From: AAAI Technical Report SS-02-07. Compilation copyright © 2002, AAAI (www.aaai.org). All rights reserved.

What Monkeys See and Don't Do: Agent Models of Safe Learning in Primates

Joanna J. Bryson and Marc D. Hauser
Harvard Primate Cognitive Neuroscience Laboratory
33 Kirkland Ave, Room 970
Cambridge, MA 02138
jbryson@wjh.harvard.edu, hauser@wjh.harvard.edu

Abstract

This abstract suggests that some of the unusual findings in the primate learning literature may reflect evolved solutions to the problem of safe learning in intelligent agents. We also propose a research program for trying to model this learning, and incorporate it into agent architectures and development methodology.

This extended abstract is actually a summary of a research proposal we have submitted for funding. If we don't actually wind up being funded to do this work, we will probably withdraw this paper from your symposium, unless you are very interested and we can find travel funding elsewhere!

Introduction

Biological intelligence is fascinating because it performs so well with so many constraints. Any agent embedded in the real world must be able to make decisions under time pressure, nearly always with incomplete information. As Artificial Intelligence (AI) has moved to consider the problems of real-time agents in complex dynamic environments, it has been forced to abandon the search for provably-optimal or even provably-correct plans to govern decision making.

Instead, real-time AI agents now generally rely on *reactive planning*: a process whereby the next action is chosen by a look-up indexed on the agent's perception of the current environment. For well-ordered behavior in complex agents (e.g. those capable of pursuing multiple, potentially-conflicting goals), the agent generally also uses stored state on recent decisions to focus attention on a subset of possible actions.

This abstract proposes research to examine how to safely incorporate new behavior into established reactive plans. We are particularly concerned with modeling *how* plans are updated, and *when*. Existing primate research shows that although an animal may possess a skill that is applicable in a situation, it may never attempt to apply it, preferring established solutions even if they are reliably failing. On the other hand, particular learning situations can result in the animals changing these preferences. We intend to build functioning AI models of these situations to test several current theories of specialized learning for action-selection control.

Copyright © 2001, American Association for Artificial Intelligence (www.aaai.org). All rights reserved.

The primary expected benefit of this research is a set of new idioms for agent architectures which allow for safe, autonomous extension by the agent of its existing reactive plans. Notice that in order to assure safety, we expect that, like the monkeys, these agents will sometimes fail to exploit potentially useful behaviors. However, we hope to develop a framework for helping an agent determine when new behaviors are likely to be useful.

The remainder of this abstract is in two parts. We will first discuss the primate research which we wish to model. We will then describe our proposed approach.

Primate Learning

Why Model Primates?

Having adaptable artificial agents is an obvious and significant goal for AI. We would like agents to be able to learn from experience, adapt to individual users, correct their own software, and so forth. Unfortunately, the qualities 'safe' and 'adaptive' are generally in opposition. Behavior that can change is, almost definitionally, less reliable than behavior that cannot. Extensive research in constraints on animal learning indicate that:

- learning is a relatively easy adaptation to evolve (Marler, 1991), however,
- it is used only sparingly, and with severe constraints, even in mammals (Roper, 1983; Gallistel *et al.*, 1991).

In general, although individual adaptation can be a useful mechanism for allowing a species to move between niches, evolution tends to favor hardcoding solutions of high general utility, because this guarantees *every* individual finding the solution (Baldwin, 1896; Turney, 1996; Belew & Mitchell, 1996).

Primates are truly exceptional in their capacity of individual adaptation, a fact that no doubt facilitates the incredible cultural development of our own species. But even within human and non-human primates, there are significant restrictions on learning. There is some evidence that the reason for primates' unique intelligence is actually social / sexual selection (Byrne & Whiten, 1988; Whiten & Byrne, 1997). That is, primate intellect *in itself* may not be particularly 'fit' in terms of general-purpose survival, but rather an attribute of sexual selection. In proposing this research, we are hypothesizing that the still existent restrictions on primate learning may have significant utility in allowing the agent to adaptive, yet safe and reliable in dangerous, dynamic environments.

Specific Research to be Modeled

These are tasks we have selected for modeling. All of these tasks involve an agent discovering the way to reliably retrieve a desired object. The research of this tasks is reviewed by Hauser (1999), further references can be found in that article.

Boxed-Object Task In this task (originally designed by Diamond (1990)) an agent is exposed to a clear box with food inside of it and only one open side, which is facing the left or right. Human infants under a particular age and some adult primates will repeatedly reach straight for the food, despite repeated failure due to contact with the transparent face of the box. More mature children and some other primate species succeed in this task by finding the open face of the box. For a short intermediate period of human development and for mature cotton-top tamarins (one of the species in our lab) subjects learn the task if they are first exposed to an opaque box, and then transferred to the transparent one. This shows that some tasks are easier to learn than others, and that knowledge about some tasks can be relevant to learning others, but that both of these factors are age and species dependent.

We intend to model the tamarins' performance on this task. In this case, the high salience of the visible reward blocks the exploratory behavior that *might* find another solution, but does not block the adoption of a relatively certain solution that had been learned in a different framework. Modeling this requires first modeling the interaction between perceptually driven motivation and the operation of control plans and the operations of a meta-learning behavior that provides both for exploration and for incorporating discoveries into the behavior repertoire. We hope to develop a parameter-based model of these interactions and attempt to match individual tamarin learning profiles to account for individual differences as well as gross behavior.

Cloth-pulling Task In this task, tamarins learn to discriminate relevant cues as to whether a cloth can be used to retrieve a piece of food. The primary relevant cue is whether the food is on a piece of cloth within grasp of the tamarin — distractors include color, texture, shape and so forth. Here tamarin can fully learn the task, but still be persuaded to try ineffective strategies by a particularly tempting reward stimulus. This research will provide further testing of the motivation integration model described in the boxed-object task.

Tube-Following Task When a piece of food is launched down an opaque, s-shaped tube, tamarins incorrectly expect it to land directly beneath the release point if the test apparatus is standing upright. However, tamarins can easily learn to trace the tube path in the case where the food moves through a horizontal apparatus. In the vertical case, the learning seems biased by a strong prior expectation (of either experiential or genetic origin) for the effect of gravity. This research will provide further testing of the behavior integration model described in the boxed-object task.

Analysis: The Safe Incorporation of New Behavior

There are two reasons an agent might apply an appropriate behavior rather than an inappropriate one:

1. it might not know the appropriate behavior, or
2. it might fail to inhibit the inappropriate one.

Similarly, there are two reasons why an agent might fail to inhibit an inappropriate behavior:

1. there may be a general failure of the inhibition mechanism, or
2. it may be incorrectly assigning the inappropriate behavior higher priority in the present behavioral context.

Notice that the process of exploring (searching for a new appropriate behavior) is itself a behavior.

These latter two options will be the primary focus of our research: we will use standard machine learning (e.g. Bishop, 1996) for developing new categories of perception and high-level abstractions in Artificial Life (ALife) simulations for the mechanics of newly learned behaviors. What we consider key is how a new behavior comes to be integrated into ordered action selection.

A Safe Adaptive Artificial Agent Architecture

Specialized Learning in AI

The central problem of AI is search. An agent must find a way to behave that reliably increases the probability of its goals being met (Albus, 1991). This search includes finding ways to act, finding appropriate situations in which to act, and finding the correct information to attend to for determining both situation and action.

Unfortunately, the problem of search is combinatorially explosive, and thus cannot be solved optimally by a resource-bounded agent (Chapman, 1987; Gigerenzer & Todd, 1999). Consequently, the intelligence of an agent is dependent on what knowledge and skills can be provided to it at its inception. The less search an agent has to perform itself, the more likely it is to succeed. For an animal, this information is provided either genetically or, for a few species, culturally (that is, from other agents.) In an artificial agent, this information is provided by the programmer. Providing for an adaptive agent requires focusing (also called *constraining* or *biasing*) search in such a way that it is both likely to succeed and unlikely to interfere with other, established behavior patterns. Such provisions are currently best performed in an artifact via state-of-the-art software engineering, incorporating modular design and specialized representations.

Our Agent Engineering Approach

We intend to use and possibly extend Behavior-Oriented Design (BOD) (Bryson & Stein, 2001a; Bryson, 2001), which incorporates a modular agent architecture in the tradition of Behavior-Based AI (e.g. Brooks, 1991; Arkin, 1998). BOD

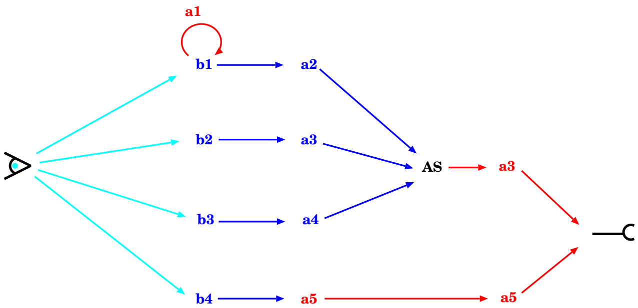

Figure 1: BOD *behaviors* ($b_1 \ldots$) generate *actions* ($a_1 \ldots$) based on their own perception (derived from *sensing*, the eye icon). Actions which affect state external to their behavior (e.g. *expressed acts*, the hand icon), may be subject to arbitration by *action selection* (AS) if they are mutually exclusive (e.g. sitting and walking). (Bryson & Stein, 2001b)

exploits Object-Oriented Design (e.g Parnas, Clements, & Weiss, 1985; Coad, North, & Mayfield, 1997), currently the leading software development paradigm. BOD has explicit reactive plans which support action selection when that action selection requires arbitration between different behavior modules. However, unlike some more restricted 'hybrid' behavior-based and planning architectures (Georgeff & Lansky, 1987; Bonasso *et al.*, 1997, e.g.), behavior modules in BOD are not mere primitives controlled by the planning structures. Rather, BOD utilizes a truly modular architecture based around specialized representation (see Figure 1).

In BOD, the state and representations necessary for adaptation and dynamic control are in the heart of the modules, which are coded as objects. Plan primitives are an interface to the actions (methods) of these modules. Where arbitration is necessary, a reactive plan determines *when* actions happen. But the behavior modules determine *how*. They also process the sensory information and maintain the state relevant to performing those actions.

Expected Architectural Development

To date, the emphasis of the BOD methodology has been that adaptation required by the agent should be allocated a specialized representation embedded in a behavior module charged with learning that adaptation. This approach may seem overly constrictive, but is supported by the existing animal learning literature mentioned above (Roper, 1983; Gallistel *et al.*, 1991). Also, it limits the amount of damage to critical control that adaptation might accidently inflict. It can easily be used to model such phenomena as imprinting, map learning, kin recognition and perceptual categorization. In other words, BOD currently facilitates modification of existing behavior, but not the creation of *new* behavior.

Learning a completely new behavior is the kind of combinatorially implausible search discussed in the introduction. Primates learn *partially* new behaviors, by exploiting and adapting behaviors they already possess, and biasing search to reasonable representations for the reorganization. We propose to model this process by extending BOD to allow the learning of new reactive plan structures in combination with specialized behaviors for supporting perceptual and control modifications specific to the new context.

References

Albus, J. S. 1991. Outline for a theory of intelligence. *IEEE Transactions on Systems, Man and Cybernetics* 21(3):473–509.

Arkin, R. C. 1998. *Behavior-Based Robotics*. Cambridge, MA: MIT Press.

Baldwin, J. M. 1896. A new factor in evolution. *The American Naturalist* 30:441–451.

Belew, R. K., and Mitchell, M., eds. 1996. *Adaptive Individuals in Evolving Populations: Models and Algorithms*, volume XXVI of *Santa Fe Institute Studies in the Sciences of Complexity*. Reading, MA: Addison-Wesley.

Bishop, C. M. 1996. Neural networks: a principled perspective. Technical Report NCRG/95/014, Aston University.

Bonasso, R. P.; Firby, R. J.; Gat, E.; Kortenkamp, D.; Miller, D. P.; and Slack, M. G. 1997. Experiences with an architecture for intelligent, reactive agents. *Journal of Experimental & Theoretical Artificial Intelligence* 9(2/3):237–256.

Brooks, R. A. 1991. Intelligence without representation. *Artificial Intelligence* 47:139–159.

Bryson, J., and Stein, L. A. 2001a. Modularity and design in reactive intelligence. In *Proceedings of the 17th International Joint Conference on Artificial Intelligence*, 1115–1120. Seattle: Morgan Kaufmann.

Bryson, J., and Stein, L. A. 2001b. Modularity and specialized learning: Mapping between agent architectures and brain organization. In Wermter, S.; Austin, J.; and Willshaw, D., eds., *Emergent Neural Computational Architectures Based on Neuroscience.*, 98–113. Springer.

Bryson, J. J. 2001. *Intelligence by Design: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*. Ph.D. Dissertation, MIT, Department of EECS, Cambridge, MA.

Byrne, R. W., and Whiten, A., eds. 1988. *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans*. Oxford University Press.

Chapman, D. 1987. Planning for conjunctive goals. *Artificial Intelligence* 32:333–378.

Coad, P.; North, D.; and Mayfield, M. 1997. *Object Models: Strategies, Patterns and Applications*. Prentice Hall, 2nd edition.

Diamond, A. 1990. Developmental time course in human infants and infant monkeys, and the neural bases of higher cognitive functions. *Annals of the New York Academy of Sciences* 608:637–676.

Gallistel, C.; Brown, A. L.; Carey, S.; Gelman, R.; and Keil, F. C. 1991. Lessons from animal learning for the study of cognitive development. In Carey, S., and Gelman, R., eds., *The Epigenesis of Mind*. Hillsdale, NJ: Lawrence Erlbaum. 3–36.

Georgeff, M. P., and Lansky, A. L. 1987. Reactive reasoning and planning. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI-87)*, 677–682.

Gigerenzer, G., and Todd, P. M., eds. 1999. *Simple Heuristics that Make Us Smart*. Oxford University Press.

Hauser, M. D. 1999. Perseveration, inhibition and the prefrontal cortex: a new look. *Current Opinion in Neurobiology* 9:214–222.

Marler, P. 1991. The instinct to learn. In Carey, S., and Gelman, R., eds., *The Epigenesis of Mind*. Hillsdale, NJ: Lawrence Erlbaum. 37–66.

Parnas, D. L.; Clements, P. C.; and Weiss, D. M. 1985. The modular structure of complex systems. *IEEE Transactions on Software Engineering* SE-11(3):259–266.

Roper, T. J. 1983. Learning as a biological phenomena. In Halliday, T. R., and Slater, P. J. B., eds., *Genes, Development and Learning*, volume 3 of *Animal Behaviour*. Oxford: Blackwell Scientific Publications. chapter 6, 178–212.

Turney, P. 1996. How to shift bias: Lessons from the Baldwin effect. *Evolutionary Computation* 4(3):271–295.

Whiten, A., and Byrne, R. W., eds. 1997. *Machiavellian Intelligence II: Evaluations and Extensions*. Cambridge University Press.