Cognitive Psychology 85 (2016) 43–77

Contents lists available at ScienceDirect

Cognitive Psychology

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

Learning and choosing in an uncertain world: An investigation of the explore–exploit dilemma in static and dynamic environments

Daniel J. Navarro [a,*], Ben R. Newell [b], Christin Schulze [b]

[a] School of Psychology, University of Adelaide, Australia
[b] School of Psychology, University of New South Wales, Australia

ARTICLE INFO

Article history:
Accepted 2 January 2016
Available online 21 January 2016

Keywords:
Decision making
Dynamic environments
Explore–exploit dilemma
Decisions from experience

ABSTRACT

How do people solve the explore–exploit trade-off in a changing environment? In this paper we present experimental evidence from an "observe or bet" task, in which people have to determine when to engage in information-seeking behavior and when to switch to reward-taking actions. In particular we focus on the comparison between people's behavior in a changing environment and their behavior in an unchanging one. Our experimental work is motivated by rational analysis of the problem that makes strong predictions about information search and reward seeking in static and changeable environments. Our results show a striking agreement between human behavior and the optimal policy, but also highlight a number of systematic differences. In particular, we find that while people often employ suboptimal strategies the first time they encounter the learning problem, most people are able to approximate the correct strategy after minimal experience. In order to describe both the manner in which people's choices are similar to but slightly different from an optimal standard, we introduce four process models for the observe or bet task and evaluate them as potential theories of human behavior.

* Corresponding author.
E-mail addresses: dan.navarro@unsw.edu.au (D.J. Navarro), ben.newell@unsw.edu.au (B.R. Newell), cschulze@mpib-berlin.mpg.de (C. Schulze).

http://dx.doi.org/10.1016/j.cogpsych.2016.01.001
0010-0285/

1. Introduction

A defining characteristic of decision making under uncertainty is that people lack definitive evidence to motivate their choices, due to ambiguity, insufficient expertise or missing information. In many laboratory tasks, people are simply given a partial specification of the problem and asked to make choices. In real life, however, people are often required to expend time and effort acquiring the information needed to guide their choices. This creates a dilemma: the actions that return the most information about the world are not necessarily the same as those actions the greatest immediate reward. This disconnect plays out in many real world scenarios:

- Workers on online marketplaces (e.g., Amazon Mechanical Turk) typically do not get paid until they complete a task, and cannot know their hourly wage until later when they are paid by their employer. The action that yields rewards is to do tasks, but the action that yields most information is to read reviews (e.g., on Turkopticon) of potential employers.
- Grant agencies have limited budgets with which to fund projects. The actions (monetary allocations) that yield greatest information are to employ officers to solicit grant applications, review progress reports, and check audit trails, but the ones that yield rewards are those that allocate money directly to projects.
- Manufacturing processes that generate a stream of outputs (e.g., factory productions lines) provide rewards for the company when those goods are sold to consumers, but only if the goods are not faulty. Allocating resources to quality control processes (product testing) produces more information about the goods, but at the cost of taking resources away from the production line itself.

In these and many other scenarios, there is an inherent tension between selecting actions that maximize immediate rewards and actions that maximize immediate information gain. In an ideal world, a decision maker would not be forced to choose between information and reward, but this rarely occurs in practice. The decision maker either has resource constraints (e.g., funding agencies are often very short on money), time constraints (e.g., online workers cannot devote attention to two things at once) or physical constraints (e.g., quality control processes often require destructive tests – measuring tensile strength of a steel bar, for instance) that ensure that there is some trade off involved. In the long run, of course, information eventually works its way back to the decision maker: the online worker gets paid, the agency finds out which projects worked, the manufacturer learns which product lines had to be recalled. In the short term, however, this delayed feedback means that the decision maker must find some way to balance the search for information against the need to generate rewards. It is almost never a wise idea to forego all information-rich actions in the hope that when the rewards (and hence feedback) eventually arrive, one's reward-seeking actions will turn out to have been good ones.

Our focus in this paper is on a laboratory task which shares some of the fundamental features of these situations. The task presents a very clear distinction between information that is obtained via observation – but is not associated with any immediate reward – and actions that can lead to rewards but for which only delayed feedback about the (non) occurrence of a reward is available. Specifically we examine the "observe or bet" task introduced by Tversky and Edwards (1966), in which the decision maker has a number of options available, each of which may yield rewards or losses. On each trial, she may choose to observe the state of the world, in which case she gets to see what rewards each option provided, but receives no reward nor suffers any losses. Alternatively she may pick one of the options (i.e., bet on it) and receive the rewards/losses associated with that option at the end of the task. However, she receives no information at the time of the choice: the outcomes are hidden from her. By separating information from reward so cleanly, the task provides a very pure means by which to assay the explore–exploit dilemma (e.g., Cohen, McClure, & Yu, 2007; Hills, Todd, Lazer, Redish, & Couzin, 2015; Mehlhorn et al., 2015).

Moreover, we are interested in how people deal with this very stark form of an explore–exploit dilemma when there is some possibility that the world can change (cf. Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Gureckis & Love, 2009; Knox, Otto, Stone, & Love, 2011; Speekenbrink &