower ‘risk’ decks, with a negative correlation between selection preference and the ‘riskiness’ of the deck. However, on further analysis, two sub-groups among the VMF patients were found. One subgroup selected more from the risky decks, risky-VMF (n = 5) and another group, safe-VMF (n = 4), had a preference for low risk decks (decks 1 and 2) and a dislike of the high risk decks (decks 4 and 5). All three other groups, age matched healthy controls (n = 17), psychology students (n = 63) and frontal patients with lesions not including the VMF (n = 4), acted similarly in their preferences to the different decks.

Sanfey et al. (2003) have suggested that the behaviour in Bechara et al.'s VMF patients on the IGT is due to a preference for risk. This does not seem to be the case when considering Bechara et al.'s VMF patients across studies, because, in the EFGH variant of the IGT, VMF patients select more or less randomly (Bechara et al., 2000a), with a net score equal to 1. One would expect the net score to be higher as the best decks are also the 'riskiest'. Results from the E'F'G'H' variant (Bechara et al., 2000a, 2002b) provides further evidence against a 'risk-seeking' only hypothesis, as VMF patients have a preference for disadvantageous decks, which have lower variances than the advantageous decks. The behaviour of these patients on the IGT is more subtle than just risk-seeking. However, the suggestion that "... different mixtures of sub-types of VMPFC [VMF] patients were sampled in different studies" (Sanfey et al., 2003) is possible.

The two sub-groups defined by Sanfey et al. (risky-VMF and safe-VMF), have different behaviour profiles, but not obvious lesion location or volume differences in their VMF damage. However, the risky-VMF group did seem to have more damage to areas in the DLPFC, particularly the left DLPFC, compared to the safe-VMF group. However, contrary to this frontal patients in Sanfey et al. (2003) with lesions not including the VMF, but damage in the left DLPFC (left Brodmann area 46, (BA46)), do not show risky behaviour. This is suggestive of an interaction between VMF-DLPFC in normal 'non-risky' decision making, and is supported by Manes et al. (2002). In Sanfey et al. (2003)'s study they also found apparently more bilateral damage to the frontal pole in the risky-VMF group, however one member of the safe-VMF group also had damage in this region, making the evidence inconclusive. Another point to mention from Sanfey et al. (2003)'s study is that some members of control groups were also risk orientated, suggesting that this could be a phenomena influenced by the distribution of heterogenous risk-attitudes in the general healthy population.

3.4 Myopia for future consequences - Bechara

Bechara et al. (2000a) have described the decision making behaviour of their patients, with VMF damage, as displaying a ’myopia’ for future consequences. This is a view which the group have kept over the past few years. They have found that their VMF patients act in the short term and often make decisions which are detrimental in the long run.

4 ϵ-Greedy Action-Value Method

In this paper, we have tried to simulate the human and patient data from Bechara et al. (1999, 2000a); Clark et al. (2003); Bechara and Damasio (2002a); Bechara et al. (2002b), using basic reinforcement learning algorithms, based on ϵ-greedy action-value methods (See p27, 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 exploitive, and will select the choice with the highest estimated value. If $\epsilon = 1$, then every trial is explorative, where in this set-up, each action is equiprobable. Whether to explore or exploit is an important issue in reinforcement learning. Also, importantly, the methods for how to explore the action space are varied and can require consideration in their own right. Methods for exploration include random searches, explore the second best option, e.t.c.. This extends to methods for exploitation which are usually based on predicting which action leads to the most valued result. Prediction methods might attempt to answer questions like, which previous action among a sequence of actions should be causally linked, with a temporally separated reward?

In the simulations considered here, either one of two simple methods are used to estimate the value of an action. If we use the original ABCD variant of the IGT as an example to explain the methods used in the simulation, then the possible actions are a selection of a card from one of the four decks, A,B,C or D. The value of an action is the mean result gained when that action has been selected. This method for valuing a possible action is called *sample-averaging* and is simply the mean result gained from executing a particular action. Therefore, through sufficient exploration, and the law of large numbers, this method, for problems like the IGT, should achieve a level of optimal behaviour, minus the cost of exploration. This means that, if the contingencies of the game do not change and the exploration rate is low, over many trials, this method will become asymptotic to an optimal policy of actions for each state.

The other valuation method considered uses an *error-driven* valuation function and is explained in more detail in section 4.2.

If the exploration rate is high then the method should rapidly find its optimal strategy (i.e. arrive at the correct valuation of the choices/options), as it would more often investigate the current non-optimal strategy. Therefore, it will more rapidly discover the contingencies across all actions. The method would also be more able to manage on-going changes, in the reward/punishment schedules of the task. However, if the contingencies are slow to change, or do not change, then a high exploration rate will generally be less optimal than a low exploration rate. This is because the chance of acting optimally on an exploration trial is (no. of optimal choices) / (no. of choices available).

Using the standard notation in Sutton and Barto (1998), we can set out the method mathematically. We denote the true value of an action *a* as $Q^*(a)$. This is the value that our method will approach over time and will guide the method towards the optimal strategy for the task/problem. The estimated value of action *a*, which is used by the simulator to decide on the best action among the set of possible actions, is, on the *t*th play, denoted by $Q_t(a)$. Using this notation to model behaviour in the IGT, we can assume that every time step *t* is equivalent to each trial in the task. On each trial the player of the IGT has four possible choices/potential actions (unless decks have been exhausted). Players can select a card from each of the decks, A,B,C or D. Each deck will have an estimation of the expected value gained from selecting that particular deck. Taking the first selection, we have four action values:-

$Q_1(A)$ = 0,
$Q_1(B)$ = 0,
$Q_1(C)$ = 0
and $Q_1(D)$ = 0.

where, when *a* = A, the action is to select a card from deck A and, if *a* = B, then that is equivalent to picking from deck B and so on. All the estimations would begin at zero, unless there is a wish to force an initial exploration of the choice space, by setting the estimations very high, so, until new information specifies otherwise, untested actions have a high valence and so will be picked early in the task. In the simulations shown later in this paper, each action has an initial value of 0, which is overcome by the relatively high levels of exploration used to reproduce the human data. The very first selection on the task is always random.

If we continue our example at the beginning of *t* = 2, and deck A was selected first, at *t* = 1, then as the player would have won $100 with no loss (see figure 1), this would give the following updated estimations:-

$Q_2(A)$ = 100,
$Q_2(B)$ = 0,
$Q_2(C)$ = 0
and $Q_2(D)$ = 0.

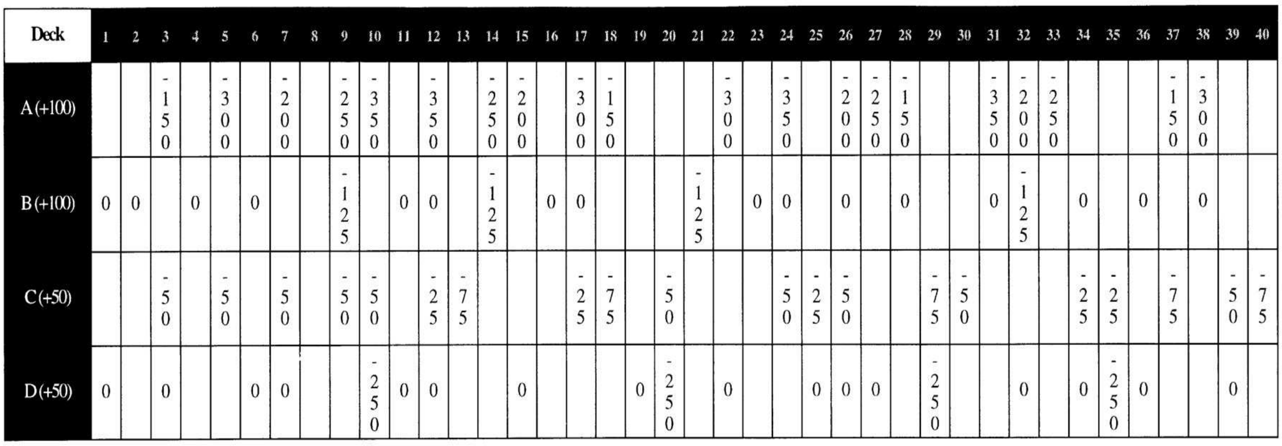
This would make deck A the next choice as it has the highest estimated value, unless the next trial is an exploration trial. Firstly, to decide whether it is an exploration trial, a random number is generated between 0 and 1 and, if it is less than or equal to the exploration rate, an exploration trial is undertaken, and each of the four choices will have an equal chance of being selected.

To reiterate, the true value of an action is the mean reward received when that action is selected. Therefore, by averaging across trials when a particular deck is selected, we will eventually achieve this. If, on the *t*th play, action *a* has been chosen k_a times prior to *t*, yielding rewards r_1,r_2,...,r_{k_a}, then its value is estimated using the equation in figure 2.

So, for k_a = 0, as stated above, we can set $Q_t(a)$ to some default value, such as $Q_0(a)$ = 0. Again, as $k_a \to \infty$, by the law of large numbers, $Q_t(a)$ converges to $Q^*(a)$.

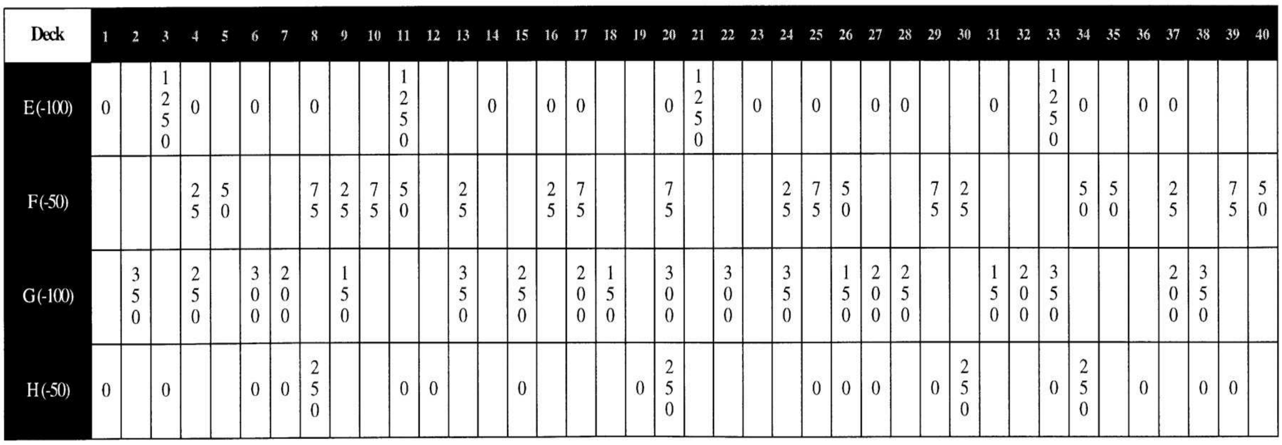
We use the simplest action selection rule (greedy, as described by the method name) on exploitation trials, which is to select the action (or one of the actions) with the highest estimated action value, that is, to select on play *t*, one of the greedy actions a^*, for which $Q_t(a^*) = max_a Q_t(a)$ = max$\{Q_t(a) \mid$ *a* $\in \{A, B, C, D\}\}$. So, if exploration is low, then it may take longer to learn the contingencies, but it will be better at exploiting them once they are learnt. Therefore, by increasing the exploration, we get a more rapid view of the value of all the possible actions. Without some level of exploration, we only get the greedy exploitation, which can be ineffective, if the contingencies are not known. This kind of over exploitation has been a problem for our previous models, where the exploration was not sufficient

Original Task (ABCD)



Deck	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
A(+100)			-150		-300		-200		-250	-350		-350		-250	-200		-300	-150				-300		-350		-200	-250	-150			-350	-200	-250				-150	-300		
B(+100)	0	0		0		0			-125		0	0		-125		0	0				-125		0	0		0		0			0	-125		0		0		0		
C(+50)			-50		-50		-50		-50	-50		-25	-75				-25	-75		-50				-50	-25	-50			-75	-50				-25	-25		-75		-50	-75
D(+50)	0		0			0	0			-250	0	0			0				0	-250		0			0	0	0		-250			0		0	-250	0			0	

Variant Task (EFGH)



Deck	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
E(-100)	0		1250	0		0		0			1250			0		0	0			0	1250		0		0		0	0			0		1250	0		0	0			
F(-50)				25	50			75	25	75	50		25			25	75			75				25	75	50			75	25				50	50		25		75	50
G(-100)		350		250		300	200		150				350		250		200	150		300		300		350		150	200	250			150	200	350				200	350		
H(-50)	0		0			0	0	250			0	0			0				0	250					0	0	0		0	250			0	250		0		0	0	

Figure 1: **Score-cards from the original and variant versions of the gambling task (Bechara et al., 2000a). This paper particularly explores the original ABCD task. Please note that $125 losses for deck B is a misprint and should actually be, -$1250 as stated in the article.**

$$Q_t(a) = \frac{r_1 + r_2 + ... + r_{k_a}}{k_a}$$

Figure 2: **Estimated action-value equation for the ϵ-greedy method using *sample-averaging***

to override highly valued options/actions in the E(')G(')F(')H(') versions of the IGT (Kalidindi et al., 2005b,a).

4.1 *Sample-averaging* valuation functions

The ϵ-*greedy* method using *sample-averaging* converges to a probability greater than 1-ϵ of selecting the optimal action as $k_a \rightarrow \infty$. This is because the chance of selecting the optimal action is 1-ϵ, the probability of exploitation during a trial, plus the chance of selecting the optimal action during an exploration trial. In the IGT, for the ABCD version, picking from C or D is optimal, therefore optimal action occurs 50% of the time during an exploration trial. So using these methods in the IGT, the probability of taking the optimal action, which is when $Q_t(a)$ converges to $Q^*(a)$, is:-

P(optimal) = $\frac{\epsilon}{2}$ + 1-ϵ
= 1 - $\frac{\epsilon}{2}$

Therefore, using an exploration rate of 0.58, the potential probability of optimal action, given that none of the decks have been exhausted, is 1-0.58/2 = 0.71. This means that, after sufficient trials, on average, 71% of the time, an advantageous card will be selected.

In the following simulations, it was found that an exploration rate of 0.58 gave a good match to behaviour in normal human subjects across the different versions of the IGT. This exploration rate does not provide the best match for each individual version of the IGT, which may be due to the normal control groups being different each time the experiment was run. Alternatively, this could be due to the variation in the contingencies in the four versions, i.e. people adjust their exploration due to the 'difficulty' of the task. This is difficult to glean because normal subjects have too high a learning effect. Once they have participated in any version of the task they learn other versions of the task very quickly. Therefore, it is impossible to use the same normal human test group again, and so there will inherently be group differences. Also, the results within study are not very robust, as can be seen from the large SEMs found in all studies, suggesting that group differences might be reasonably large, especially with the sample sizes used.

This data lends itself to the conclusion that normal humans fit in a continuum of ability to exploit the IGT. Based on the work set out so far, a way to consider this continuum might be as a range of exploitation rates. This range could be quite broad, and it may be quite arbitrary, as from the literature it is clear that even persons considered normal and healthy can fall into many sub-categories based on their performance, and sometimes other criteria. Evans et al. (2004) have found that when they tested university educated students vs. subjects educated till 16 years on the ABCD version of that IGT, there were significant differences between the groups in blocks 4 (61-80 cards) and 5 (81-100 cards). Interestingly, those with less formal education performed superiorly, suggesting education as an influencing factor on performance. Other studies have sub-grouped normals by performance (Crone et al., 2004; Bechara et al., 2002b). As a corollary to group differences, some have commented on the variation in individuals (Suzuki et al., 2003). Therefore, where this range lies would depend on the criteria used to judge normal status. Bechara and Damasio (2002a) have described advantageous behaviour as getting an overall net score greater than 0 across all 100 trials.

4.2 *Error-driven* valuation function

In the model set out above, there are a few areas where different parts of the method can be modified. These include how a choice is selected on an exploration trial (currently randomly), and how information from trials is used to calculate the expected result of a choice (so far, sample-averaging has been considered). Another possible way to calculate the estimated value of a choice, is by *error-driven* learning. This is a method heavily used in supervised learning in neural networks, with back propagation being an example (**?**).

The basis of error-driven learning in this context is that an estimation is made for each potential choice from a particular state. The estimations at the beginning of the task can be set to zero for each choice or based on predefined knowledge about the task. In the *error-driven* ϵ-greedy method, set out in this paper, the initial values for all choices are set to zero. Also, in this case, there is generally only one state, which does not even change on the last trial, as the participant does not know that there are 100 trials (See IGT participant instructions in the methods section Bechara et al. (1999)). Therefore, in the normal state there are four possible choices.

$$Q_t(a) = Q_{t-1}(a) + \gamma(R_t(a) - Q_{t-1}(a))$$

Figure 3: **Error-driven valuation function. Where γ denotes the learning rate and $R_t(a)$ denotes the reward at time t, for action a.**

Overall, *error-driven* ϵ-greedy differs from *sample-averaging* ϵ-greedy only in the valuation function. The evaluation function for error-driven learning can be found in figure 3. The main difference between *sample-averaging* and *error-driven* valuation functions is that *error-driven* learning has the full Markov property. The full Markov property requires only information about the current state, without the necessity for knowledge about how many times a certain choice has been made, as used in the *sample-averaging* valuation function (i.e. the value of k_a, see figure 2). *Sample-averaging* creates a partial Markov decision process in that the number of times a choice has been made is required rather than the exact path. Additionally, in *sample-averaging* the influence of new information is dependent on the amount of information previously presented for that choice. Therefore, if new information comes after many iterations, even if the information is very different from previous estimations, its influence will be limited. However, with the error-driven valuation function, the number of times a particular selection is made is irrelevant to the way the valuation of that choice is updated. This means, dependent on the value of γ, error-driven learning can still be responsive to individual changes in contingencies, especially later in a long task. In addition, whereas *error-driven* valuation functions will continue to value new information the same throughout a task, if γ is constant, *sample-averaging* will value early information more than later data. Therefore, *sample-averaging* is more influenced by initial stimulus-reinforcement associations.

For example, this means a reversal task using *sample-averaging* would result in progressively more perseveration at each reversal as the task continued. However, it would probably be rapid at making the initial stimulus-reward associations. In contrast, an *error-driven* valuation function with a high learning rate would learn the initial stimulus-reward associations at a similar rate to any reversals (i.e. with limited perseveration). Therefore, ϵ*-greedy* methods using *error-driven* would not get worse at the task.

In the case of a probabilistic reversal task (Cools et al., 2002), *sample-averaging* would be less influenced by the false errors (when the participant is told they have made an error when their response is actually correct, false errors are presented to increase task complexity), than a high learning rate *error-driven* ϵ-greedy set-up. However, with a low learning rate, an *error-driven* valuation would be less influenced by singular false errors and correctly adjust to genuine reversals, though, with some perseveration. The level of perseveration would not however grow as with *sample-averaging*.

4.3 Problems with Non-Linear Neural Decision Network (N-LNDN)

Ideas about valuations of a potential choice, and exploration, can help us understand some of the problems our previous model (Kalidindi et al., 2005b,a)(N-LNDN) had in recreating the human net scores in the E(')F(')G(')H(') versions of the IGT. In the N-LNDN, the chance of exploration was based on adding an element of random noise into each unit in the Response Layer, before it entered a highest activation selection procedure. This random noise came from a normal distribution and was kept constant across the 100 trials in the IGT. This makes the chance of exploration dependent on the differences between the activation valuations for each deck. For example, if the estimated deck valuation for E is much greater than the other deck's estimated valuations, then the chance of an exploration trial is lowered. Also, an exploration trial is most likely to be an exploration of the next highest 'valued' deck. More explicitly, if,

$$Q_t(E) > Q_t(G) >> Q_t(F) > Q_t(H),$$

then deck E is most likely to be chosen and there is a smaller chance of G being chosen. However, dependent upon the settings for the probability distribution of the random noise, the chances of F or H being selected is slim. If the random noise is allowed to be too large, then there is insufficient exploitation, and behaviour tends towards random.

When using the N-LNDN for the EFGH versions, this problem arises. During the task, decks E and G become much more highly valued than decks F and H, resulting in the initial, effective, setting for the random noise being no longer sufficient to generate exploration. Therefore, we get greedy action on almost every trial, which does not reflect the normal human data (Bechara et al., 2000a). So, unlike the ϵ-*greedy* method, the likelihood of an exploration trial occurring in the N-LNDN is dependent on the relative activations of the units in the Memory Layer (Kalidindi et al., 2005b), as well as the amount of random noise added to the system. This contrasts with the ϵ-*greedy* method, where exploration trials are dependent on a pre-set exploration rate, ϵ, and are not dynamically influenced by the particular contingencies encountered.

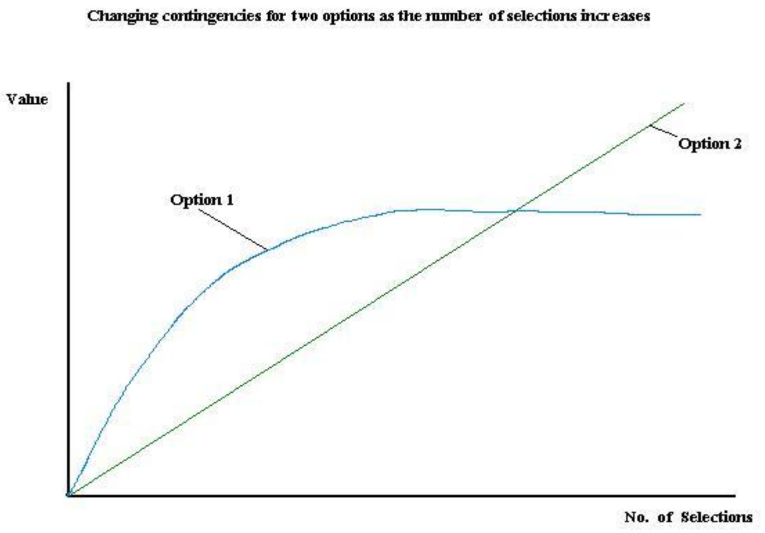


Figure 4: **Graph of the changing value functions for two options over a number of selections of that choice, where Option 2 becomes more highly valued than Option 1 after many selections.**

Even if an exploration trial does occur, there are again, differences in how the deck for that trial is selected. In the ϵ-*greedy* method, each deck/choice has an equal chance of happening, including the choice with the highest estimated value. However, in the N-LNDN, an exploration trial is best described as when the estimated optimal choice is not selected. Additionally, in the N-LNDN, the chance of a deck being chosen in an exploration trial, is weighted against currently low valued choices. Due to this, the N-LNDN may get trapped in local maxima, and not always converge to the global solution. For example, once the τ in the N-LNDN has risen sufficiently that the task contingencies become established in the network, then, unless the value of the current best choice is or becomes close to the worst choice, the worst choice is not likely to be selected. A graphical representation of this type of contingency set is shown in figure 4. It is evident that these contingencies pose a problem for the N-LNDN, as the network would mainly select from option 1, and possibly never discover that option 2 is actually the superior choice in the long run.

Returning to the IGT, the N-LNDN can manage certain reversal problems, like that found in the ABCD versions. This is where the initially higher valued choices, A and B, (see Figure 5) after a number of selections, are found to actually be worth less than the other options, C and D. Specifically, in the ABCD task, deck A and, particularly, deck B, using mean return per selection, are initially more highly valued than C and D, but, if selections continue, even if dominated by decks A and B, then, A and B's values fall below the 'resting' valuation (the value the network decays to if no selections of that option are made for some time) of C and D. Therefore, by selecting what initially seems optimal, even with limited exploration, it is possible to discover the true contingencies of the ABCD task using the N-LNDN. This is because the 'resting' valuation of any deck or option using the N-LNDN is zero. What is meant by 'resting' value, is the value that all options start with and eventually decay back to, if not selected.

In the ABCD version of the IGT, after sufficient selections of decks A and B, after card 9 in each deck, their mean returns per selection fall below zero (see Figure 5). This makes the likelihood of C and D being selected greater than A and B, because C and D's valuations are near zero or have fallen to zero. Even without decay in the information held about the decks, if A and B's valuation falls below that of C and D then C and D will be selected more frequently. Basically, if A and B's valuations fall below the initial valuation of C and D through selection of A and B's, contingencies, or the random factor is sufficient to create regular exploration throughout the task, then N-LNDN will solve the task. However, if the contingencies of the task are like those in Figure 4, where the value of Option 1 never falls below the early valuations for Option 2, then dependent on the variation in the random factor, Option 2 may never be selected sufficiently for the N-LNDN to discover that Option 2 is the optimal choice over the long-term.

From this discussion, the following question arises, can we not solve this by correctly setting the random noise, so C and D are investigated often enough to learn their contingencies? Yes, we can, but the random noise distribution would have to be set for each new task and require some knowledge of the expected differences in contingencies. Therefore it does not currently provide a good multi-purpose learning algorithm. This problem could be solved by having a dynamic addition of random noise, that was dependant on the relative activations of the units in the Memory Layer. It would then be possible to always get the chance of exploration on a trial. However, it is probably simpler to model this as two processes, like in the ϵ-*greedy* method.

This leads into a further discussion about what it means to explore. If we define exploration as 'picking alternatively from the current estimated optimal strategy', then using ϵ-*greedy* methods, the algorithm will actually pick the optimal strategy on some ϵ trials. Thus, the actual exploration rate is

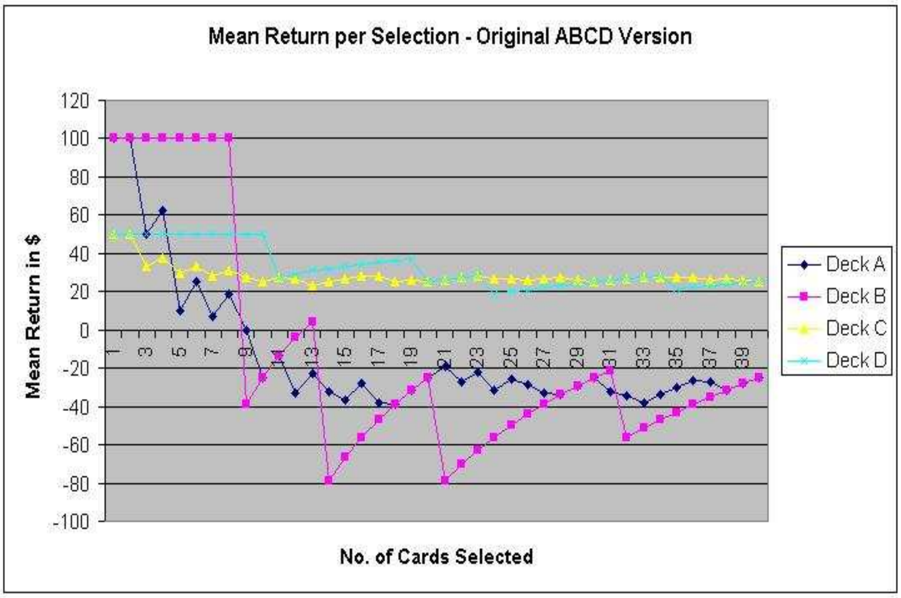


Figure 5: **Mean return per card selection from a particular deck based upon the number of selections made in a deck (using sample-averaging). For example at the 10th selection from deck B the mean return is $-25, as the total wins from deck B are $1000, so far, and the total losses are $1250.**

lower than the value of ϵ, as set out in section 4.1.

4.4 What is exploration?

In the currently used ϵ-*greedy* method, the choice of action during an exploration trial is random. This may not be the case for humans, often, at least at an intuitive level, an explorative action for a human seems to be, more like trying what is not currently normal behaviour. Alternatively expressed as, a break from the current strategy; choosing alternatively to what is currently considered optimal or re-testing the water. This re-testing could be inherent in humans, in an evolutionary sense, for a number of reasons.

(i) The contingencies of the task change.

(ii) The current strategy is not sufficiently effective.

(iii) New options may present themselves.

5 Results and analysis using *sample-Averaging*

The colour maps found in figures 6, 7, 10, 13, 8, 9, 14 and 15 show the absolute differences between the human net score profiles and the simulated data, for each of the 5, 20 card blocks in the 100 trial/card choice task, with varying exploration rates. The range of exploration rates, from 0-1, are simulated with steps of 0.01. The exploration rates used in the net score graphs, shown on the right hand side in each of the aforementioned figures, is the rate that most closely matches the human data across the five data points. This is calculated as the exploration rate that has the least absolute difference between the human data and the simulated data across all blocks. This is because the least absolute difference gave the best qualitative match when compared with least mean square error.

The data used to show the human profiles for comparison against the simulated data, in each of the right hand graphs in figures 6, 7, 10, 13, 8, 9, 14 and 15, comes from an amalgamation of data from relevant studies, comprising: Bechara et al. (2000a, 1999); Clark et al. (2003); Bechara and Damasio (2002a); Bechara et al. (2002b). The net scores can be found in table 3.

5.1 Amalgamation of IGT data across studies

From the studies considered in table 3 it was possible to get the standard error of the mean (SEM). If the standard deviation (SD) represents the variation in the values of a variable, then the SEM represents the spread of the means in the samples, were you to keep taking samples. This can be reconsidered as, the SD providing an idea of the variability in single observations, whereas the SEM gives an idea of the accuracy of the mean. These two measurements are related by:

SEM = $\frac{SD}{\sqrt{SampleSize}}$

Therefore,

SD = SEM $\times$ $\sqrt{SampleSize}$

Following this, it was necessary to add the data together from the different studies to get mean net scores and standard deviations.

Net Score Blocks	1-20	21-40	41-60	61-80	81-100	Total	No. of Participants
Task Variant							
ABCD-N	-5.9	2.3	7.1	8.6	9.7	21.8	54
t	-9.6	3.5	11.8	7.2	11.1	26.4	N/A
Sig ($P < 0.05$)	yes	yes	yes	yes	yes	yes	N/A
Sig ($P < 0.001$)	yes	yes	yes	yes	yes	yes	N/A
ABCD-V	-4.7	0.7	2.0	-3.0	-4.2	-9.2	15
t	-2.4	0.6	1.4	-2.2	-3.0	-2.8	N/A
Sig ($P < 0.05$)	yes	no	no	yes	yes	yes	N/A
Sig ($P < 0.001$)	no	no	no	no	no	no	N/A
EFGH-N	1.0	6.5	10.0	9.0	8.5	35	20
t	1	3.3	6.7	4.5	4.3	6.3	N/A
Sig ($P < 0.05$)	no	yes	yes	yes	yes	yes	N/A
Sig ($P < 0.001$)	no	yes	yes	yes	yes	yes	N/A
EFGH-V	-2.0	0.0	0.0	1.0	3.0	2	10
t	-1.0	0.0	0.0	0.5	1.0	0.4	N/A
Sig ($P < 0.05$)	no	no	no	no	no	no	N/A
Sig ($P < 0.001$)	no	no	no	no	no	no	N/A
A’B’C’D’-N	-4.7	3.7	7.0	7.4	7.7	21.1	66
t	-5.2	4.8	7.7	7.5	6.3	21.1	N/A
Sig ($P < 0.05$)	yes	yes	yes	yes	yes	yes	N/A
Sig ($P < 0.001$)	yes	yes	yes	yes	yes	yes	N/A
A’B’C’D’-V	-6.1	0.0	-3.4	-3.1	-4.7	-17.3	18
t	-4.6	-1.5	-2.8	-2.2	-2.0	-10.1	N/A
Sig ($P < 0.05$)	yes	no	yes	yes	yes	yes	N/A
Sig ($P < 0.001$)	yes	no	no	no	no	yes	N/A
E’F’G’H’-N	1.4	2.3	11.4	10.1	9.6	34.8	36
t	1.0	4.4	11.5	6.7	6.5	8.3	N/A
Sig ($P < 0.05$)	no	yes	yes	yes	yes	yes	N/A
Sig ($P < 0.001$)	no	yes	yes	yes	yes	yes	N/A
E’F’G’H’-V	-1.6	-3.9	-3.1	0.0	-5.5	-14.1	16
t	-1.0	-1.5	-1.1	0.0	-2.9	-3.2	N/A
Sig ($P < 0.05$)	no	no	no	no	yes	yes	N/A
Sig ($P < 0.001$)	no	no	no	no	no	no	N/A

Table 3: **An amalgamation of net score data from relevant studies, comprising, Bechara et al. (1999, 2000a); Clark et al. (2003); Bechara and Damasio (2002a) and (Bechara et al., 2002b). The nomenclature used to describe the task variants is their version names with the -N for normals and -V for VMF net scores. It also includes the results of significance testing (using Student’s t-test) with a null hypothesis (NH), that the net scores are random. 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 $P < 0.001$.**

The mean net scores were just taken as weighted averages and added together.

mean net score = $\sum_{i=1}^{n} w_i * \mu_i$

where w denotes the Number of participants in study / Total number across studies for that IGT variant,

n denotes the number of studies added together.

To calculate the SD across multiple studies, the following definition of the variance is used.

V(X) = E($X^2) - \mu^2$

where V(X) denotes the variance in the data set X,

E(X) denotes the expected value for the data set X (the mean),

μ denotes the mean.

So, if we rearrange this and apply the sample corrected sum of squares then,

V(x) = $\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2$

where

$\sum_{i=1}^{N} (x_i - \overline{x})^2$ = $\sum_{i=1}^{N} x_i^2 - \frac{(\sum_{i=1}^{N} x_i)^2}{N}$.

Using the definition of the mean,

$\mu = \frac{\sum_{i=1}^{N} x_i}{N}$

therefore,

$\mu^2 = \frac{(\sum_{i=1}^{N} x_i)^2}{N^2}$

$\frac{(\sum_{i=1}^{N} x_i)^2}{N} = N\mu^2.$

Now our definition of the variance becomes,

V(X) = $\frac{1}{N}(\sum_{i=1}^{N} x_i^2 - N\mu^2)$

= $\frac{1}{N}\sum_{i=1}^{N} x_i^2 - \mu^2.$

By definition

V(X) = $(SD)^2$

$\frac{1}{N}\sum_{i=1}^{N} x_i^2 = (SD)^2 + \mu^2$

$\sum_{i=1}^{N} x_i^2 = N((SD)^2 + \mu^2).$

Now, as we have the SD and the mean net score for each of our samples, it is possible to calculate $\sum_{i=1}^{N} x_i^2$ for each data sample. In addition, because adding the samples together is assumed to be the same as taking a bigger sample, we can just add the $\sum_{i=1}^{N} x_i^2$ terms across the same IGT versions. Earlier, a method was set out to show how the mean between two samples can be calculated, and we have the values for the number of participants. Finally, these values can be plugged into one of our original definitions of variance and then square rooted to provide the new SD for the group of samples together.

5.2 Consideration of the Human Data

In table 3, an attempt has been made to see where behaviour is significantly different from a random binomial series. This is possible because the net scores are based on a binary score, either a selection is advantageous, 1 (deck C or D), or disadvantageous, -1 (deck A or B). Of particular interest is whether the VMF patients are acting significantly differently from a random agent. An assumption is made that the human data is normally distributed and that an agent acting randomly has a mean net score of 0 and can be described with a binomial series with probability, p = 0.5. Then Student's t-test was used to measure the likelihood of a type I error occurring in our acceptance that the means of our two data sets are significantly different. The following equations and procedures were applied to produce the significance test results found in table 3.

To apply Student's t-test it is necessary to have three values: (1) t-value; (2) degrees of freedom (d.f.) and (3) α level.

t-value

$$\frac{Signal}{Noise} = \frac{difference-between-group-means}{variability-of-groups}$$

$$= \frac{\overline{x_H}-\overline{x_B}}{SE(\overline{x_H}-\overline{x_B})}$$

= t-value.

Where $\overline{x_H}$ denotes the mean of the data set X_H and $\overline{x_B}$ denotes the mean of the data set X_B

$$SE(\overline{x_H}-\overline{x_B}) = \sqrt{\frac{Var_H}{n_H} + \frac{Var_B}{n_B}}$$

therefore,

$$t = \frac{\overline{x_H}-\overline{x_B}}{\sqrt{\frac{Var_H}{n_H}+\frac{Var_C}{n_C}}}$$

degrees of freedom

$d.f. = n_H + n_B - 2$

α **-value** = probability of finding a Type I error and is equivalent to the acceptable power value (P-value),

where,

$\overline{x_H}$ denotes the mean for the human data,

$\overline{x_B}$ denotes the mean for the binomial distribution data, which is zero,

Var_H denotes the variance in the human data,

n_B denotes the number of data points in the binomial distribution, as this is actually equivalent to having an infinite number of data points, this means the whole term $\frac{Var_C}{n_C} \rightarrow 0$. Therefore, for the comparison to a distribution the t-value is reduced to

$$t = \frac{\overline{x_H} - 0}{\sqrt{\frac{Var_H}{n_H} + 0}}$$

$$t = \frac{\overline{x_H}}{\sqrt{\frac{Var_H}{n_H}}}$$

It follows from this that the degrees of freedom variable will also have a value which tend to ∞ because $n_C \rightarrow \infty$ so d.f. tends to infinite. The values selected for the power, P, can be found in table 3.

In other cases the standard equation set out above is used, where both the standard deviations and sample sizes for both groups under consideration are used.

The results of these calculations, which are in table 3, show that the normal control profiles are both significant and highly significantly different from a random Bernoulli process, apart from in block 1 on the EFGH and E’F’G’H’ versions. The convention for this paper is that significant reflects a P $<$ 0.05 and highly significant reflects a P $<$ 0.001. However, for the VMF profiles, only the 1st block (cards 1-20) for the A’B’C’D’ version is highly significantly different from random. One would expect that this would also be the case for the ABCD version, as the contingencies for the first ten cards for each deck are the same. However, the first block for the ABCD version is only significantly (P $<$ 0.05) different from random, and not highly significant (P $<$ 0.001). A possible explanation for this might be that VMF patients actually seem to become worse at the task (in that VMF patients are more influenced by the regular rewards and punishments than the uncertain results) as a function of repetition over time (see figure 3, Bechara et al. (2000b)), and, because much of the data for the A'B'C'D' variant was collected after testing on the original ABCD version, with often the same VMF patients, this could be the case.

For the VMF patients on the EFGH and E'F'G'H' versions of the IGT, only the net score on the 5th block on the E'F'G'H' version is significantly different from random. The differences between the E'F'G'H' version and the EFGH version is that E' and G' become better and F' and H' become worse per 10 cards in each deck. Therefore, for the E'F'G'H' version, if the patients were making any kind of reasonable estimation of the uncertain positive outcomes, then you would expect them to have larger positive net scores, compared to the EFGH version. This is not the case, as the VMF net scores for the EFGH versions are more positive than those for the E'F'G'H' variant. This may be due to the increased complexity in the changing contingencies in the E'F'G'H' version, compared to the EFGH version. The EFGH version net score profile for the VMF patients can actually be well described by the ϵ-greedy method using *sample-averaging*, with ϵ set to 0.9. However, using *sample-averaging*, the closet match to VMFs on the E'F'G'H' version uses an exploration rate of 1, which makes it random, and so does not properly describe the final block or the overall qualitative shape of the curve (getting worse, better and then worse again, see figure 15). In addition, across all blocks there is an overall significant tendency for VMF patients to select more from the disadvantageous decks on the E'F'G'H' version.

The VMF profiles for the ABCD and A'B'C'D' versions of the IGT are both significantly different from random, for the first block, and highly significant for the A'B'C'D' version, as aforementioned. However, human VMF profiles are not significantly different from the normal human profiles in this first block (see table 5). Hence, early on in the IGT for the ABCD and A'B'C'D' versions when decks A(') and B('), on the basis of known results, are the best decks VMF and normal subjects are actually selecting advantageously. Although, this could be for different reasons in VMF patients than for normal healthy controls. The VMF patients could be more driven by the consistent $100 and $50 wins, than by the uncertain losses, as they seem to be later in the task. As discussed earlier, it has also been suggested that initial stimulus-reinforcement associations are preserved in VMF patients (Fellows and Farah, 2003). The influence of regular results is shown by the VMF's, significantly different from random, disadvantageous net scores gained in blocks 3, 4 and 5 for the A'B'C'D' version, and those

Net Score Blocks	1-20	21-40	41-60	61-80	81-100	Total	No. of Participants
Task Variant							
ABCD-N	-5.9	2.3	7.1	8.6	9.7	21.8	54
ABCD-V	-4.7	0.7	2.0	-3.0	-4.2	-9.2	15
t	-0.6	1.3	3.28	5.4	8.5	10.9	N/A
Sig ($P < 0.05$)	no	no	yes	yes	yes	yes	N/A
Sig ($P < 0.001$)	no	no	no	yes	yes	yes	N/A
EFGH-N	1.0	6.5	10.0	9.0	8.5	35	20
EFGH-V	-2.0	0.0	0.0	1.0	3.0	2	10
t	1.3	2.3	4.0	2.8	1.5	4.4	N/A
Sig ($P < 0.05$)	no	yes	yes	yes	no	yes	N/A
Sig ($P < 0.001$)	no	no	yes	no	no	yes	N/A
A’B’C’D’-N	-4.7	3.7	7.0	7.4	7.7	21.1	66
A’B’C’D’-V	-6.1	0.0	-3.4	-3.1	-4.7	-17.3	18
t	0.9	4.0	6.8	6.0	4.8	18.8	N/A
Sig ($P < 0.05$)	no	yes	yes	yes	yes	yes	N/A
Sig ($P < 0.001$)	no	yes	yes	yes	yes	yes	N/A
E’F’G’H’-N	1.4	2.3	11.4	10.1	9.6	34.8	36
E’F’G’H’-V	-1.6	-3.9	-3.1	0.0	-5.5	-14.1	16
t	1.4	3.3	5.0	3.1	6.4	8.2	N/A
Sig ($P < 0.05$)	no	yes	yes	yes	yes	yes	N/A
Sig ($P < 0.001$)	no	no	yes	no	yes	yes	N/A

Table 4: **Comparison of mean for the normal controls vs. the VMF patients for each version of the IGT. The nomenclature used to describe the task variants is their version names with the -N for normals and -V for VMF net scores. It also includes the results of significance testing (using Student’s t-test) with a null hypothesis (NH), that the net scores have the same mean. The alternate hypothesis (AH) is that they have different means. The answer of yes or no signifies whether the NH can be rejected at $P < 0.05$ and $P < 0.001$.**

found in blocks 4 and 5 for the ABCD version. As discussed earlier, it has been suggested by Fellows and Farah (2005) that the IGT problem is due to difficulties, found in orbitofrontal patients, with reversal learning (Rolls, 2000). Although this explanation is consistent with the results gained by, Fellows and Farah, it does not seem to completely satisfy the ABCD profiles found in Bechara et al.'s data sets, in that, for blocks 2 and 3, the VMF patients select 'randomly' and then return to net scores similar to block 1 for blocks 4 and 5. If it were solely a simple reversal learning problem, where the VMF patients perseverate at the point of reversal but after a few errors learn the new contingencies, one might expect there to be a slow gradual improvement in performance after the reversal. This not what happens though. Therefore, raising the question of what might be occurring during blocks 2 and 3 in the ABCD version and block 2 in the A'B'C'D'. This question one of the main areas of investigation in this paper.

5.3 Normal controls on the original ABCD version (see figure 6)

If we look at the right hand graph (see figure 6), it is clear that the simulated data is a reasonable fit for the human controls' net scores, apart from the slightly faster learning/exploitation of the advantageous decks in blocks 2 and 3. There is also an earlier leveling off in block 3 as the amount of exploitation of the advantageous decks in the simulated data set, compared to the human data. In the human data, it seems that there would be increased selection of the good decks, were the task to continue beyond 100 trials. The leveling off of the simulated data is due to the level of exploration. The height at which the net score curve levels off is inversely proportional to the exploration rate. This is due to a higher exploration rate reducing the chance of greedy exploitation on any trial. The best match to the human data, using *sample-averaging*, was found with an exploration rate of 0.58, i.e. 58% of the time the simulation will pick a deck at random, rather than selecting the deck with the highest expected return. At an intuitive level, this seems high, in that, if all information is held and accurately calculated, why would exploration occur more than half the time. This will be further examined in the discussion, section 7.

5.4 VMF patients on the original ABCD version (see figure 7)

From the human VMF patient net score profile, there is seemingly a stage of learning in blocks 1-3 (0-60 cards), after which, however, there is a move back to disadvantageous behaviour. Therefore overall, the VMF patients 'favour' the disadvantageous decks. This may be because in the early parts of the

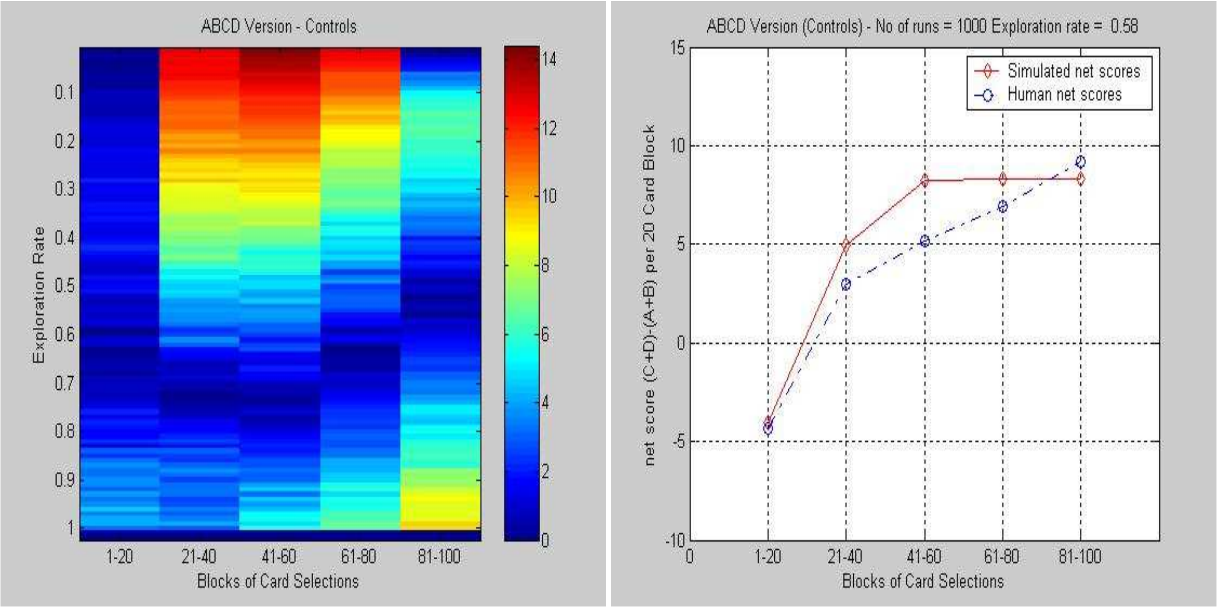


Figure 6: **(Left) Colour map showing the differences between the human control net scores on the ABCD version and those produced by the simulation, with the y-axis representing the varying exploration rate. (Right) Comparison of normal human control net scores (n = 54) on the ABCD version of the IGT with the 0.58-greedy action-value algorithm**

task, until some losses are encountered, decks A and B have higher average wins than decks C and D. (The negative net score in both normal controls and VMF patients in the first block is a result of the contingencies set out in the original ABCD score-card, see figure 1.) Is it also this early evaluation that is difficult to override and later becomes dominant again (from 60-100 cards)? Alternatively, could it be VMF patients' tendency to prefer regular results than uncertain ones.

5.5 Normal controls on the A'B'C'D' version (see figure 8)

There seems to be a good match between the simulated data and the normal human net scores. A closer match could be gained by increasing the exploration rate to 0.66. This would reduce the probability of picking the optimal action, even when a fairly accurate estimation of the true value of each action exists. The exploration rate controls the level at which the net score curve levels off.

5.6 VMF patients on the A'B'C'D' version (see figure 9)

The simulated data does not match the VMF patients' net scores for this version of the IGT. This is because the worst mean performance possible for this simulation method, as aforementioned, is a net

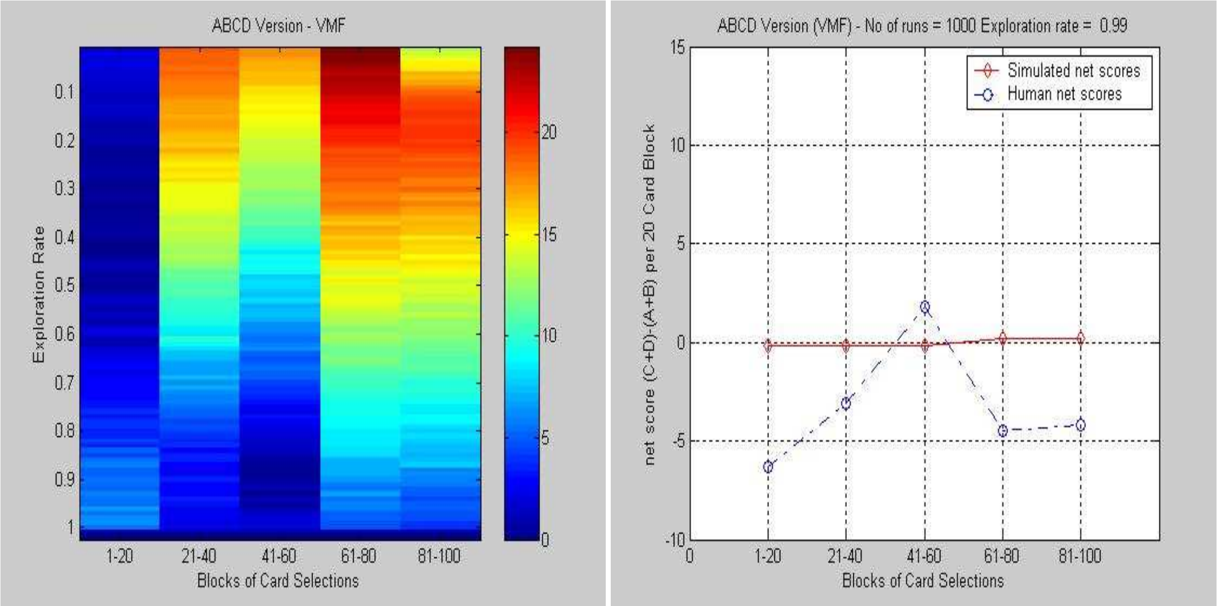


Figure 7: **(Left) Colour map showing the differences between the human VMF patient net scores on the ABCD version and those produced by the simulation, with the y-axis representing the varying exploration rate. (Right) Comparison of human VMF patient net scores (n = 15) on the ABCD version of the IGT with the 0.99-greedy action-value algorithm**

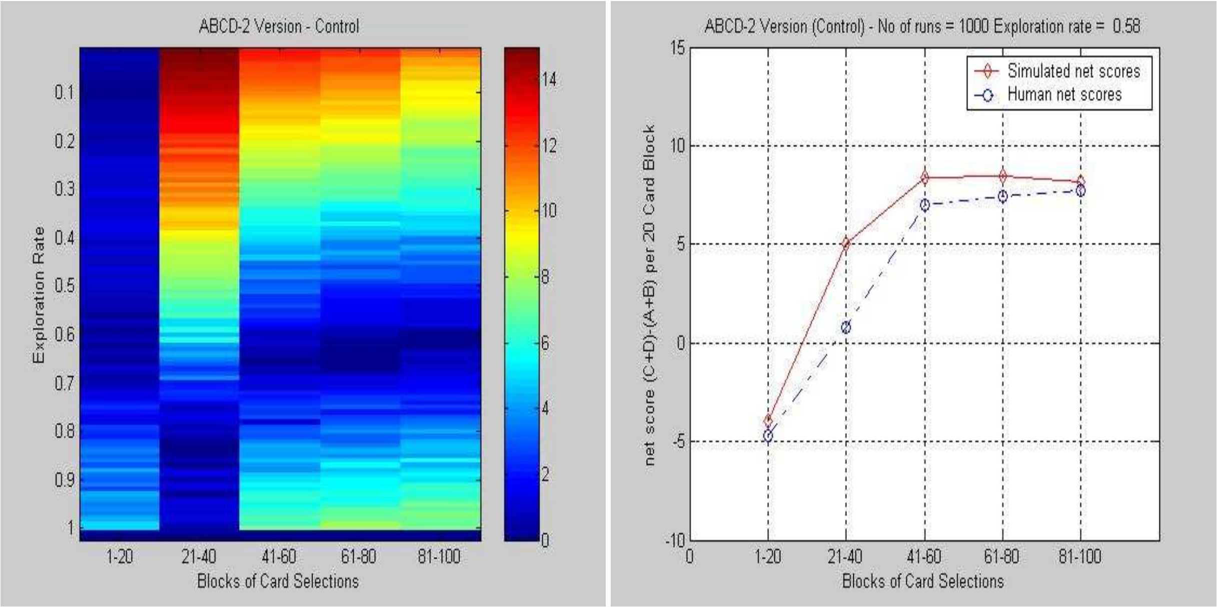


Figure 8: **(Left) Colour map showing the differences between the human control net scores on the A’B’C’D’ version and those produced by the simulation, with the y-axis representing the varying exploration rate. (Right) Comparison of normal human control net scores (n = 66) on the A’B’C’D’ version of the IGT with the 0.58-greedy action-value algorithm**

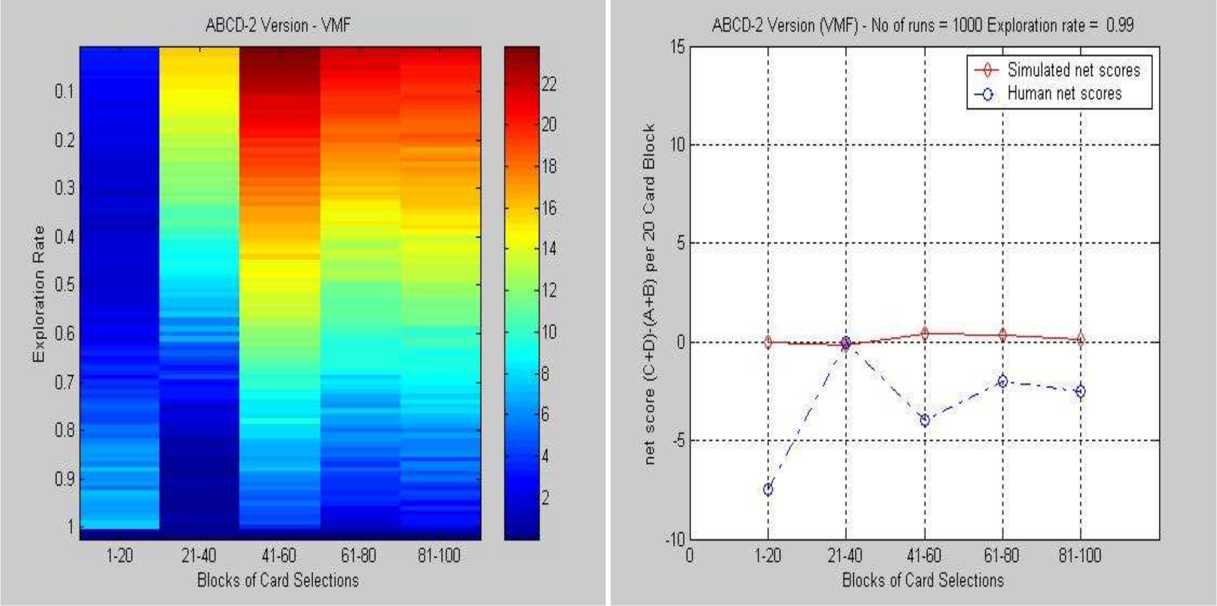


Figure 9: **(Left) Colour map showing the differences between the human VMF patient net scores on the A'B'C'D' version and those produced by the simulation, with the y-axis representing the varying exploration rate. (Right) Comparison of human VMF patient net scores (n=18) on the A'B'C'D' version of the IGT with the 0.99-greedy action-value algorithm**

score of 0 (random behaviour). However,, the VMF patients' net score profile is regularly negative, since it contains more bad selections than good ones. There is some improvement in initial performance for the VMF patients (block 2), this could just be increased exploration, or simply a move away from deck B after the large reversal. This improvement is then followed by a return to more disadvantageous selections, as in the original ABCD version.

5.7 Normal controls on the EFGH version (see figure 10)

Here the simulated data is a good match for normal human controls' net score profiles. It is interesting to note that in the EFGH version, normals begin the task at random or slightly better. This is probably due to the score card for deck E, where participants quickly receive a large win $1250, on the 3rd card in that deck (see figure 1), and so can adjust their evaluation of deck E, to 'good', more rapidly. This is in contrast to the ABCD version where the first large uncertain loss, in magnitiude, in deck B, is not until the 9th card in that deck. Therefore, it is much longer before the partcipant reveals the true nature of deck B in the ABCD version compared to deck E in the EFGH version. In the net score profiles, this difference is probably the cause of normal human controls beginning the ABCD task with an average

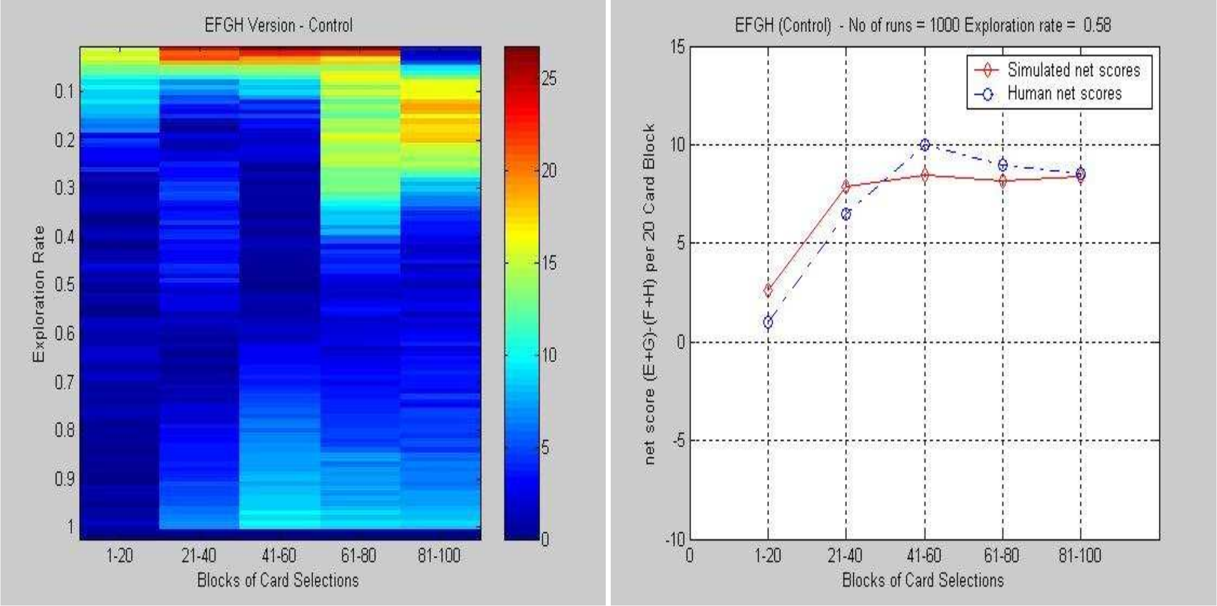


Figure 10: **(Left) Colour map showing the differences between the human control patient net scores on the EFGH version and those produced by the simulation, with the y-axis representing the varying exploration rate. (Right) Comparison of normal human control net scores (n = 20) on the EFGH version of the IGT with the 0.58-greedy action-value algorithm**

score of, approximately, -5 in the first block (1-20 cards).

5.8 VMF patients on the EFGH version (see figure 13)

For this variant of the task there only seems to be one set of data (Bechara et al., 2000a), but what we find is that the VMF patients are acting more or less randomly throughout this version of the task. This allows a close match between the *sample-averaging ϵ-greedy* method and the human VMF populations' net scores. Note, after sufficient trials to learn the contingencies, the *sample-averaging ϵ-greedy* method will not perform worse than random while the estimates of $Q^*(a)$ are based on *sample-averaging*. However, as in the early portions (1-20 cards) of the ABCD and A'B'C'D' versions, when there is limited sampling, estimates of $Q^*(a)$ favour the disadvantageous packs, as the samples have yet to reveal their uncertain punishments. This occurs particularly on one of the decks (B or B'). It actually turns out that deck B or B' is the best deck until card 9 in the deck is reached and the player losses $1250.

The net score profile for Bechara et al.'s VMFs on the EFGH variant are similar to Fellows and Farah (2005) on the original 'unshuffled' ABCD variant. Therefore, if we assume that a reversal learning deficit is the main cause of poor VMF patient performance on the IGT, then, as the reversal comes very early in

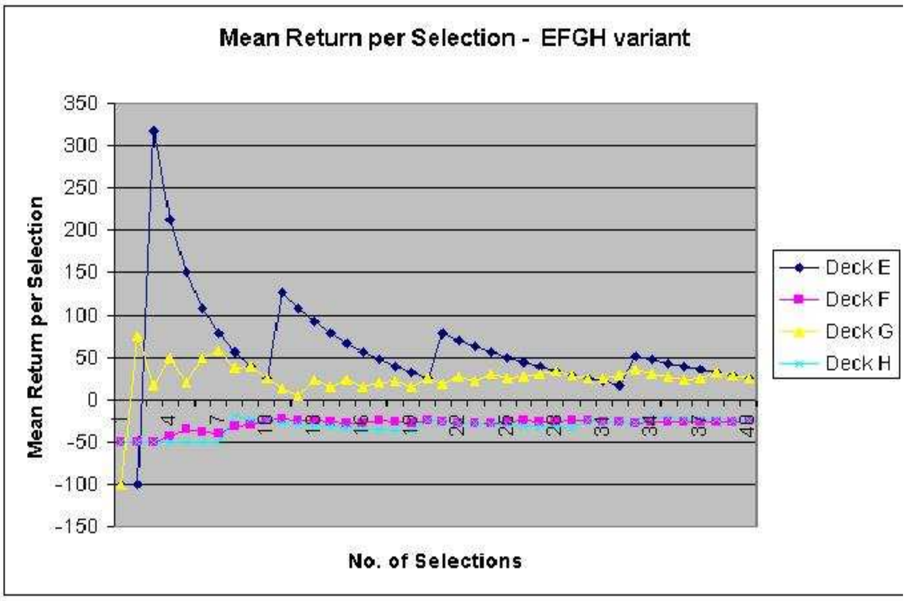


Figure 11: **Mean return per card selection, EFGH variant, where the x-axis is the number of selections made from a particular deck. Using sample-averaging the mean return per selection for a deck is calculated based on, rewards minus punishments gained after x selections from certain deck (based on the predefined contingencies of the IGT (?)) and then divided by x.**

the EFGH variant (see figure 11), by card 3 for deck E and card 2 for deck G, we would expect the VMF patients to perform better than when the reversal happens later, i.e. when the pre-reversal contingencies are less established. If we ignore variance, this seems to be the case as VMF patients perform less disadvantageously on the EFGH version, than on the original 'unshuffled' ABCD variant. However, there still is a reversal, unlike in Fellows and Farah (2005)'s 'shuffled' version of the ABCD IGT, where A and B are never, on average, better than decks C and D (see figure 12). So, it might be reasonable to expect the VMF patients to perform more closely to normals in this EFGH variant than VMFs do in the original 'unshuffled' ABCD variant. This is the case. But Bechara et al.'s patients on the EFGH variant act similarly to Fellows and Farah (2005)'s VMF patients on the original 'unshuffled' ABCD variant, and not better than them, which provides more support for the claim that they may be from differently performing populations. However, when compared to the profiles for the E'F'G'H' variant, then this random (or slow learning) profile is not repeated, even though the reversal occurs at the same time as in the EFGH version. In addition, the EFGH variant is the only study which has not been repeated, whereas the E'F'G'H' variant has been repeated, with consistency, and follows a pattern more reminiscent of the ABCD and A'B'C'D variants.

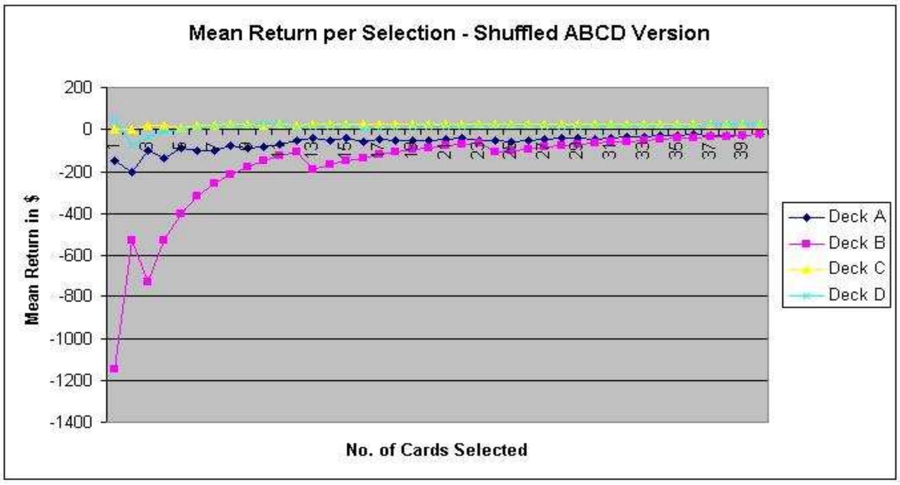


Figure 12: **Mean return per card selection, 'shuffled' ABCD variant (Fellows and Farah, 2005), where the x-axis is the number of selections made from a particular deck. Using sample-averaging the mean return per selection for a deck is calculated based on, rewards minus punishments gained after x selections from certain deck (based on the predefined contingencies of the IGT (?)) and then divided by x.**

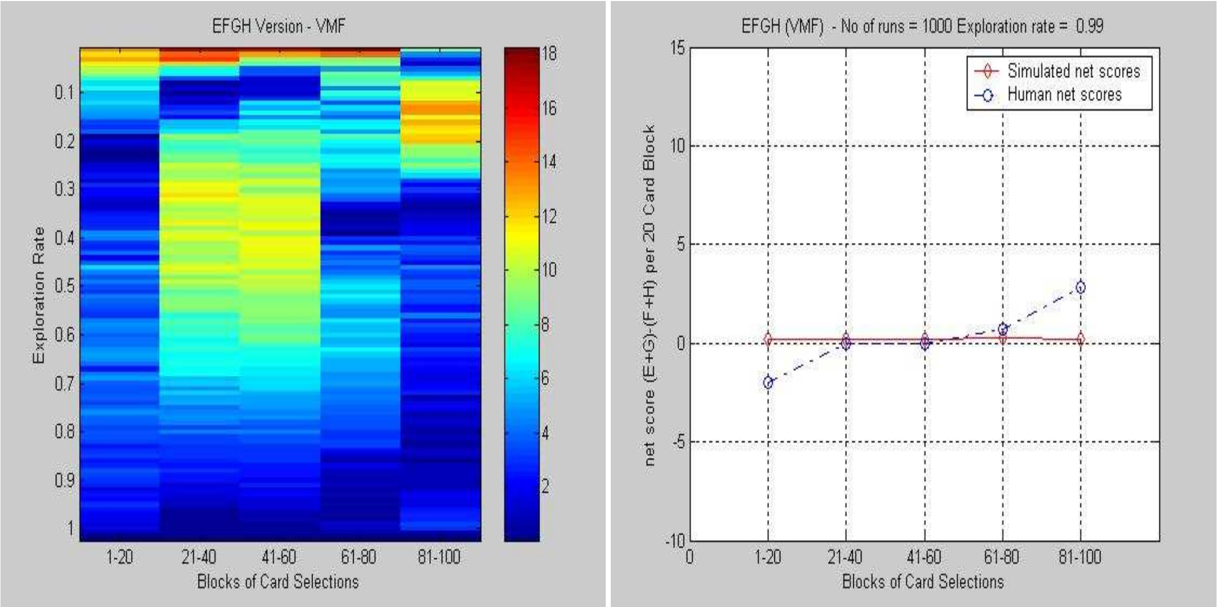


Figure 13: **(Left) Colour map showing the differences between the human VMF patient net scores on the EFGH version and those produced by the simulation, with the y-axis representing the varying exploration rate. (Right) Comparison of human VMF patient net scores (n = 10) on the EFGH version of the IGT with the 0.99-greedy action-value algorithm. (Note: an exploration rate = 0.9 provides a very good match to the human data)**

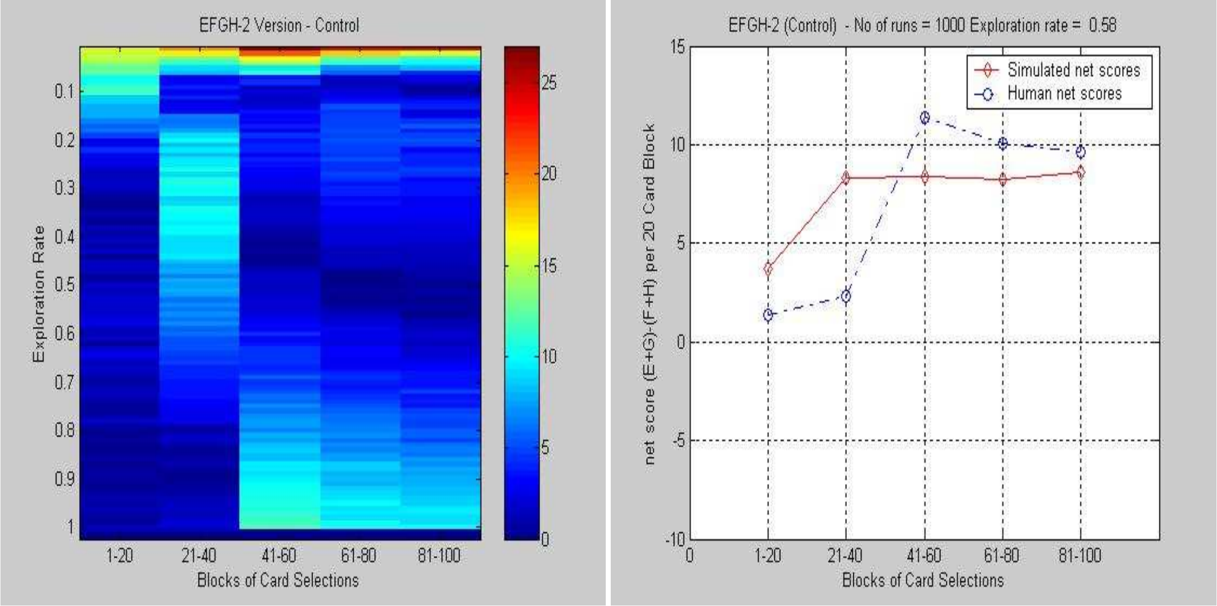


Figure 14: **(Left) Colour map showing the differences between the human control patient net scores on the E'F'G'H' version and those produced by the simulation, with the y-axis representing the varying exploration rate. (Right) Comparison of normal human control net scores (n=36) on the E'F'G'H' version of the IGT with the 0.58-greedy action-value algorithm**

5.9 Normal controls on the E'F'G'H' version (see figure 14)

Again, like the EFGH simulations, there is a reasonably good match to the normal human data, but a better match would be obtained with an exploration rate of 0.65. This variation in the exploration rate, could be due to the participant groups being different for each version of the task. This is because, once normal controls have learnt one version of the task, they understand the set-up, and can rapidly learn other versions of the IGT. Normal subjects become too good at the task.

Variations in the required exploration rate across IGT variants, could also be due to normal subjects working within a range of exploration rates, which are adjusted to the difficulty of the task. A single value $\epsilon = 0.58$ has been used to show that one parameter setting can produce a good replication of all the normal subject results.

5.10 VMF patients on the E'F'G'H' version (see figure 15)

The replication of the VMF data for E'F'G'H is not so close, but it is important to note that, as shown in table 3, only block 5 is significantly different from random selection. However, the net scores are consistently negative and overall have a probability $P < 0.002$ of being random. This is impossible to

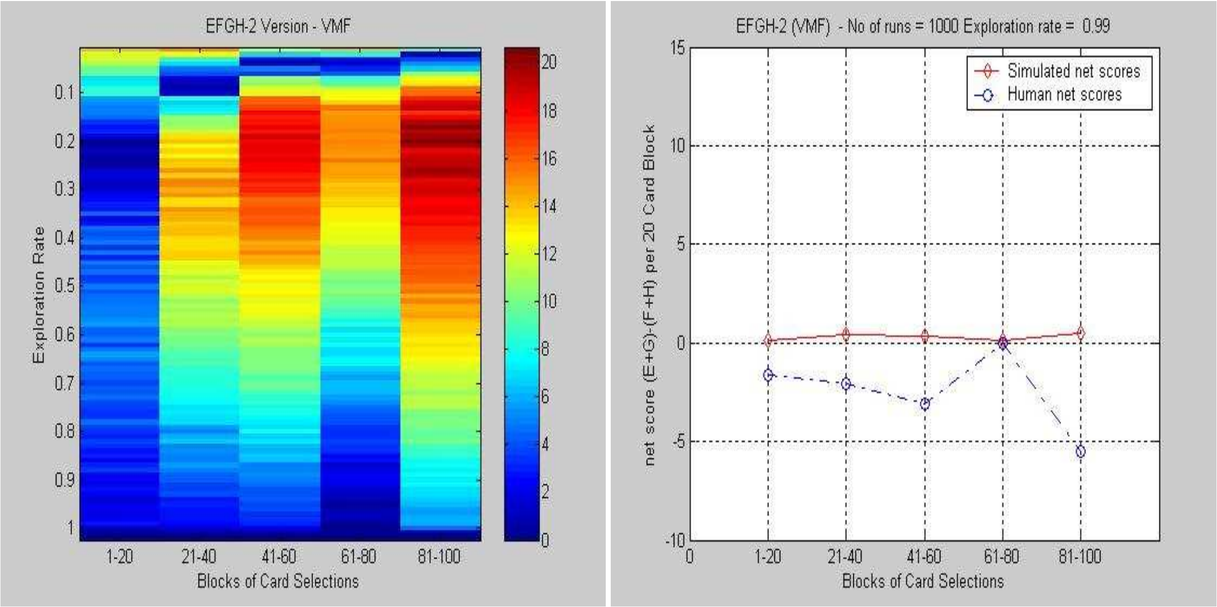


Figure 15: **(Left) Colour map showing the differences between the human VMF patient net scores on the E'F'G'H' version and those produced by the simulation, with the y-axis representing the varying exploration rate. (Right) Comparison of human VMF patient net scores (n=16) on the E'F'G'H' version of the IGT with the 0.99-greedy action-value algorithm**

obtain with *sample-averaging*, after the reversals in deck E' and G'.

6 Results and analysis using *error-driven* learning

In the results for *error-driven* learning a comparison has been made between the simulated data with two settings for the learning rate and, the normal control and VMF patient, net score profiles. To recreate the normal healthy control net scores, the exploration rate $\epsilon = 0.55$, and the learning rate $\gamma = 0.2$. To simulate the VMF data, the learning rate was changed to $\gamma = 0.95$, while the exploration remained at 0.55.

6.1 Normal controls on the original ABCD version (see figure 16)

This provides a good match to the human data.

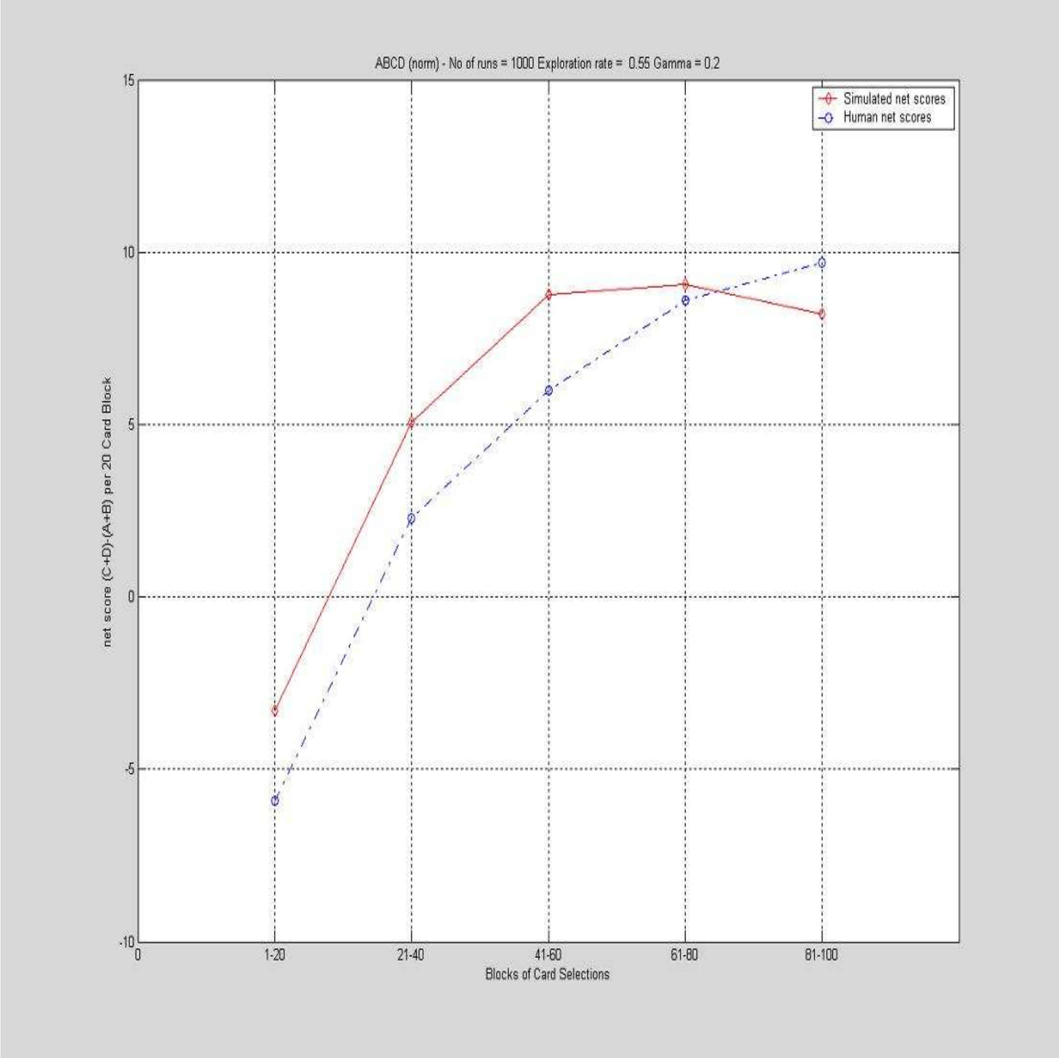


Figure 16: **Comparison of normal healthy human controls' net scores (n=54) on the ABCD version of the IGT with the 0.55-greedy error-driven valuation (γ = 0.2) algorithm (n = 1,000).**

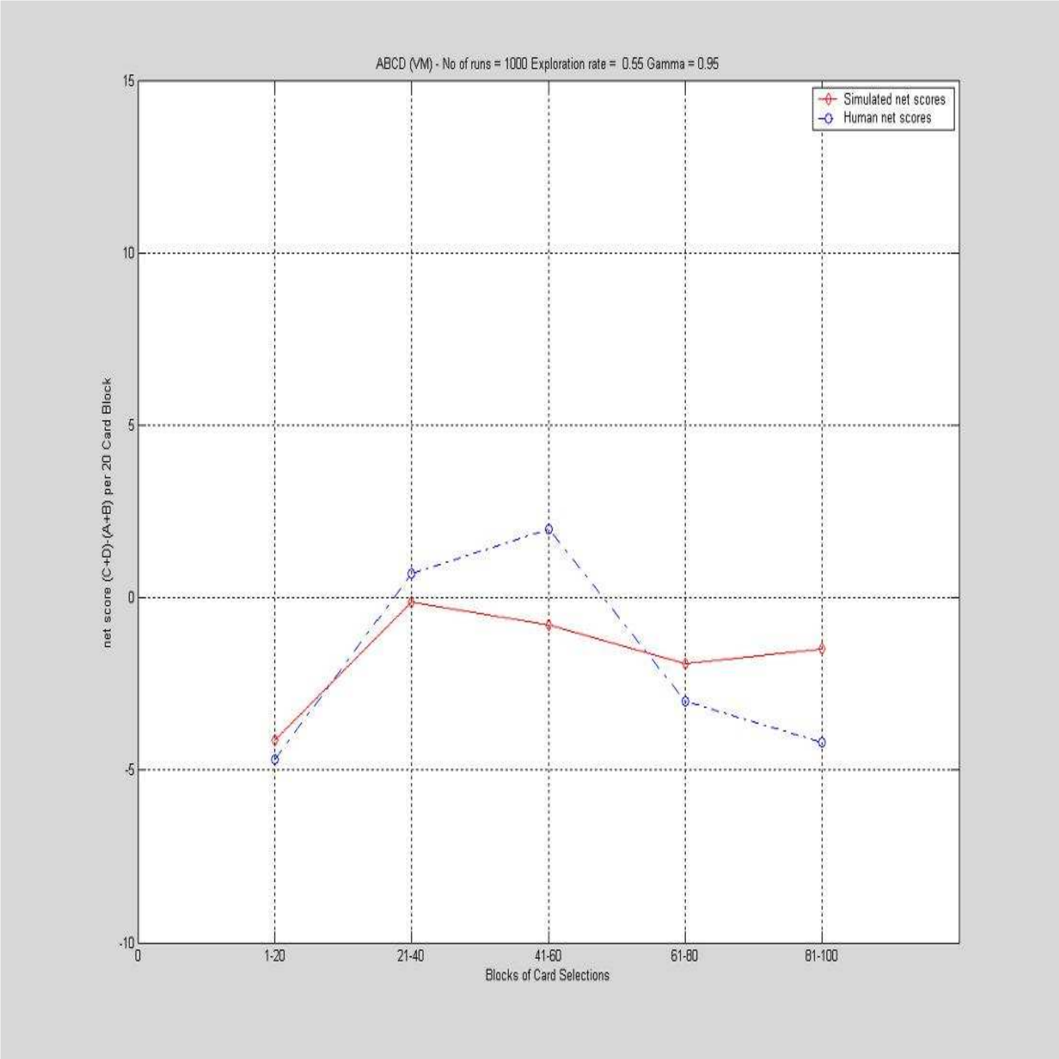


Figure 17: **Comparison of VMF patients' net scores (n=15) on the ABCD version of the IGT with the 0.55-greedy error-driven valuation (γ = 0.95) algorithm (n = 1,000).**

6.2 VMF patients on the original ABCD version (see figure 17)

The simulated profile is similar to the human patient profiles, and interestingly, on some simulated runs, with n = 100, almost perfect matches between the data sets can be generated in 1 in 4 attempts. However, a much larger number of simulated runs is used to try and minimize the variation. There will always be some variation, because of the random elements present in the simulation models.

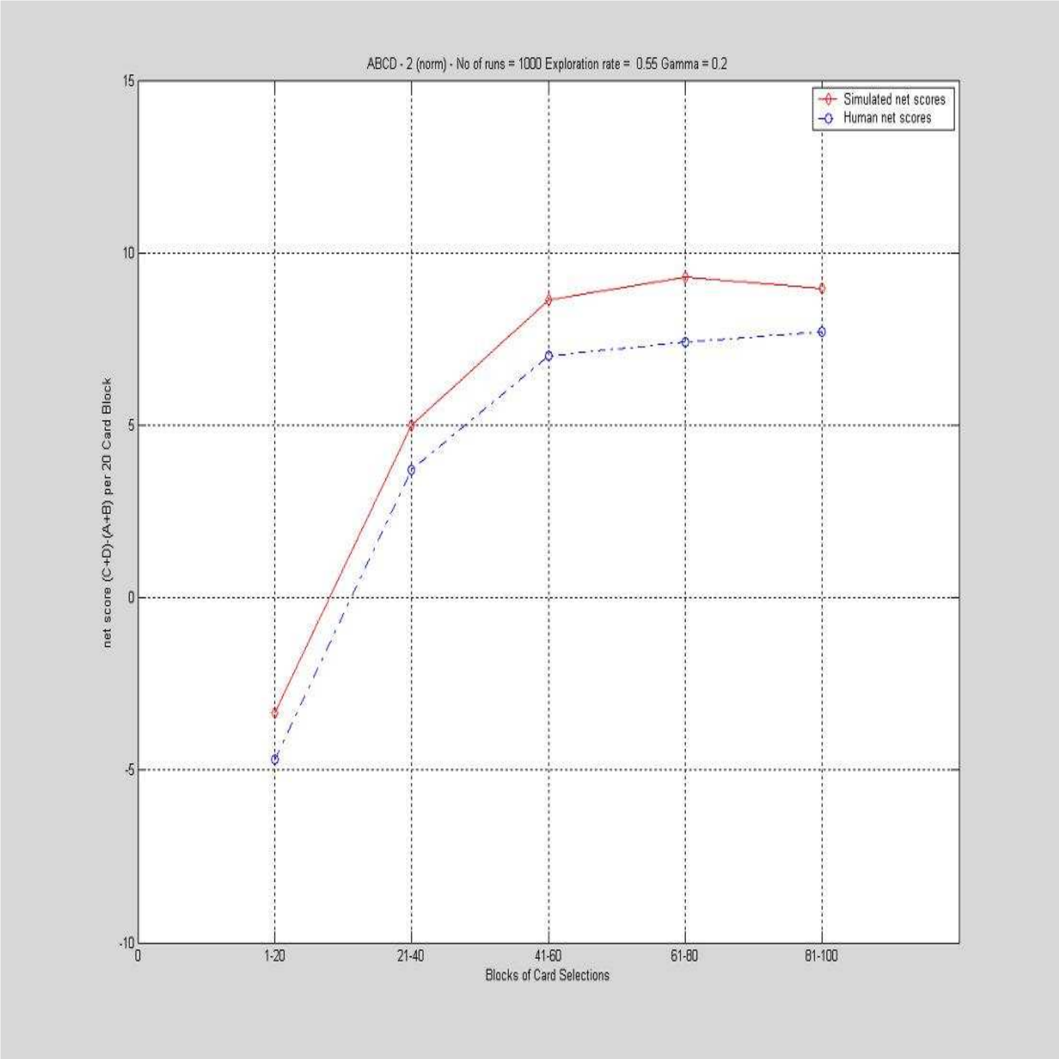


Figure 18: **Comparison of normal healthy controls' net scores (n=66) on the A'B'C'D' version of the IGT with the 0.55-greedy error-driven valuation (γ = 0.2) algorithm (n = 1,000).**

6.3 Normal controls on the A'B'C'D' version (see figure 18)

Again, with the same settings for the normal control simulations, a good match to the human healthy control data is generated.

6.4 VMF patients on the A'B'C'D' version (see figure 19)

For A'B'C'D' VMF profiles, a very good qualitative correlation to the human data has been generated with 1000 runs of the simulation.

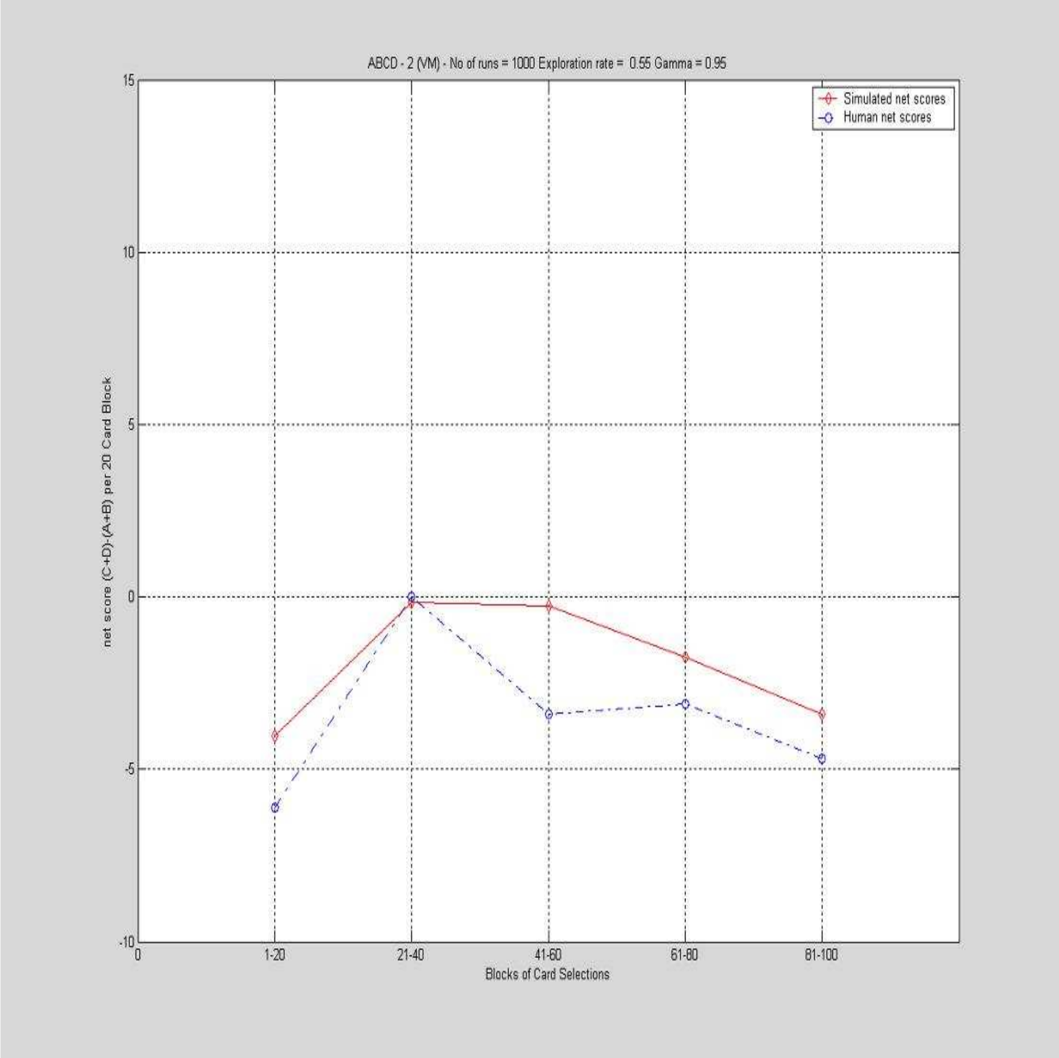


Figure 19: **Comparison of VMF patients' net scores (n=18) on the A'B'C'D' version of the IGT with the 0.55-greedy error-driven valuation (γ = 0.95) algorithm (n = 1,000).**

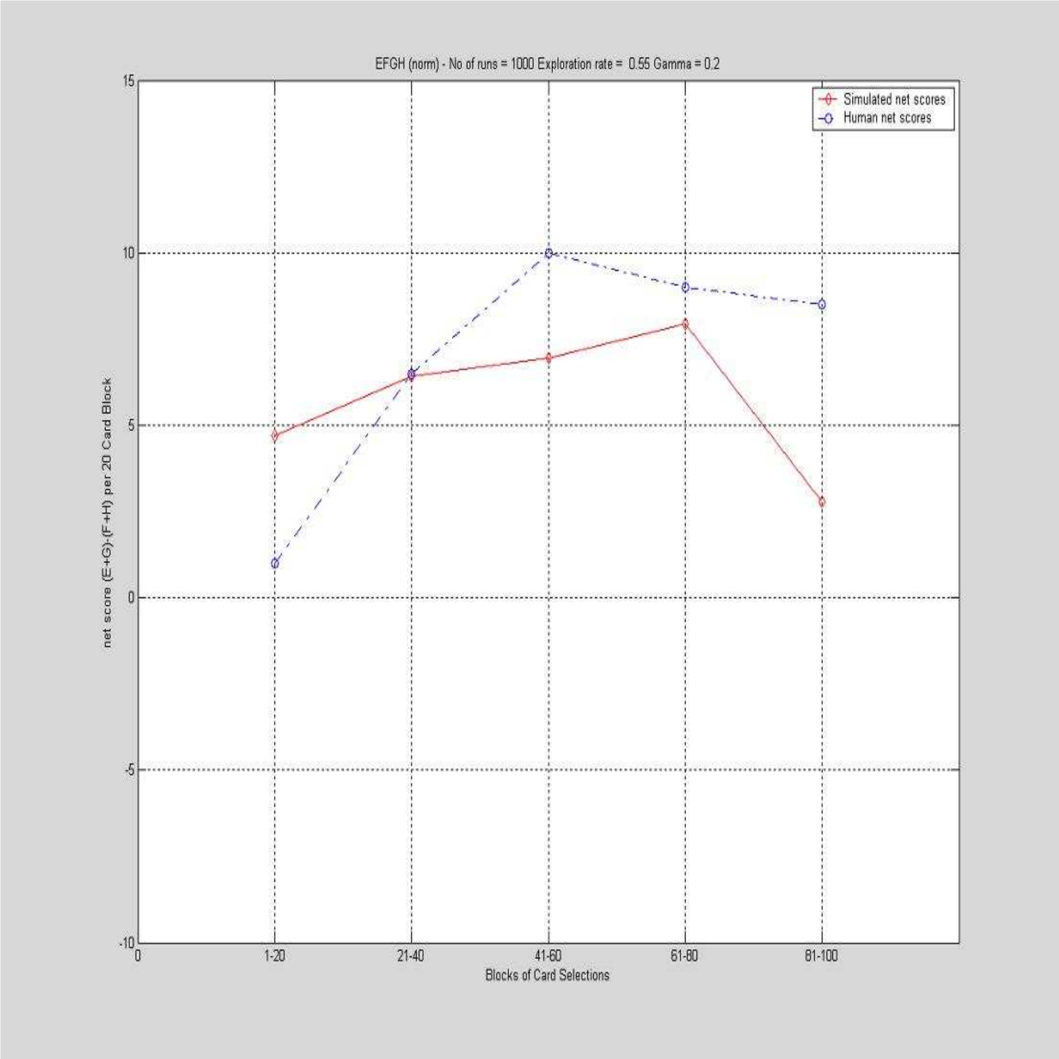


Figure 20: **Comparison of normal healthy controls' net scores (n=20) on the EFGH version of the IGT with the 0.55-greedy error-driven valuation (γ = 0.2) algorithm (n = 1,000).**

6.5 Normal controls on the EFGH version (see figure 20)

The simulated data provides a reasonable match to normal controls. The *error-driven* ϵ-greedy method is slightly weak in its match to the human data in blocks 1 and 3, but the general trend of the curve is similar.

6.6 VMF patients on the EFGH version (see figure 21)

The curve shape of the simulated data is very similar to the human data, but to actually fit the data points accurately, a rotation of about 45 degrees about block 1 would be required. This is the only VMF

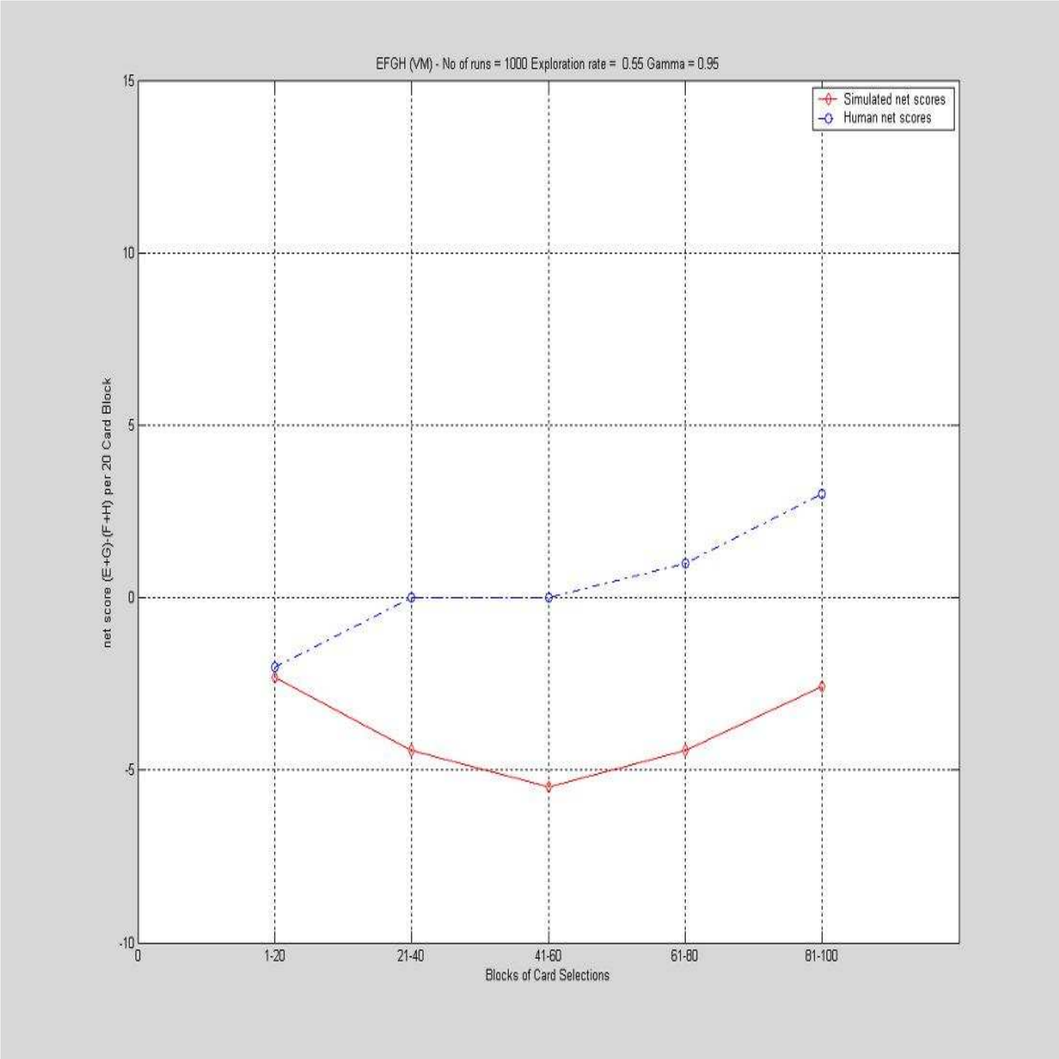


Figure 21: **Comparison of VMF patients' net scores (n=10) on the EFGH version of the IGT with the 0.55-greedy error-driven valuation (γ = 0.95) algorithm (n = 1,000).**

example where using *sample-averaging* for the valuation function is more effective than using the *error-driven* valuation function. This is because the total net score for VMF patients for the EFGH version is positive and *error-driven* with γ = 0.95, favours the regular rewards that are the least detrimental.

6.7 Normal controls on the E'F'G'H' version (see figure 22)

Here the simulated profile follows the general improvement shown in the human data, but performance does not dramatically improve in block 2 as it does in healthy humans. Blocks 1-3 are substantially different from the mean data points for normal humans and the simulated data, but blocks 4 and 5 are

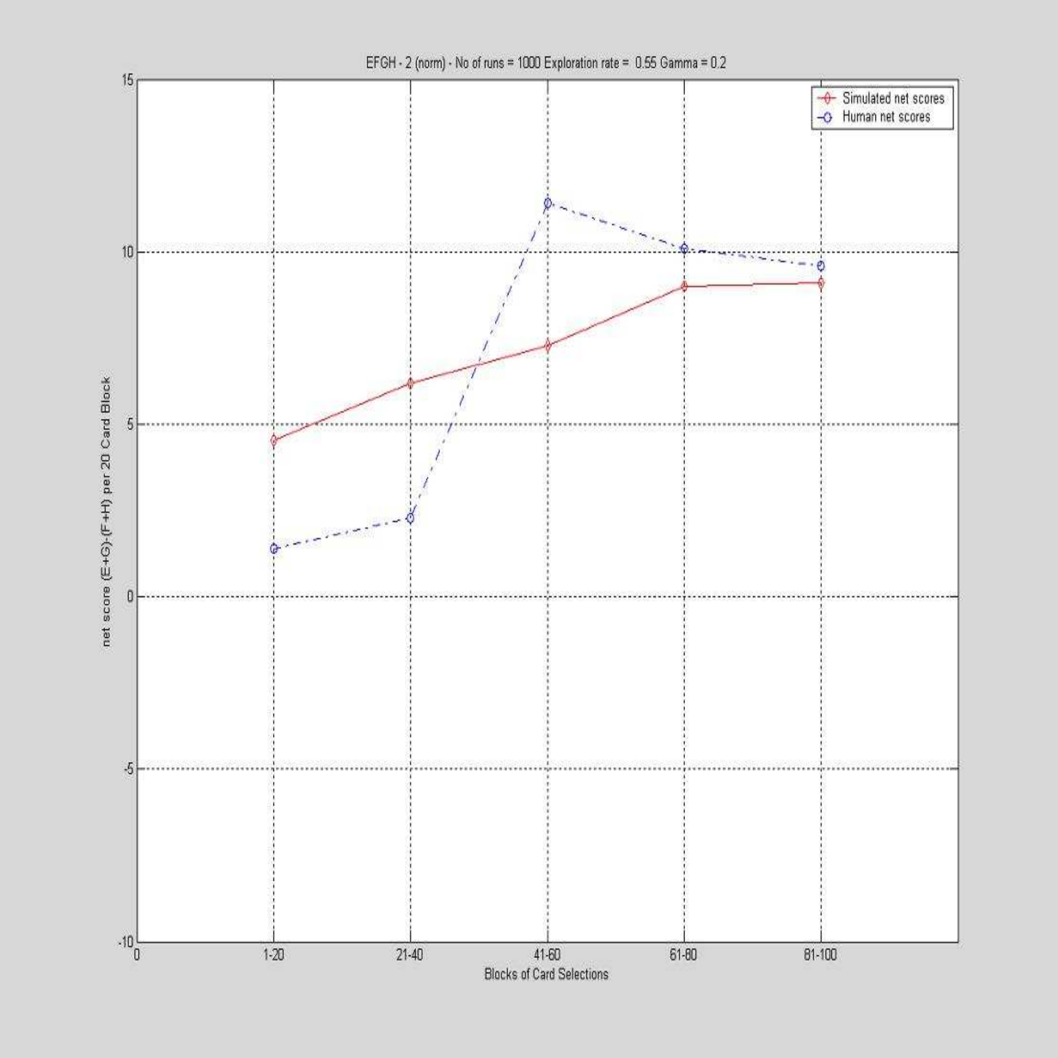


Figure 22: **Comparison of normal healthy controls' net scores (n=66) on the E'F'G'H' version of the IGT with the 0.55-greedy error-driven valuation (γ = 0.2) algorithm (n = 1,000).**

very similar.

6.8 VMF patients on the E'F'G'H' version (see figure 23)

In the final comparison, blocks 1, 2 and 5 are very similar, but data points in blocks 3 and 4 are somewhat dissimilar. The simulation misses the improvement in performance followed by a decline, shown in the human data (blocks 4 and 5). In the simulation there is a decline, followed by an improvement, so in a qualitative sense, some similarities are lost.

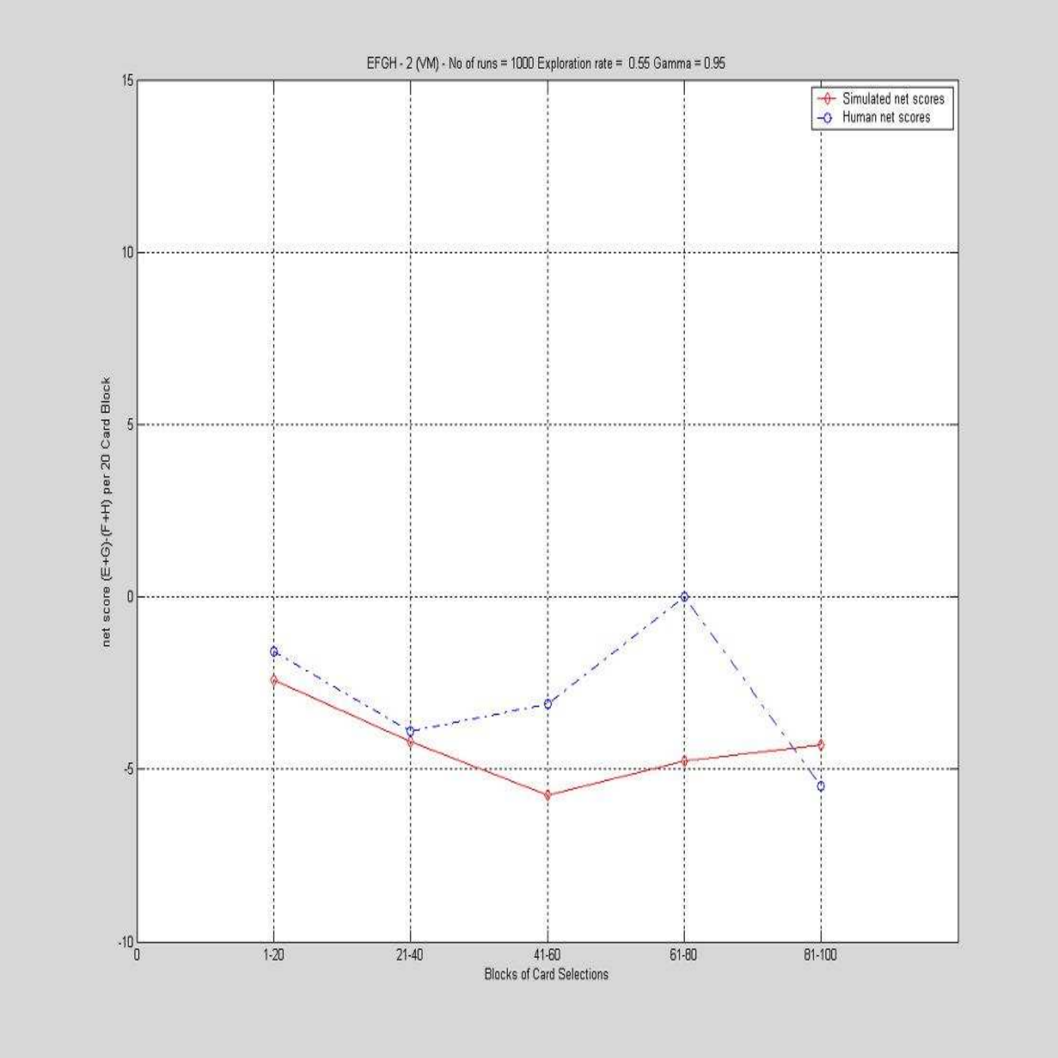


Figure 23: **Comparison of VMF patients' net scores (n=16) on the E'F'G'H' version of the IGT with the 0.55-greedy error-driven valuation (γ = 0.95) algorithm (n = 1,000).**

	Total Square error	sim to human/human VMF to Norm
Sim ABCD (γ = 0.2) - Human ABCD Norm	20	6
Sim ABCD (γ = 0.95) - Human ABCD VMF	21	6
Human ABCD Norm - Human ABCD VMF	354	N/A
Sim A'B'C'D' (γ = 0.2) - Human A'B'C'D' Norm	12	3
Sim A'B'C'D' (γ = 0.95) - Human A'B'C'D' VMF	20	5
Human A'B'C'D' Norm - Human A'B'C'D' VMF	405	N/A
Sim EFGH (γ = 0.2) - Human EFGH Norm	57	23
Sim EFGH (γ = 0.2) - Human EFGH VMF	110	45
Human EFGH Norm - Human EFGH VMF	245	N/A
Sim E'F'G'H' (γ = 0.2) - Human E'F'G'H' Norm	28	4
Sim E'F'G'H'(γ = 0.95) - Human E'F'G'H' VMF	22	3
Human E'F'G'H' Norm - Human E'F'G'H' VMF	649	N/A

Table 5: **This table compares the total square error (second column) between the relevant simulated data and the human data, and between the two human groups, to show how much more similar the relevant simulated data is to its target, than the human data sets are to one-another. The comparison is then shown as percentage ratio in the third column.**

6.9 Analysis of the *error-driven* ϵ-greedy methods and results

In the table 5, we present comparison of the total sum square errors between our simulated data sets and different human data sets. In addition, the table 5 shows the ratio of difference between the total sum square errors for normal human controls profiles and VMF profiles verses the simulated data and the human data the simulation is trying to replicate to provide another measure of the quality of fit. Table 5 shows that, in terms of square error, the simulated data is around 20 times closer to its' target data set, than the VMF human patient and NHC (Normal Healthy Controls) sets are from one-another. This suggests a fairly good match, but this is a relative measure, and is dependent on the differences between the two human data sets.

Overall, the *error-driven* ϵ-greedy model (γ = 0.95) suggests that, if there is a rapid update of information, a fairly consistent replication of the VMF human IGT data can be generated. This idea of VMF patients having a more rapid information update than healthy controls, is in conflict with the reversal learning literature, in that our model suggests normal human controls update their information less rapidly than VMF patients on the IGT. A simplistic deduction from our model is that VMF patients would be better at reversal learning tasks than healthy controls. This is generally disconfirmed in the literature, therefore suggesting that our model may be incorrect or not easily extendable to these simpler reversal learning paradigms. However, a simple reversal learning deficit does not seem to well describe the behaviour of Bechara et al.'s bilateral VMF patients, on the IGT, because they quickly change choice when they incur a loss (Bechara et al., 2005). Their problem is that they go back to the source of the loss (decks A and B) more rapidly than healthy controls. A quicker return to sources of loss compared to NHCs is replicated in Fellows and Farah (2005), where after the big loss in deck B (card 9) 8 out of 9 VMF patients stop selecting from deck B, and have a mean return to deck B after only 4 selections, compared to 10 selections for NHCs. This is more in line with the behaviour of the *error-driven* ϵ-greedy model (γ = 0.95) than a simple reversal learning deficit, based on perseveration, which would result in, immediate, repeated selection of the previously positive but now negative stimulus. It should be noted that the *error-driven* ϵ-greedy model with VMF patient settings (γ = 0.95) will not return to deck B, on average, any quicker than with normal controls settings (γ = 0.2), but would be more inclined to continue selecting from deck B once they had returned to that deck. This is because the speed of return to a deck that does not have the highest expectation of reward is only dictated by the exploration rate, which in the current set-up is the same for both the VMF patients and normal controls simulations.

The results of Fellows and Farah (2003, 2005)'s studies do suggest strongly that the main cause of deficits for Fellows and Farah's VMF patients on the IGT, are due to underlying reversal learning errors. On the simple reversal task the VMF patients deficit is perseverative, in that they continue to select from the previously rewarded stimulus for a few trials after the reversal. This suggests that the more complex IGT reveals a different type of deficit in VMF patients than straight forward perseveration. Below is a synopsis of the evidence Fellows and Farah present to support their reversal learning deficits theory.

1. During initial stimulus-reinforcement learning, VMFs make nearly twice as many errors as normal healthy controls in a simple reversal learning task (but this is not significant), and VMFs make

twice as many reversal errors as normal human controls (which is significant) (Fellows and Farah, 2003).

2. The number of reversal errors increases with lesion volume in the OFC, but not with total lesion volume (Fellows and Farah, 2003).

3. Improvement in performance on the ‘shuffled’ ABCD variant of the IGT, is correlated to severity of deficit in the reversal learning task. This is mainly due to a large improvement by 3 VMF patients. Two of which are among the lowest performers on the ‘unshuffled’ version, who then fall among the joint best performers (along with a 3rd participant originally, a good ‘unshuffled’ IGT performer, total net score $\approx$ 30) on the ‘shuffled’ version (Fellows and Farah, 2005).

4. Interestingly, reversal learning performance is correlated with day-to-day functional ability (IADL score). The worse their reversal learning performance, the greater their problems with day-to-day living (Fellows and Farah, 2003).

The current *error-driven* ϵ-greedy model presented here, does not seem to easily extend to the simple reversal learning task (Fellows and Farah, 2003). Interestingly, our *error-driven* ϵ-greedy model with the VMF settings, (ϵ = 0.55 and γ = 0.95) gives similar overall results to Fellows and Farah (2003) on the reversal task, but the errors do not necessarily occur at points of reversal as most errors do in the human VMF data. It also does not perform precisely like Fellows and Farah (2005)’s VMF patients in the original ‘unshuffled’ IGT, in that, the models performance is more dynamic, since net scores vary much more over the 5 blocks of the task. Although the simulation does match Fellows and Farah’s VMF patient data for blocks 3 and 4, and sits between Bechara et al’s VMF profiles and Fellows and Farah’s VMF profiles on blocks 4 and 5, the *error-driven* ϵ-greedy model more closely matches the data it originally tries to simulate, from Bechara et al.’s VMF patients performance on the original ‘unshuffled’ IGT. Fellows and Farah (2005)’s VMF patients, in both the ‘shuffled’ and ‘unshuffled’ versions of the ABCD IGT, barely change their behaviour from block 1. There is a slight gradual improvement in net scores over trials, from -2 in block 1 to +2 in block 5. Whereas in Bechara et al.’s studies, the VMF patients begin with a negative net score (-4.7) in block 1, rising to +2 in block 3, and then falling to -4.2 in block 5.

The model cannot replicate both Bechara et al.’s and Fellows and Farah’s, VMF patients’ net scores precisely on the original ‘unshuffled’ ABCD, which both groups have been tested on. This is because both groups act differently, under the same contingencies.

Interestingly, the normal controls in the Fellows and Farah (2005) study act differently from normals shown the in other studies (Bechara et al., 2000a,b, 1999; Clark et al., 2003; Bechara and Damasio, 2002a; Bechara et al., 2002b). They start with a positive, not a negative net score in block 1 (Fellows and Farah (2005) net score = 2 and Bechara et al. (2000a,b, 1999); Clark et al. (2003); Bechara and Damasio (2002a); Bechara et al. (2002b) net score = -5.9), where a negative score is actually advantageous (Maia and McClelland, 2004). In addition, in both the ‘shuffled’ and ‘unshuffled’ variants of the task, Fellows and Farah’s normal controls do not change their performance over blocks, apart from in block 5, on the ‘unshuffled’ version. However, in Bechara et al. (2000a,b, 1999); Clark et al. (2003); Bechara and Damasio (2002a); Bechara et al. (2002b) the normal controls increase their net scores over time, with the sharpest gradients between, blocks 1 and 2 and blocks 2 and 3, with the gradient attenuating across the final blocks. The reason for this difference, apart from simply sample variation, is not known, but the difference in performance in block 1, between normals across many studies (n=54) compared to the Fellows and Farah (2005)’s normals (n = 14) is highly significant ($P < 0.001$, t = 3.92), but is not across other blocks. In addition, only in block 5 on the ‘unshuffled’ version are the normals from Fellows and Farah (2005) significantly different from a random agent ($P < 0.05$, t = 2.25). The net scores of the normal controls also improves, when the reversal is removed in the ‘shuffled’ version, but not as much as the VMF patients. Although, if the 3 best improvers on the ‘shuffled’ version are excluded, the improvement in normal controls and the other VMF patients is similar.

When the *error-driven* ϵ-greedy model (ϵ = 0.55 and γ = 0.95) is applied to the ‘shuffled’ version of the ABCD IGT, then it does not act as Fellows and Farah (2005)’s VMF patients did. It begins with a positive net score (+4) in block 1, but is quickly influenced into selecting from the disadvantageous decks. The model is most strongly influenced by the regular wins or losses, rather than uncertain events, as these are quickly overridden by new results. This is consistent with rodent OFC lesions studies, where lesions cause increased preference for small immediate rewards over larger delayed rewards, and increased preference for small certain reinforcers over larger uncertain ones (Mobini et al., 2002). Therefore, *error-driven* ϵ-greedy model (ϵ = 0.55 and γ = 0.95) does not replicate Fellows and Farah (2005)’s VMF profiles, on either the ‘shuffled’ or ‘unshuffled’ variants of the ABCD IGT. The reasons are not completely evident, although, as suggested earlier, Fellows and Farah's patients with the worst reversal learning deficits, have a focus of lesions in the left posteromedial orbitofrontal cortex. This is in contrast with Bechara et al.'s patients who have a slightly greater focus on the right hand side of the orbitofrontal cortex. The authors suggest that these two groups of VMF patients are, possibly, different in their focus of lesions and, certainly, in their IGT performance. In both the 'shuffled' and 'unshuffled' versions, Fellows and Farah (2005)'s VMFs seem to follow their initial 'inclination' (block 1 net score) throughout the task, with a slight increase in their net scores (increase of 4 across all blocks in the 'unshuffled' version and between 4 and 5 in the 'shuffled' version). Bechara et al.'s VMF patients have a more dynamic profile, during the first 3 blocks they have profiles similar to normals and then begin to select disadvantageously again in block 4, with block 1 and block 5, and block 2 and block 4 being more or less symmetrical with one-another in terms of net scores.

7 Discussion

7.1 Comparing the two valuation functions: *error-driven* and *sample-averaging*

Sample-averaging gives a slightly better match than the *error-driven* valuation function does for normal healthy controls' profiles on the IGT. This is, especially, evident for the EFGH variant, and less so for the E'F'G'H' variant of the task. But apart from the EFGH variant, where it is superior to the *error-driven* learning, *sample-averaging* is poor at reproducing Bechara et al.'s VMF patients' profiles, as it cannot, regularly, select disadvantageously. In contrast, the *error-driven* valuation function gives a fairly good match for all Bechara et al.'s IGT variants, for both normals and VMFs, apart from on the EFGH variant. This shows that normal controls can be described using a number of different possible valuation functions, and VMFs can be well described using a strongly updating *error-driven* valuation function. The *error-driven* learning ϵ-greedy model seems to describe VMF behaviour well on the IGT, across a number of studies, and could suggest reasons for their more general behaviour deficits, where VMF patients are often whimsical, and lack the ability to pursue long and medium term goals through coherent strategy. VMF patients are easily distracted from on-going tasks, and can need prompting to perform basic maintenance tasks like cleaning themselves. Our *error-driven* ϵ-greedy model would predict this sort of behaviour, in that, unless new information is provided, the task is continued, but current strategies are easily influenced or ignored due to new information. Unlike normals, who are less influenced by new information, and take more previous data into account when making decisions about future action, VMF patients, we propose, are strongly biased by recent information. These ideas could extend to higher level goals, as seems to be evident in real life VMF behaviour, where VMFs often have many ideas or plans that never come to fruition (old ideas have limited influence on current decision making). The *error-driven* ϵ-greedy model would also suggest that multi-tasking is difficult for VMF patients, as external influences are likely to override intentions to perform the different on-line tasks, as high lighted in a case study (Wood and N.A., 2004). Following on from this, patients would be expected to perform better in situations where the contingencies are not rapidly changing, otherwise they would be too influenced by each new piece of information, rather than the general trend. It may be a cause of their social deficits, as information is rapidly being updated in social situations, and so may pose a serious processing problem for VMFs, since they would overly react to small information changes.

7.2 Discussions of Hypotheses

Four hypotheses have been set out in this paper.

1. VMF patients have 'no preferences' (Sloman, 2004).
2. VMF patients are 'risk-seeking' (Sanfey et al., 2003).
3. VMF patients have a reversal learning deficit (Fellows and Farah, 2003, 2005).
4. VMF patients have a 'myopia' for future consequences (Bechara et al., 2000a)

7.2.1 VMF patients have no preferences

The 'no preference' theory is supported by Fellows and Farah (2005)'s VMF patients on the 'unshuffled' ABCD IGT, as they do not act significantly differently from a random agent, apart from in block 1. However, on examining Fellows and Farah's VMF patients on the 'shuffled' ABCD version it seems they are acting with a preference for the advantageous decks, but this is not significantly different from a random agent, due to the small sample (n = 9) and the large standard errors of the mean. However, this 'preference' for the advantageous cards on the 'shuffled' ABCD version is not significantly different from the normal controls in this study (Fellows and Farah, 2005). Thus, although not significantly different from random, it seems that, on a qualitative and normative level, these VMF patients have some preference, which seems dependent on their initial estimations (block 1), as they improve very little after block 1. If only the statistical interpretation of not being significantly different from random is used, then the ‘no preference’ theory has validity for Fellows and Farah’s VMF patients. However, from the normative approach, Fellows and Farah’s VMF patients are not significantly different from their normal healthy controls on the ‘shuffled’ ABCD IGT. In addition, the data from Bechara et al.’s VMF patients is different, in that, selections in blocks 1, 4 and 5 are significantly different from a random agent, and significantly different from Fellows and Farah (2005) VMF patients in block 5. Therefore, Bechara et al.’s data weakens the ‘no preference’ theory, both statistically and qualitatively. When one inspects the changes in the curve, over the blocks in the ‘unshuffled’ ABCD IGT, there is a fairly consistent pattern over studies, with a distinctive non-random curve. This original ‘no preference’ hypothesis presented in Sloman (2004), probably comes from anecdotal evidence, where bilateral orbitofrontal lesions patients have stated they have “no preferences” (Sacks, 1998). As this study is particularly interested in Bechara et al.’s VMF patients, which have featured more heavily in the literature of the last decade, the ‘no preference’ theory does not seem to hold with these patients (apart from on the EFGH variant (Bechara et al., 2000a)), but it seems to have some validity with the patients in Fellows and Farah (2005).

7.2.2 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 human VMF patient data, where their selections are considerably more disadvantageous than normals human controls. In a weak sense, if risk-seeking were a small factor in VMF patients’ IGT deficit, then one would expect them to select more from the advantageous decks in the E(’)F(’)G(’)H(’) variants than normals compared to their corresponding A(’)B(’)C(’)D versions. This is because, in the E(’)F(’)G(’)H(’) versions, E(’) and G(’) are the best decks, and have higher variances than the bad decks, F(’) and H(’). However, in the A(’)B(’)C(’)D versions, the bad decks A(’) and B(’) have the highest variance. This weak argument is supported by VMF patients’ profiles as for the EFGH variant (total net score = 2) they have a higher overall net score than on the ABCD version (total net score = -9.2), and again for E’F’G’H’ version (total net score = -14.1), compared to the A’B’C’D’ version (total net score = -17.3).

A corollary of this, is that, the net scores on the computerised version, E’F’G’H’, should be higher, than for the non-computerised versions, EFGH, because the variance in the good decks, E’ and G’, increases over trials, but is constant for, E and G. In comparison, one would expect the net scores to be lower for VMF patients on the A’B’C’D’ version than on the ABCD version, as the variance in A’ and B’ increases, and decreases in C’ and D’, but is constant for the ABCD versions. This is the case for the A(’)B(’)C(’)D versions, but not for the E(’)F(’)G(’)H(’) versions, where EFGH has a higher total net score than for the E’F’G’H’ version. A reason for this discrepancy could be due to the E’F’G’H’ variant being more complex than the EFGH version, as the contingencies change throughout the task. However, this confound of task complexity does not effect the predictions for the A(’)B(’)C(’)D versions, where increasing variance in A’ and B’, increases their overall selection, and so, is not consistent with the data from these two versions. Unless, we assume an increased cognitive load, caused by changing contingencies, which leads to a greater influence of regular rewards. If this assumption is accepted, then the human data supports a cognitive load theory.

Therefore, when considering the profiles of VMF patients on the different IGT variants the evidence supporting a risk-seeking theory is weak, and even with the addition of a task complexity proviso the theory is only slightly strengthened. This theory in terms of the IGT, is only well supported if the A(’)B(’)C(’)D(’) versions are considered alone, as VMF patients in these versions pick more from the high variance decks, A(’) and B(’), than normal human controls.

The evidence better supports other theories, like VMF patients being more influenced by the regular rewards and punishments, than the uncertain ones (’myopia’ for future consequences theory). Under this theory patients would select more from, A(’) and B(’), in the A(’)B(’)C(’)D(’) versions, and more from, G(’) and H(’), in the E(’)F(’)G(’)H(’) versions. The ’myopia’ for future consequences theory is therefore supported by most of Bechara et al.’s data, except, the data produced using the EFGH variant.

Overall, there is little supporting evidence for the risk-seeking being the main cause of VMF patients IGT deficit.

Sanfey et al. (2003) found that their were two subgroups within their VMF population, one risk-VMF and safe-VMF, this could be the case in Bechara et al’s VMF population, but this is not obvious from the data. Therefore, with the present data, this interpretation does not seem to be the main cause for Bechara et al’s VMF patients’ deficits on the IGT.

7.2.3 VMF patients have a reversal deficit as the cause of VMF IGT deficits

Fellows and Farah (2005)'s VMF patients suggest that reversal learning deficits are the main cause of their disadvantageous behaviour in the IGT, at least for 3 of them. However, neither set of VMF patients on the IGT have the same type of deficit evident in the simple reversal task (Fellows and Farah, 2003), in that, on the IGT they do not perseverate after the large reversal, which happens early in the task. It is actually their rapid return to the bad decks that is their deficit (Bechara et al., 2005), not immediate perseveration. In addition, one might suspect that the earlier the occurrence of the reversal, the less influential it would be on future behaviour, since initial contingencies are less established. The data suggests that an earlier reversal is less detrimental to net scores, that is, Bechara et al.'s VMF patients act less disadvantageously on the E'F'G'H' version (total net score = -14.1), than on the A'B'C'D' version (total net score = -17.3) of the task. The EFGH version (total net score = 2) results are also supportive of the reversal learning deficit theory, as VMF patients perform better on this task than on the ABCD task (total net score = -9.2). There are some problems with this interpretation, as it could be confounded by risk-seeking behaviour, because, as mentioned in the risk-seeking section, E and G are the best decks and the most risky (highest variance). However, overall, the earlier reversal in the E(')F(')G(')H(') versions clearly correlates with increases in net scores compared to corresponding A(')B(')C(')D(') versions. The only problem with this interpretation is that there is not an obvious reason why, in the E'F'G'H' version, VMF patients perform much worse than on the EFGH variant, when the positions of the reversals are the same in both IGT versions. To account for net scores being much lower in the E'F'G'H' version compared to the EFGH version, one might look to an increased cognitive load argument. E'F'G'H' might be considered to be more complex, and so, create more cognitive load, due to the continually shifting contingencies. Therefore, one might argue that reversal learning has a very high cognitive load as stimulus-reinforcement associations change considerably, and that reduced contingency changes would have a lesser affect on performance. This seems to be the case as VMF patients are worse at the computerised versions ((') where the contingencies change throughout the task), than they are at the card based versions (see table 5). Reasons for the better performance by VMFs in the E(')F(')G(')H(') variants, compared to the A(')B(')C(')D(') versions, are not so easily definable and could be due to influence from risk-seeking, an earlier reversal, or just a high learning rate in an error-driven system working with different contingencies.

It should be made clear that 3 of the 9 VMF patients in Fellows and Farah (2005)'s study substantially improved their performance when tested on the 'shuffled' ABCD IGT, so much so that, if removed from the study, improvements in normals and VMF patients is about the same. This might be problematic in a statistical sense, as these patients probably drive Fellows and Farah (2003)'s correlation between OFC lesion size and reversal errors, and the further correlation between reversal errors and improvement in the 'shuffled' version compared to the 'unshuffled' version (Fellows and Farah, 2005). This all seems to follow a stimulus-reinforcement associations argument for the role of the OFC, with damage reducing that ability. On the simple reversal task (Fellows and Farah, 2003), VMF patients perseverate and could be described as slow to adjust to the changing contingencies. However, on the IGT, VMF patients are as quick as NHCs to stop selection from deck B after a large loss, 8 out of 9 VMF patients in Fellows and Farah (2005). Therefore, in the IGT, a large loss is sufficient to alter behaviour, but the reversal in the simple reversal task is often not sufficient to quickly shift behaviour. This suggests that if the *error-driven* ϵ-greedy model were used to simulate humans on the simple reversal learning task (Fellows and Farah, 2003), the VMF patient model would have a lower learning rate than the version of the model used to describe normal controls.

Another theory is that initial stimulus associations have a special place and are less dependent on the OFC than the management of changing associations. This might help clarify the large improvement in the three good VMF patient performers on the 'shuffled' version of the IGT. Using *sample-averaging*, with $\epsilon = 0.9$, provides a good simulation of the VMF patient data for the 'unshuffled' ABCD version, while setting $\epsilon = 0.7$, gives a good match to the 'shuffled' version VMF data. Importantly, for Fellows and Farah (2005)'s VMF patients' profiles, *sample-averaging* emphasizes the value of initial stimulus-reinforcement associations more than *error-driven* learning. This is because it accurately evaluates the actually value of the first item. This is similar to *error-driven* learning with a high learning rate, although then information would be quickly overwritten. However, in the first few selections, *sample-averaging* can be quite different from *error-driven* with a low learning rate, whereas initial stimulus will not be learnt unless it is repeatedly received. Therefore, one needs to examine the differences between these models. It is evident that removing the reversal in the IGT improves performance in both NHCs and VMF patients, particularly those with larger OFC lesions. We suggest that those with larger OFC lesions are more likely to be influenced by initial stimulus-reinforcement associations, as well as having the equivalent of a high learning rate, due to their reduced ventromedial prefrontal cortex capacity. Initial stimulus-reinforcement is preserved, as it is mediated, at least to a degree, by a different neural substrate, in the temporal cortex (Daum et al., 1991). Although a *sample-averaging* valuation function better matches Fellows and Farah's data, the increased influence of initial stimulus-reinforcement associations is not sufficient to explain why two different exploration rates are required to model the 'shuffled' and 'unshuffled' version of the IGT Fellows and Farah (2005). Therefore, the moving of the reversal in the 'shuffled' version, must have a greater affect on behaviour than described by *sample-averaging*

One of the important aspects we take from our *error-driven* learning valuation function, is that VMF patients can only utilise a limited set of past information, as we have previously suggested in another model (Kalidindi et al., 2005b). A further note to be considered about Fellows and Farah (2005)'s data to be considered is that only 3 of her VMF patients actually fall into the category of selecting disadvantageously (total net score $<$ 0) in the 'unshuffled' ABCD IGT, only one of which is among the 3 big improvers. This is unlike Bechara et al.'s VMF patients who all selected disadvantageously. In terms of our models, we would suggest that a number of Fellows and Farah's VMF patients have a lower learning rate than modelled in this paper. It is difficult to take these ideas further without individual data sets, which would certainly be useful in further simulations and analysis.

Reversal learning deficits could be influential on IGT performance, particularly with certain VMF patient groups (Fellows and Farah, 2005), but it does not in its simple interpretation help us decipher the actual behaviour displayed by Bechara et al.'s VMF patients on the IGT.

7.2.4 VMF patients have a 'myopia' for future consequences

This hypothesis, suggested by Bechara et al. (2000a), is based directly on the VMF patient group who are the main focus of consideration in this paper. A 'myopia' for future consequences theory would predict that patients would select more from the decks which give the best regular returns. Decks A(') and B(') give the best regular returns for the A(')B(')C(')D(') versions, while decks F(') and H(') provide the least losses (best regular returns) on the E(')F(')G(')H(') variants. The IGT is set up so that the decks with the best regular returns are actually the worst decks. Bechara et al.'s VMF patients do generally follow this pattern of selecting more from disadvantageous decks, apart from on the EFGH version (total net score = 2). In addition, with this theory, using the *error-driven* ϵ-greedy model with set learning parameters for all IGT versions, one would expect VMFs on the ABCD and EFGH versions to have similar or lower net scores than on the corresponding A'B'C'D' and E'F'G'H' versions. The differences in net scores would be especially evident in later blocks, when good decks get better and bad decks get worse. The human data does not support this prediction, with net scores being worse in the A'B'C'D' and E'F'G'H' versions. One would expect normal healthy controls (NHC) to perform better on the A'B'C'D' and E'F'G'H' versions than on the corresponding ABCD and EFGH versions, as differences between good and bad decks are made wider in the A'B'C'D' and E'F'G'H' versions. But NHCs perform very similarly on the corresponding computerised vs. the original card based, versions of the task. Therefore, we suggest normals can manage the increased task complexity from changing contingencies, and the decreased task complexity provided by increased separation of mean rewards, between advantageous and disadvantageous decks. Interestingly, even more accurate human profiles can be simulated by increasing the learning rates (from the presented learning rates) for the normal subject simulations on the computerized versions, and by reducing the learning rate for the card based versions. This is also the case for simulations of the VMF patient profiles, which is suggestive of changes in cognitive load.

However, in VMF patients, the increased complexity of the computerised versions, A'B'C'D' (total net score = -17.3) and E'F'G'H' (total net score = -14.1), is not managed so well, as they have reduced net scores for these versions compared to the ABCD (total net score = -9.2) and EFGH (total net score = 2) versions. We suggest the increased complexity in changing contingencies raises the cognitive load on VMF patients, pushing their cognitive limits. This limitation in cognitive ability would correlate with an increase in the learning parameter, reducing the influence of temporally distant information on current decision making. This might suggest that the EFGH variant is actually easier than the other versions of the IGT, requiring less cognitive resources, reflected by a lower learning rate to reproduce the VMF patient data. This is supported by the EFGH version having, a positive risk-seeking orientation (best decks have the highest variance) in terms of net scores; an early reversal compared to the two A(')B(')C(')D(') versions and a lower cognitive load than the E'F'G'H' variant.

In general the 'myopia' for future consequences hypothesis is supported by much of the data. The *error-driven* ϵ-greedy model with fixed parameters for the two patient groups recreates nearly all the data from the different IGT versions. However, when slightly different learning rates are used for the different versions of the task, an even more accurate set of profile replications can be produced. This does reduce some of the simplicity of the model presented so far, which gave a very parsimonious account of the differences between normal controls and bilateral VMF patients. Such simple models are rarely devoid of inaccuracies in the details, but can still prove to be useful accounts of complex real world behaviour.

8 Conclusion

Of the four theories, the 'myopia' for future consequences best describes the behaviour of Bechara et al's VMF patients on the IGT, compared to normals. Our simple *error-driven* ϵ-greedy model also supports the 'myopia' hypothesis, as increasing the learning rate from $\gamma = 0.2$ (normals) to $\gamma = 0.95$ (VMFs) increases the influence of recent information. We suggest that 'myopia' for future consequences is caused by the limited use of past information in present decision making. Therefore, the contribution of this work is in presenting more support for a present theory, but also an extension to the causes of 'myopia' for future consequences, in both a quantitative and qualitative sense. In the quantitative sense, the work offers a simple numerical model, which can easily be tested on the IGT and possibly other paradigms, and produces a good approximation to the human data sets, giving us insight into how the information in the task might be processed. In the qualitative sense, as mentioned earlier, it extends the 'myopia' for future consequences theory, by suggesting that it is caused by VMF patients being too influenced by recent and current information, and not giving enough credence to, older, past events.

Currently, based on the evidence and models presented in this paper, Bechara et al.'s patients suffer from an over-zealous learning rate in their adjustment to rewards and punishment, particularly on the Iowa Gambling Task, but this may be extendable to a much wider set of behavioural deficits observed in VMF patients. An interesting final thought is that our model and interpretation would suggest that VMF patients would probably benefit from routine, as then unexpected events would be less likely, and so less influential on behaviour. This would also be the treatment suggested by a reversal learning deficit theory, as routine would reduce the requirement for reversals in stimulus-reinforcement associations. Perhaps, a theory of cognitive load could incorporates both of these ideas, without contradiction.

References

N. Alderman, P.W. Burgess, C. Knight, and C. Henman. Ecological validity of a simplified version of the multiple errand shopping test. *Journal of International Neuropsychological Society*, 9:31–44, 2003.

J. Barrash, D. Tranel, and S.W. Anderson. Acquired personality disturbacnes associated with bilateral damage to the ventromedial prefrontal region. *Developmental Neuropsychology*, 18:355–381, 2000.

A. Bechara. The role of emotion in decision-making: Evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, 55:30–40, 2004.

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

A. Bechara and H. Damasio. 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, 2002a.

A. Bechara, H. Damasio, and A.R. Damasio. Emotion, decison making and the orbitofrontal cortex. *Cerebral Cortex*, 10:295–307, 2000b.

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

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

A. Bechara, H. Damasio, D. Tranel, and A.R. Damasio. Deciding advantageously before knowing the advantageous strategy. *Science*, 275(5304):1293–1295, 28 February 1997.

A. Bechara, H. Damasio, D. Tranel, and A.R Damasio. The iowa gambling task and the somatic marker hypothesis: some questions and answers. *Trends in Cognitive Science*, 9(4):159–162, April 2005.

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

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

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

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

R. Cools, R.A. Barker, B.J. Sahakian, and T.W. Robbins. Enhanced and impaired cognitive function in parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral Cortex*, 11:1136–1143, 2001.

R. Cools, L. Clark, A.M. Owen, and Robbins T.W. Defining the neural mechanisms of probablisitic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience*, 22(11):4563–4567, 2002.

E.A. Crone, R.J.M. Somsen, B. Van Beek, and M. Van Der Molen. Heart rate and skin conductance analysis of antecedents and consequences of decision making. *Psychophysiology*, 41:531–540, 2004.

A.R. Damasio. *Descartes Error: Emotion, Reason and the Human Brain*. New York: Grosset/Putnam, 1994.

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

R. Dias, T.W. Robbins, and A.C. Roberts. Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, 380:69–72, 1996.

J.J. Downes, A.C. Roberts, B.J. Sahakian, J.L. Evenden, R.G. Morris, and Robbins T.W. Impaired extra-dimensional shift performance in medicated and unmedicated parkinson's disease: Evidence for a specific attentional dysfunction. *Neuropsychologia*, 27:1329–1343, 1989.

C.E.Y. Evans, K. Kemish, and O.H. Turnbull. Paradoxical effects of education on the iowa gambling task. *Brain and Cognition*, 54:240–244, 2004.

L.K. Fellows. The cognitive neuroscience of human decison making: a review and conceptual framework. *Behavioural and Cognitive Neuroscience Reviews*, 2(10):1–14, 2005.

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

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

J.J. Gallo, T. Fulmer, G.J. Paveza, and W. Reichel. *Handbook of geriatric assessment*. Gaithersburg (MD):Aspen Publications, 3rd edition, 2000.

S. Grant, C. Contoreggi, and E.D. London. Drug abusers show impaired performance in a laboratory test of decision making. *Neuropsychologia*, 38:1180–1187, 2000.

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

D. Kahneman, P. Slovic, and A. Tversky, editors. *Judgment under uncertainty: heuristics and biases*. Cambridge; New York: Cambridge University Press, 1982. ISBN 0521240646.

K. Kalidindi, Bowman H., and Wyble B. A consideration of decision-making, motivation and emotions within dual process theory: supporting evidence from somatic-marker theory. In Dylan Evans and Lola Canamero, editors, *Proceedings of Symposium on Agents that Want and Like: Motivational and Emotional Roots of Cognition and Action - SSAISB 2005 Convention*, pages 51–54. www.aisb.org.uk, The Society for the Study of Artificial Intelligence and Simulation of Behaviour, 2005a.

K. Kalidindi, Bowman H., and Wyble B. An investigation of the myopia for future consequences theory of vmf patient behaviour on the iowa gambling task; an abstract neural network simulation. In Guido Bugmann Angelo Cangelosi and Roman Borisyuk, editors, *Proceedings of the Neural Computation and Psychology Workshop 9*, 2005b. In press.

M.L. Kringelbach, Rolls E.T. O'Doherty, J., and C. Andrews. Activation of the human orbitofrontal cortex to a liqiud food stimulus is correlated with its subjective pleasantness. *Cerebral Cortex*, 13: 1064–1071, 2003.

T.V. Maia and J.L. McClelland. 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, November 2004.

F. Manes, B. Sahakian, L. Clark, R. Rogers, N. Antoun, M. Aitken, and T. Robbins. Decision-making processes following damage to the prefrontal cortex. *Brain*, 125:624–639, 2002.

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

J. O'Doherty, M.L. Kringelback, Rolls E.T., J. Hornak, and C. Andrews. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, 4:95–102, 2001.

J. O'Doherty, E.T. Rolls, S. Francis, R. Bowtell, F. McGlone, and et al. Kobal G. Sensory-specific satiety-related olfactory activation of the human orbitofrontal cortex. *Neuroreport*, 11:399–403, 2000.

R.D. Rogers, T.C. Andrews, P.M. Grasby, D.J. Brooks, and T.W. Robbins. Contrasting cotical and subcortical activations produced by attentional set-shifting and reversal learning in humans. *Journal of Cognitive Neuroscience*, 12:142–162, 2000.

R.D. Rogers, B.J. Everitt, A. Baldacchino, A.J. Blackshaw, R. Swainson, K. Wynne, N.N. Baker, J. Hunter, T. Carthy, E. Booker, M. London, J.F. Deakin, B.J. Sahakian, and T.W. Robbins. 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, 1999.

E.T. Rolls. *The Brain and Emotion*. Oxford University Press, 1999.

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

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

E.T. Rolls, H.D. Critchley, R. Mason, and E.A. Wakeman. Emotion related learning in patients with social and emotional changes associated with frontal lobe damage. *Journal of Neurology, Neurosurgery and Psychiatry*, 57:1518–1524, 1994.

O. Sacks. *The man who mistook his wife for a hat.* Touchstone, 1998. ISBN 0684853949.

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

W.A. Schmitt, Brinkley C.A., and Newman J.P. Testing damasio's somatic marker hypothesis with psychopathic individuals: risk takers or risk averse. *Journal of Abnormal Psychology*, 108(3):538–543, 1999.

T. Shallice and P.W. Burgess. Deficits in strategy application following frontal lobe damage in man. *Brain*, 114:727–741, 1991.

B. Shurman, W.P. Horan, and K.H. Nuechterlein. Schizophrenia patients demonstrate a distinctive pattern of decision-making impairment in the iowa gambling task. *Schizophrenia Research*, 72:215–224, 2005.

A. Sloman. What are emotion theories about? Invited talk at cross-disciplinary workshop on Architectures for Modeling Emotion at the AAAI Spring Symposium at Stanford University in March 2004, March 2004. See - http://www.cs.bham.ac.uk/research/cogaff/sloman-aaai04-emotions.pdf.

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

R.S. Sutton and A.G. Barto. *Reinforcement Learning: an introduction.* A Bradford Book. The MIT Press, 1998. ISBN 0-262-19398-1.

A. Suzuki, A. Hirota, N. Takasawa, and K. Shigemasu. Application of the somatic marker hypothesis to individual differences in decision making. *Biological Psychology*, 65:81–88, 2003.

I. Tomb, M. Hauser, P. Deldin, and A. Caramazza. Do somatic markers mediate decisions on the gambling task? *Nature Neuroscience*, 5(11):1103–1104, 2002.

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

R.L. Wood and Rutterford N.A. Relationships between measured cognitive ability and reported psychosocial activity after bilateral frontal lobe injury: An 18-year follow-up. *Neuropsychological Rehabilitation*, 14(3):329–350, 2004.