Available online at www.sciencedirect.com

SCIENCE DIRECT®

Cognitive Systems Research 6 (2005) 27–40

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Adapting to the task environment: Explorations in expected value

Action editor: Christian Schunn

Wayne D. Gray *, Michael J. Schoelles, Chris R. Sims

Rensselaer Polytechnic Institute, Carnegie Building, 110 8th St. Troy, NY 12180, USA

Received 16 September 2004; accepted 16 September 2004

Abstract

Small variations in how a task is designed can lead humans to trade off one set of strategies for another. In this paper we discuss our failure to model such tradeoffs in the Blocks World task using ACT-R's default mechanism for selecting the best production among competing productions. ACT-R's selection mechanism, its *expected value* equation, has had many successes (see, for example [Anderson, J. R., & Lebiere, C. (Eds.). (1998). *Atomic components of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.]) and a recognized strength of this approach is that, across a wide variety of tasks, it tends to produce models that adapt to their task environment about as fast as humans adapt. (This congruence with human behavior is in marked contrast to other popular ways of computing the utility of alternative choices; for example, Reinforcement Learning or most Connectionist learning methods.) We believe that the failure to model the Blocks World task stems from the requirement in ACT-R that all actions must be counted as a binary success or failure. In Blocks World, as well as in many other circumstances, actions can be met with mixed success or partial failure. Working within ACT-R's expected value equation we replace the binary success/failure judgment with three variations on a scalar one. We then compare the performance of each alternative with ACT-R's default scheme and with the human data. We conclude by discussing the limits and generality of our attempts to replace ACT-R's binary scheme with a scalar credit assignment mechanism.

Keywords: Expected utility; Expected value; Credit assignment; Strategy acquisition; Adaptive behavior; Reinforcement learning

1. Introduction

Few tasks are so new as to require the invention of strategies that have never been used by the task performer. Hence, in many situations, settling on a

* Corresponding author. Tel.: +1 518 276 3315; fax: +1 518 276 8268.

E-mail address: grayw@rpi.edu (W.D. Gray).

1389-0417/$ - see front matter
doi:10.1016/j.cogsys.2004.09.004

strategy or set of strategies for performing a task is not so much a matter of learning new strategies as it is learning which strategy, out of a set of already acquired strategies, is best adapted to the current task environment.

Two remarkable aspects of this adaptation are that it is usually continuous and often unguided. Strategy selection continues to change and evolve even when the task performed is a routine act such as making photocopies of a book chapter (Agre & Shrager, 1990). This process occurs despite the absence of supervision or explicit guidance. In fact, performance improves far beyond what would be expected if, for each step, the choice among *n* possible alternatives were based solely on local considerations of utility. Generally, the class of non-local cumulative-effects models (Davis, Staddon, Machado, & Palmer, 1993) required to explain this behavior is known as unsupervised learning (Sutton & Barto, 1998). This is in contrast to supervised learning (such as is used in most neural networks) where the learning agent is told not only when it errs, but also how it should have behaved differently.

This paper is motivated by our attempts to model strategy selection in a Blocks World paradigm using ACT-R. First we introduce Blocks World and the empirical phenomena we seek to model. Second, in ACT-R a type of non-local cumulative effects model referred to as the *expected value* equation (Anderson et al.; Anderson & Lebiere, 1998) determines which of two or more alternative strategies will be selected. After introducing the expected value equation, we present data from two variations of a model that uses the default equation. The variations differ by whether we update expected value after each strategy is executed, or whether we update after the entire task is completed. We then discuss reasons why the ACT-R mechanism is inadequate for modeling Blocks World. Third, we present three variations of ACT-R's expected value equation and present data from the original model run with each variation. For each, we discuss how the variation influenced model behavior as well as its fit or misfit to the empirical data. Fourth, we present results from two variations of an abstract model that uses ACT-R's expected value equation, but replaces nearly all else with estimates obtained directly from the human data. (As before, the variations differ in terms of when expected value is computed.) Fifth and finally, we summarize our work and draw conclusions regarding the Blocks World task specifically, our variations for calculating expected value, as well as the implications of our results for ACT-R.

2. Blocks World

Blocks World is a simple task that has been used to study the tradeoff between interaction-intensive and memory-intensive strategies (Ballard, Hayhoe, & Pelz, 1995; Fu & Gray, 2000; Gray & Fu, 2000; Gray, Sims, Fu, & Schoelles, in preparation). The task is to copy a pattern of colored blocks shown in the *Target window* to the *Workspace* window, using the colored blocks in the *Resource* window (for our version see Fig. 1).

2.1. The Blocks World studies

Each trial begins with a random placement of 8 colored blocks into empty spaces (defined by an invisible 4×4 grid) in the Target window. Unlike

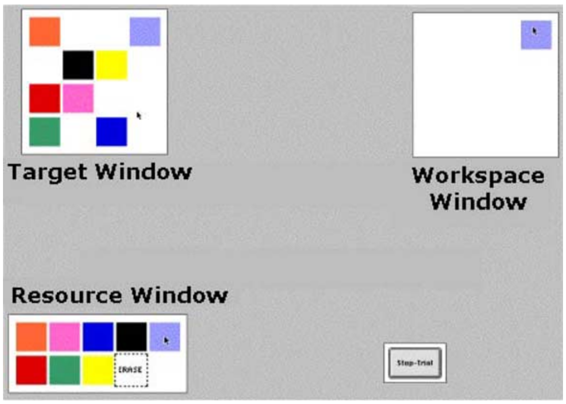

Fig. 1. The Blocks World task at the start of a new trial. In the actual task all windows are covered by gray boxes and at any time only one window can be uncovered. (The labels do not appear in the actual task. The Start/Stop button is shown at the lower right.)