Biologically Inspired Cognitive Architectures (2014) 10, 10–16

Available at www.sciencedirect.com
ScienceDirect
journal homepage: www.elsevier.com/locate/bica

Application of a neural network model of prefrontal cortex to emulate human probability matching behavior

Suhas E. Chelian [a,*], Ryan M. Uhlenbrock [a], Seth Herd [b], Rajan Bhattacharyya [a]

[a] HRL Laboratories, LLC, Malibu, USA
[b] e-Cortex, Boulder, USA

Received 7 November 2014; accepted 7 November 2014

KEYWORDS
Probability matching;
Prefrontal cortex;
n-arm bandit

Abstract
Probability matching behavior occurs in a variety of decision-making domains that can be mapped to the *n*-arm bandit problem. Prefrontal cortex has been implicated in executive control over several tasks including the *n*-arm bandit problem. Previously the Prefrontal cortex Basal Ganglia Working Memory (PBWM) model has been used to replicate other decision-making functions of prefrontal cortex such as recognizing sequences of symbols or visual scenes. In this work, we emulate probability matching behavior from human subjects using the PBWM model in *n*-arm bandit-like problems. Possible extensions to the current work such as including other biases like loss aversion and misperception of both large gains and losses are also discussed.

Introduction

The *n*-arm bandit problem—a class of problems where one must repeatedly chose among several alternatives with unknown and possibly dynamic payoffs—arises in several psychological and technological domains (see Lee, Zhang, Munro, & Steyvers, 2011 for a review). The Bayesian optimal solution is to always pick the option with the highest expected payoff. However, humans often choose options in proportion to the expected payoff of each alternative; this is known as probability matching. A number of neuroscience studies have implicated several regions of prefrontal cortex in the *n*-arm bandit task in rats (Sul, Kim, Huh, Lee, & Jung, 2010), monkeys (Walton, Behrens, Buckley, Rudebeck, & Rushworth, 2010) and humans (Wunderlich, Rangel, & O'Doherty, 2009). A recent model of prefrontal

* Corresponding author.
E-mail addresses: sechelian@hrl.com (S.E. Chelian), rmuhlenbrock@hrl.com (R.M. Uhlenbrock), seth.herd@e-cortex.com (S. Herd), rbhattac@hrl.com (R. Bhattacharyya).

http://dx.doi.org/10.1016/j.bica.2014.11.002
2212-683X/

cortex, Prefrontal cortex Basal Ganglia Working Memory, or PBWM, has been used to recognize sequences of symbols (O'Reilly and Frank, 2006) and perform visual scene recognition (Chelian, Bhattacharyya, & O'Reilly, 2011). Here, we adapted PBWM networks to *n*-arm bandit tasks derived from a geospatial intelligence setting where one must choose to defend (e.g., arm 1) or not defend (e.g., arm 2) against an opponent. Greater degrees of conservatism with greater probability matching bias make agents pick options closer to an even distribution than the rational winner-take-all distribution of decisions. E.g., a conservative agent might select the action with the highest expected payoff (e.g., not defend) 60% of the time versus 40% of the time for the other action (e.g., defend). Conversely, lesser degrees of conservatism and probability matching bias correspond to less distance from the optimal distribution of choices; i.e., select the action with the highest expected payoff (e.g., not defend) 90% of the time. These varying degrees of conservatism were found in human data and modeled with PBWM networks.

Materials and methods

We summarize the tasks here but full details can be found in the MITRE Technical Report (in preparation). The tasks were adversarial games set in a geospatial intelligence context with two players, blue and red. Blue was controlled by a human or a neurocognitive model agent while red was a computer opponent. Blue agents received information through various sources of intelligence, or INTs, about red's potential actions. In each trial, blue was informed of the strategic utility (*U*) of a potential attack location and the probability (*P*) of winning a showdown there. Red chose to attack based on *P* and *U* and blue estimated the probability of red attack (*Pa*) with the INTs. Given, *Pa*, *P*, and *U*, blue decided to divert or not divert against a potential attack; we refer to this decision as $D/{\sim}D$. In the event that blue did not divert (or did divert) a potential attack and red did attack (or did not attack), no points are lost for either side. If blue diverted and red did not attack, blue has unnecessarily committed resources and loses a small amount of points. If blue did not divert and red attacked, the winner was decided probabilistically using *P* and the winner was awarded *U* points. This is summarized in the payoff Table 1: *U* was 2, 3, 4, or 5. *P* was a real value between 0 and 0.5, and *Pa* was a real value between 0 and 1.

Table 1 Payoff table of the *n*-arm bandit-like tasks.

	Red attacked	Red did not attack
Blue diverted	0	−1
Blue did not divert	±U	0

The optimal strategy is to take the action with the highest expected payoff. From a rational basis, the decision to divert can be decided using the inequality: $-1 + Pa > U \cdot Pa \cdot (2 \cdot P - 1)$. This inequality defines a decision boundary in *Pa*, *P*, *U* space which is illustrated in Fig. 1 for $U = 2$ and $U = 5$. As *Pa* increases along the horizontal axis (or *P* decreases along the vertical axis), red is more likely to attack (or less likely to win a showdown) and hence blue should divert. As *U* increases from Fig. 1a to b at the same *Pa*, *P* point, potential losses increase so blue should also divert. In short, for points below (or on or above) the curve, blue should divert (or not divert) a potential attack.

Blue agents played 5 variations or missions of the geospatial intelligence task. In the first mission, blue practiced estimating *Pa* given *P* and *U* and did not have to make a divert/not divert ($D/{\sim}D$) decision. In missions 2 through 5, blue agents made the $D/{\sim}D$ decision with: mission 2, a basic red opponent; mission 3, a red opponent who could attack in one of two locations but not both; and missions 4 and 5, a red opponent who could vary his strategy in *P*, *U* space in two different ways. All missions had 10 trials except 4 and 5 which had 30 and 40 trials respectively. Data from humans was collected in two rounds, first with 20 subjects (subjects 1–20), then with 30 different subjects (subjects 21–50).

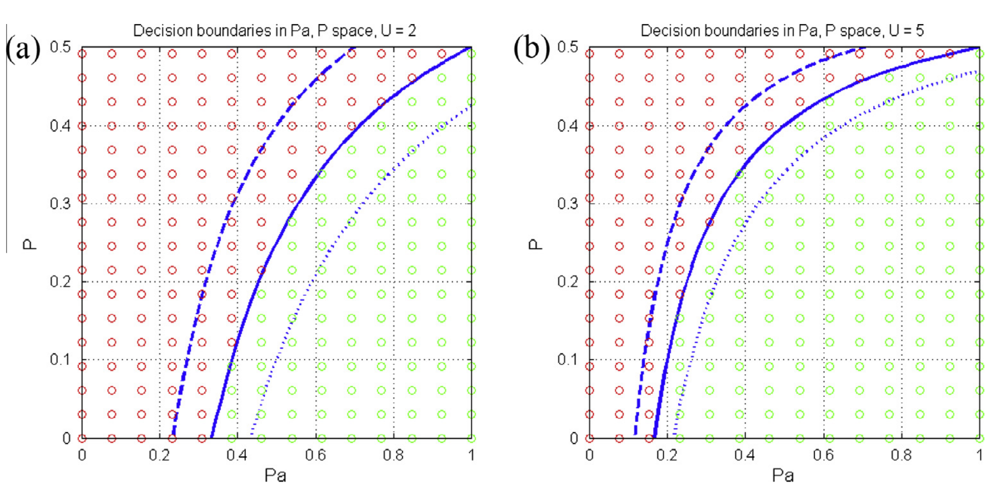

Fig. 1 Decision boundaries in *Pa*, *P*, *U* space for (a) $U = 2$ and (b) $U = 5$. For points below (or on or above) the curve, blue agents should divert (or not divert) a potential attack to minimize expected losses. Diverts and not diverts are depicted as a green or red circles in that order. The rational decision boundary is depicted with a solid line. An aggressive (or conservative) decision boundary with fewer (or more) diverts is depicted with a dotted (or dashed) line; see Section 'Human subjects' for more details.