Behavioural Processes 147 (2018) 33–37

Contents lists available at ScienceDirect

Behavioural Processes

journal homepage: www.elsevier.com/locate/behavproc

Is state-dependent valuation more adaptive than simpler rules?

Joseph Y. Halpern[a,*], Lior Seeman[b]

[a] Cornell University, United States
[b] Uber, United States

ARTICLE INFO

Keywords:
State-dependent valuation
Rational behavior
Finite automaton
Computational limitations

ABSTRACT

McNamara et al. (2012) claim to provide an explanation of certain systematic deviations from rational behavior using a mechanism that could arise through natural selection. We provide an arguably much simpler mechanism in terms of computational limitations, that performs better in the environment described by McNamara et al. (2012). To argue convincingly that animals' use of state-dependent valuation is adaptive and is likely to be selected for by natural selection, one must argue that, in some sense, it is a better approach than the simple strategies that we propose.

1. Introduction

Although much animal behavior can be understood as rational, in the sense of making a best response in all situations, some systematic deviations from rationality have been observed. For example, Marsh et al. (2004) presented starlings with two potential food sources, one which had provided food during "tough times", when the birds had been kept at low weight, while other had provided food during "good times", when the birds were well fed. They showed that the starlings preferred the food source that had fed them during the tough times, even when that source had a longer delay to food than the other source. Similar behavior was also observed in fish and desert locusts (Aw et al., 2009; Pompilio et al., 2006).

McNamara et al. (2012) claim to provide an explanation of this behavior using a mechanism that could arise through natural selection. They provide an abstract model of the bird-feeding setting where a decision maker can choose either a "risky" action or a "safe" action. They also provide a mechanism that takes internal state into account and can lead to good results (where, in the example above, the internal state could include the fitness of each source). However, as we observe, for the particular parameters used in their model, there is a *much* better (in the sense of getting a higher survival probability) and much simpler approach than their mechanism that does not take the internal state into account: simply playing safe all the time. It is hard to see how the mechanism proposed by McNamara et al. could arise in the model that they use by natural selection; the simpler mechanism would almost surely arise instead.

The fact that always playing safe does well depends on the particular parameter settings used by McNamara et al. Playing safe would not be a good idea for other parameter settings. However, we show that a simple 2-state automaton that more or less plays according to what it last got also does quite well. It does significantly better than the McNamara et al. mechanism, and does well in a wide variety of settings. Although our automaton also takes internal state into account (the internal state keeps track of the payoff at the last step), it does so in a minimal way, which does not suffice to explain the irrational behavior observed.

It seems to us that to argue convincingly that the type of mechanism proposed by McNamara et al. is adaptive and is likely to be selected for by natural selection, and thus explains animals' use of state-dependent valuation, then one must argue that, in some sense, it is a better approach than the simple strategies that we propose. Now it could be that the simple strategies we consider do not work so well in a somewhat more complicated setting, and in that setting, taking the McNamara et al.'s approach does indeed do better. However, such a setting should be demonstrated; it does not seem easy to do so. In any case, at a minimum, these observations suggest that McNamara et al.'s explanation for the use of state-dependent strategies is incomplete.

We should add that we are very sympathetic to the general approach taken by McNamara et al., although our motivation has come more from the work of Wilson (2015) and Halpern et al. (2012, 2014), which tries to explain seemingly irrational behavior, this time on the part of humans, in an appropriate model. That work assumes that people are resource-bounded, which is captured by modeling people as finite-state automata, and argues that an optimal (or close to optimal) finite-state automaton will exhibit some of the "irrational" behavior that we observe in people. (The 2-state automaton that we mentioned above is in fact a special case of a more general family of automata

[*] Corresponding author.
E-mail addresses: halpern@cs.cornell.edu (J.Y. Halpern), lior.seeman@gmail.com (L. Seeman).

https://doi.org/10.1016/j.beproc.2017.12.001
Received 1 May 2017; Received in revised form 10 October 2017; Accepted 4 December 2017
Available online 13 December 2017

considered in Halpern et al. (2012); see Section 3.3.) We believe that taking computational limitations seriously might be a useful approach in understanding animal behavior, and may explain at least some apparently irrational behavior.

The rest of this paper is organized as follows. In Section 2, we review the model used by McNamara et al. (2012) and compare it to that of Halpern et al. (2012). In Section 3, we describe four strategies that an agent can use in the McNamara et al. model, under the assumption that the agent knows which action is the risky action and which is the safe action. One is the strategy used by McNamara et al.; another is a simplification of the strategy that we considered in our work; the remaining two are baseline strategies. In Section 4, we evaluate the strategies under various settings of the model parameters. In Section 5, we consider what happens if the agent does not know which action is risky and which is safe and, more generally, the issue of learning. We conclude in Section 6.

2. The model

McNamara et al. (2012) assume that agents live at most one year, and that each year is divided into two periods, winter and summer. Animals can starve to death during a winter if they do not find enough food. If an agent survives the winter, then it reproduces over the summer, and reproductive success is independent of the winter behavior.

A “winter” is a series of T discrete time steps. At any given time, the environment is in one of two states: G (good) or S (sparse); the state of the environment is hidden from the agent. At every time step there is a small probability z of the environment switching states. At each time step, there are two actions potentially available to the agent, A (which we think of as the “risky” action) or B (the “safe” action). (The names “safe” and “risky” are due to the fact that the reward swings, depending on whether the environment is good or scarce, are greater for A than for B.) With probability γ, both options are available to the agent; with probability $(1-\gamma)/2$, the agent must play A; and with probability $(1-\gamma)/2$, the agent must play B. The payoff of actions A and B depends on whether the state of the environment is G or S.

An agent has a certain level of “energy reserves”, denoted by an integer between 0 and 10. The maximum level of energy reserves is thus 10; an agent dies if his energy reserve level is 0. At each time step, one unit of energy reserves is consumed. At each time step, an agent receives 0, 1 or 2 units of energy. The probability of each of these amounts is drawn from a binomial distribution $bin(2, p)$ (so that the probability of receiving 0 units is $(1-p)^2$, the probability of receiving 1 unit is $2p(1-p)$, and the probability of receiving 2 units is p^2), where p depends on the current environment state and the choice of action. P_{GA} is used to denote the probability p when the environment is in state G and the agent plays action A; we can similarly define P_{GB}, P_{SA}, and P_{SB}. McNamara et al. assume that rewards are higher in expectation in the good environment for both actions, that is, $P_{GA} \geq P_{SA}$ and $P_{GB} \geq P_{SB}$); moreover, A is the better action in the good environment, while B is better in the sparse environment, so $P_{GA} \geq P_{GB}$ and $P_{SB} \geq P_{SA}$.

It is interesting to compare this model to that used by Halpern et al. (2012). Although, we mentioned in the introduction, their goal was to study irrational behavior in humans, and the kinds of behaviors considered were quite different from those considered by McNamara et al. (the focus was on modeling the behavior in game playing reported by Erev et al. (2010)), the models are surprisingly similar. The main differences between the two models is that, in the model of Halpern et al. (2012), an agent's objective is to maximize his expected average payoff over rounds (rather than just to maximize the probability of surviving for a year). Agents never die; and an agent's utility is taken to be the limit of his average reward per round over an infinite time horizon. In the language of McNamara, Trimmer, and Houston, Halpern, Pass, and Seeman take $\gamma = 1$, so that both actions are always available. Moreover, instead of observing the payoff (which is deterministically dependent on the state of the world), an agent gets a signal correlated with the real state of the environment when he plays A, and no signal when he plays B. As discussed by Halpern et al. (2012), getting a noisy payoff as in the McNamara et al. (2012) model has essentially the same effect as getting a signal correlated with the environment's state. As we discuss later, in the scenarios we study here, P_{GB} and P_{SB} are very close and thus the signal we get about the environment by playing B is very weak in this model as well.

3. Four strategies

In this section, we describe four strategies that an agent can use in the McNamara et al. model. We will be interested in the probability that an agent survives a “winter” period using each of these strategies. Note that the higher this probability is, the greater the probability that this strategy will emerge as the dominant strategy in an evolutionary process.

3.1. Baseline strategies

We consider two baseline strategies. The first is called the *oracle strategy*. With this strategy, we assume that an agent knows the true state of the environment before choosing his action, and thus can make the optimal choice in each round. While this strategy cannot be implemented by an agent in this model, we use it to provide an upper bound on the the survival probability of the agent. Clearly, no strategy can do better than the oracle strategy.

The second baseline strategy we consider is the *safe strategy*. With the safe strategy, the agent always plays the safe action (B) when that choice is available (recall that in some rounds the agent is forced to play A).

3.2. The value strategy

The strategy studied by McNamara et al. (2012), which we call the *value strategy*, is based on keeping a value $V(\cdot)$ for each of the actions and choosing the action with the highest value in each round. This value is updated in every round using the formula $V_{\text{new}} = (1-\beta)V_{\text{old}} + \beta w$, where β is a fixed parameter controlling the learning rate and w is the *perceived reward*, which is defined in more detail below.

In a little more detail, $V(i)$ is initialized to the expected energy reward of action i. In a round where action i is performed, $V(i)$ is updated using the formula above (taking $V_{old}(i)$ to be the currently stored value of $V(i)$ and $V_{new}(i)$ to be the updated value), where $w = \text{u}e^{k(r-5)}$, u is the number of energy units received as a result of performing action i, r is the current reserve level, and k is a fixed constant that might evolve to match the scenario parameters.

In a round where the agent can play both A and B (which will be the case with probability $1-\gamma$), it plays whichever one has higher value. That is, it plays A if $V(A) \geq V(B)$, and otherwise plays B.

3.3. The automaton strategy

The last strategy we consider is inspired by the strategy used by Halpern et al. (2012); we call it the *automaton strategy*. With the automaton strategy, an agent keeps an internal state that is correlated with the number of good and bad signals it has seen recently. Thus, the internal state is not determined by the agent's internal reserves, but rather by recent observations. This strategy is described by a finite automaton with n states denoted $[0, \ldots, n-1]$. If the automaton has a choice, then it plays action B (the “safe” action) in state 0 and plays action A in all the remaining states. (Thus, if the automaton has only one state, it plays the safe strategy.) The automaton changes state depending on the signal it observes. Halpern, Pass, and Seeman assumed that an automaton in