Problem Set 4 (due Tues Mar. 9)
Operant matching and games

Mar. 2, 2004

1. Variable interval schedule, discrete time. Consider a single target with two possible states: baited or unbaited. If the animal chooses an unbaited target, it receives nothing. If it chooses a baited target, it harvests the reward, and the target switches to the unbaited state.

    If a target is baited at the end of a trial, it remains baited for the next trial. If a target is unbaited at the end of a trial (either because reward was harvested, or because the target was unbaited to begin with), then it is rebaited according to the toss of a coin with bias P.

    (a) Suppose that a target is unbaited at the end of a trial. Show that rebaiting occurs after a number of trials n drawn from the geometrical distribution $\mathcal{P}(n) = (1 - P)^{n-1} P$.

    (b) Show that the average of n is given by $1/P$.

    Note that this last result makes sense in two limiting cases. If $P = 1$, then the target is rebaited immediately, so that the average n is one. In the limit $P \to 0$ the time until rebaiting diverges to infinity.

2. Concurrent VI schedule, discrete time. Suppose there are two targets, both running independent VI schedules with baiting probabilities P and $\bar{P}$. Note that these two probabilities are generally chosen rather small (low overall rate of reward), so that $P + \bar{P} < 1$. In practice, the schedules would not be entirely independent, because a changeover delay would be imposed. But here no changeover delay will be assumed, so as to simplify the mathematics.

    Consider the probabilistic strategy of choosing between the targets by tossing a biased coin with odds $p : \bar{p}$, where $p + \bar{p} = 1$. Within this class of strategies, what is the optimal p? We will refer to this as the "optimal strategy", though there are other strategies outside this class that are superior.

    To solve this problem, let's calculate the expected reward as a function of p, and then maximize with respect to p. The total expected reward can be written as

    $$H_{tot} = H + \bar{H}$$

    where H and $\bar{H}$ are the expected rewards at the two targets. Assuming that all rewards have unity magnitude, these can be written as

    $$H = \frac{1}{T} \qquad \bar{H} = \frac{1}{\bar{T}}$$

    where T and $\bar{T}$ are the average time intervals between harvestings at the two targets.

    (a) Show that

    $$T = 1 + \frac{1 - P}{P} + \frac{1 - p}{p} \tag{1}$$

    $$\bar{T} = 1 + \frac{1 - \bar{P}}{\bar{P}} + \frac{1 - \bar{p}}{\bar{p}} \tag{2}$$

    Hint: The waiting time T is the sum of two times, the time required for another reward to appear at the target, and then the time required for the target to be chosen,

(b) Consider H_{tot} as a function of p, and assume that it is smooth and differentiable. Suppose that H_{tot} is maximized for $0 < p < 1$. Show that this implies

$$\frac{H}{p} = \frac{\bar{H}}{\bar{p}}$$

In other words, the matching and maximizing strategies are one and the same for concurrent VI in discrete time. This is a special case; in general, matching and maximizing are different.

(c) Show that the matching/maximizing solution p^* satisfies

$$\frac{p^*}{\bar{p}^*} = \frac{P}{\bar{P}}\frac{1 - \bar{P}}{1 - P}$$

This means that the optimal choice probabilities are approximately in the same ratio as the baiting probabilities P and $\bar{P}$, when the baiting probabilities are small. In general, this is not the case.

3. **Learning matching.** Consider a simple learning model for operant matching behavior. Let the action taken in trial t be given by the binary variables a_t and $\bar{a}_t$, which satisfy $a_t + \bar{a}_t = 1$. After each action, the animal receives reward h_t, which is also a binary variable. The incomes in the recent past are computed by

$$\begin{aligned} H_{t+1} &= \beta H_t + (1 - \beta)h_t a_t & (3) \\ \bar{H}_{t+1} &= \beta \bar{H}_t + (1 - \beta)h_t \bar{a}_t & (4) \end{aligned}$$

where $0 < \beta < 1$ is a discount factor. Based on these historical incomes, the model chooses its action randomly, with odds given by

$$\frac{p_t}{\bar{p}_t} = \frac{H_t}{\bar{H}_t}$$

(a) Program a concurrent VI schedule on two targets with baiting probabilities 0.1 and 0.2. Simulate the learning model for various values of β. What value of β is good for convergence to the optimal?

(b) Graph the results of the learner by plotting cumulative choices of target 2 vs. cumulative choices of target 1. Also plot cumulative rewards harvested from target 2 vs. cumulative rewards harvested from target 1. Matching behavior corresponds to equal slopes of these two curves.

(c) Suppose that halfway through your simulation, you exchange the baiting probabilities of the two targets. How quickly can the learning model adapt to this situation? How does your answer depend on β?

Explain your answers with words, figures, and/or equations. Also submit your MATLAB code.

4. **Pure strategy Nash equilibria**

(a) Solve the following game by eliminating dominated strategies. In what order are they eliminated? What is the Nash equilibrium?

	N	C	J
N	73,25	57,42	66,32
C	80,26	35,12	32,54
J	28,27	63,31	54,29

(b) Battle of the sexes. Robert (row) and Cecilia (column) are in love. They desperately want to spend every moment together, and are considering watching a video together this evening. Robert would like to see the horror flick *Attack of the Killer Tomatoes*, while Cecilia prefers *Babette's Feast*, based on the Isak Dinesen novel. The payoff matrix for this strategic interaction is

	A	B
A	2,1	0,0
B	0,0	1,2

What are the pure strategy Nash equilibria? Is the game dominance solvable?

5. Mixed strategy Nash equilibria. The battle of the sexes also has a mixed strategy Nash equilibrium, in which Robert chooses A with probability $0 < p < 1$ and Cecilia chooses A with probability $0 < q < 1$.

   (a) Find p and q. Hint: From the *Fundamental theorem of mixed strategy Nash equilibria* (see textbooks if you want to know it), Robert is indifferent at equilibrium, i.e. his expected payoffs from his two choices are the same. The same holds for Cecilia. From these conditions, solve for p and q.

   (b) What are the average payoffs for Robert and Cecilia? Are they better or worse than the payoffs for the pure strategy Nash equilibria?