Available online at www.sciencedirect.com

ScienceDirect

Neural Networks 20 (2007) 676–689

Neural Networks

www.elsevier.com/locate/neunet

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

Kiran Kalidindi*, Howard Bowman

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

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

Abstract

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

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

1. Introduction

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

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

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

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

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

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

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

2. Versions of the Iowa Gambling Task (IGT)

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

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

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

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

3. Competing theories of VMF patient deficits

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

3.1. *'No preferences' — A. Sloman*

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

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