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 paper describes a research program designed to use agent models to understand how primates incorporate new, learned behavior safely into their established, reliable behavior repertoires. We suggest that some of the findings in the primate learning literature that we currently find surprising may in fact reflect evolved solutions to the problem of safe learning in intelligent agents. We sketch an approach to trying to model this learning, with the expectation that our experience will also lead to insights into idioms and strategies for incorporating safe learning into artificial agents.

Introduction

Many Artificial Intelligence (AI) researchers have made the mistake of thinking that the main problems of learning are problems of quantity, such as providing adequate capacity or sufficiently rapid recall. One well-known exception is the concept of overfitting, that the function that most accurately records learned perception/behavior pairings is not necessarily the one that best predicts good behavior from novel perception (Hertz, Krogh, & Palmer, 1991, p. 147). But there is another sense in which more is not necessarily better. Agent intelligence can be complex and intricate, an island of local optimality surrounded by oceans of ineffective and even hazardous behavior (Schneider, 1997). If an agent has unlimited learning, then it can learn a version of behavior control that moves it off that island and into deep water. The challenge of safe learning is allowing an agent to adapt only enough to explore the safe territory.

One popular strategy for safe 'adaptive' behavior[1] in artificial agents is *reactive planning*. An agent using reactive planning chooses its next action through a look-up indexed on the agent's perception of the current environment. This sort of planning has been popularized because it is robust and efficient, allowing rapid, opportunistic response to complex, dynamic environments (Brooks, 1991; Georgeff & Lansky, 1987). It is 'adaptive' in that novel sequences of actions can be generated as appropriate. However, reactive planning does not involve long-term changes in the animal's behavior repertoire: in the same environmental and behavioral context, the animal will select the same action regardless of its previous outcome. Consequently, agents that rely completely on reactive planning for their adaptation are not said to be able to learn.

If we consider instead natural agents, we find that some amount of learning is ubiquitous. But we also find persistent failures to learn in even the most adaptable species, such as the primates. These failures cannot always be explained in terms of failures of perception, nor as a lack of capacity for the complexity of the learning task. Attempting to explain these failings has lead us to the hypothesis that failing to learn in some contexts may actually be an adaptive strategy (in the Darwinian sense) for protecting the overall quality of an agent's intelligence.

The AI research described in this paper reflects a new program funded by an exploratory grant under the NSF Biological Information Technology & Systems (BITS) initiative. We are attempting to gain new insights into safely incorporating learned behavior into established agent intelligence by modeling the learning of non-human primates. At the same time, we expect both the modeling itself and insights into the nature of the problem gleaned from the safe-learning agent community will help us to understand some of the unusual learning behavior witnessed in primates.

The following section of this paper discusses learning and its limits in nature, paying particular attention to primate learning. The next section describes from a more computational perspective why constraints can be advantageous, and how modular intelligence might be used to provide these limitations. We then describe our proposed course of research, and finally we discuss the relevance of our work to date to the more general questions of safe learning in agents.

Why Study Primates?

Introduction

Many researchers have a pre-Darwinian fallacy in their thinking about learning, particularly in nature. The assumption is that learning is ideally unconstrained, unbiased and

[1] I use 'adaptive' (with scare quotes) in the sense of *nouvelle AI*: "behaviors and underlying mechanisms that allow animals and, potentially, robots to adapt and survive in uncertain environments (Meyer & Wilson, 1991, p. *ix*)." Elsewhere in this paper I use the term (without scare quotes) in its Darwinian sense. I apologize for this confusion.

without limits, and that if evolution has not yet achieved this, it is well on its way, as witnessed by the incredible adaptability of humans. What makes this a pre-Darwinian fallacy is that it assumes that humans are the most evolved species on the planet. In fact, humanity is just one out of very many species all descended from the same incident of life and all species have been subject to the pressures of selectivity for the same amount of time. If the extent of the human ability to learn is special (which seems to be true), then there is a significant question as to why only one species learns to that extent.

In this section, we review evidence that learning in nature is generally restricted and specialized to particular tasks. We then discuss why primates may have come to be exceptionally adaptable, and therefore particularly interesting as models of safe learning. We conclude by reviewing some of the specific learning results we plan to model.

Limits on Learning in Nature

Earlier this century, behaviorists (a group of psychologists and animal researchers who concentrated on laboratory experiments), proposed that animals learn only through a general process of being able to create associations.

> The [behaviorists'] position is that all learning is based on the capacity to form associations; there are general laws of learning that apply equally to all domains of stimuli, responses, and reinforcers; the more frequent the pairings between the elements to be associated, the stronger the associative strength; the more proximate the members of an association pair, the more likely the learning. (Gallistel *et al.*, 1991)

Learning by association (conditioning) does appear to be a general learning mechanism with parameters that hold across species, presumably indicating a common underlying mechanism. However, behaviorist research itself eventually demonstrated that animals *cannot* learn to associate any arbitrary stimulus with any arbitrary response. Pigeons can learn to peck for food, but cannot learn to peck to avoid a shock. Conversely, they can learn to flap their wings to avoid a shock, but not for food (Hineline & Rachlin, 1969). In related experiments, rats presented with "bad" water learned different cues for its badness depending on the consequences of drinking it. If drinking lead to shocks, they learned visual or auditory cues and if drinking lead to poisoning they learned taste or smell cues (Garcia & Koelling, 1966).

These examples demonstrate highly specific, constrained and ecologically-relevant learning mechanisms. For example, the content of the associations rats are able to make biases their learning toward information likely to be relevant: poison is often indicated by smell or taste, while acute pain is often the consequence of something that can be seen or heard. Such results were originally interpreted as constraints placed on general learning to avoid dangerous associations, but research has since indicated the inverse. Specialized systems exist to form important associations (Roper, 1983). For example, poison avoidance in rats is handled by a specific one-shot-learning mechanism in the olfactory section of their amygdala.

The current ethological hypothesis is that learning by an individual organism serves as a last resort for evolution (Roper, 1983; Gallistel *et al.*, 1991; Marler, 1991). Interesting explorations and demonstrations of this hypothesis (including those using artificial models) can be found in the literature examining the Baldwin effect (Baldwin, 1896; Turney, 1996; Belew & Mitchell, 1996). The Baldwin effect indicates that individual adaptability can sustain useful genetic variations before they are fully and reliably encoded. Nevertheless, although individual learning may be sufficient to sustain such a transitional genetic trait, there is still selective pressure for full genetic coding to replace individual learning. This is because genetic coding provides a more reliable guarantee that *every* individual of the species will have the advantageous feature. In general, learning only persists when a behavior cannot be fully predetermined, because the competence involved requires flexibility on a less than evolutionary time scale[2] (Gallistel *et al.*, 1991). Examples of such competencies include landmark navigation (e.g. in bees) and calibrating perceptual systems (e.g. in barn owl chicks calibrating their hearing to their vision (Brainard & Knudsen, 1998)). Because the details of the owl's head-size or the bee's landmark location are dependent on unpredictable environmental factors, this knowledge must be learned rather than inherited.

If learning is typically restricted in nature, then why are some species, such as many primates, relatively capable? One possibility is that some ancestral primates came to inhabit a niche in which a greater capacity for learning *was* advantageous. The selective pressure favoring greater learning capacity may not have concerned standard dangers such as avoiding predators or finding food. Some genetic characteristics are selected for primarily along sexual grounds, and may otherwise be actively counter-adaptive (Zahavi & Zahavi, 1997). Byrne & Whiten (1988) have suggested that primate intelligence is driven by adaptation to complex social constraints (see also Cosmides & Tooby, 1992; Whiten & Byrne, 1997). Primates show varying degrees of sophistication for deceiving social authorities and for detecting such deception — skills that may have required exceptional intellectual resources including sophisticated learning.

If the enhanced extent of primate learning is a consequence of sexual selection, then the fact primates exhibit a greater capacity for learning than most other animals does not contradict our hypothesis that learning is inherently dangerous. In some species, such as the Giant Irish Deer and some variants of birds of paradise, over-development of sexually selected traits actively contrary to ordinary fitness may have contributed to the species extinction (Fisher, 1930). If primate learning is special in its extent, and learning is generally a dangerous thing that must be managed, then primate intelligence may well incorporate especially useful attributes for making that learning safer. Thus primate learning is particularly relevant engineering safe learning for

[2]Though note that Hinton & Nowlan (1987) have demonstrated that there is no selective pressure to genetically encode learning which *is* completely reliable. An example here might be retinotopic mapping (von der Malsburg & Singer, 1988).

agents.

Specific Research to be Modeled

We now describe some of the tasks we have selected as good candidates for modeling. In each of these tasks, a subject is presented with the task of retrieving a desirable object. The mechanism for retrieval is not straight-forward for the subject — to be consistently successful they must discover and learn a new strategy. Each of these examples shows that the animals seem to have the capacity to learn the task given a particular context, but will not in some other contexts. The experiments presented below are described more thoroughly by Hauser (1999); further references can be found in that article.

The Object-Retrieval 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. The orientation of the open side is varied from trial to trial. Human infants under the age of approximately 7 months and adult primates of at least one species (cotton-top tamarins) will repeatedly reach straight for the food, despite repeated failure due to contact with the transparent face of the box. More mature children and adult rhesus macaques (another primate species) succeed in this task by finding the open face of the box. For a short intermediate period of human development children learn the task if they are first exposed to an opaque box, and then transferred to the transparent one. This strategy also helps the adult Tamarins.

These experiments demonstrate:

- that some tasks are easier to learn than others (e.g. finding openings in opaque vs. clear boxes),
- that knowledge about some tasks can be relevant to learning others (e.g. finding an opening in a clear box after the skill has been learned with opaque ones), and
- that learning capacities can vary by both age and by species.

The task is interesting because it shows that learning is divided into at least two sub-problems: learning a new skill, and learning when to apply it. In this case, the high salience of the visible reward seems to block the exploratory behavior that *might* find a better solution, but does *not* block the adoption of a relatively certain solution that had been learned in a different framework. Thus modeling the object-retrieval task requires modeling the interaction between perceptually driven motivation and the operation of control plans, as well as modeling the operations of a behavior for controlling learning, one that provides for exploration and for incorporating discoveries into the behavior repertoire.

The Cloth-Pulling Task In this task, tamarins learn to discriminate relevant cues as to which of two pieces of 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, but initially the tamarins don't know this. They are provided with many possible distracting features for determining which cloth to pull, such as the color, texture and shape of the cloth. Tamarins may be fooled into attending to a distractor such as color if it reliably covaries with the right answer, but quickly learn to attend to the food's location on a contiguous piece of cloth in normal circumstances. Yet even after the tamarins successfully show competence at selecting the correct cloth, they can still be fooled into choosing the wrong one. This is done by placing a large but inaccessible reward in one box in contrast to a small but accessible reward in the other. The tamarins are literally tempted into doing the wrong thing — they will pull the cloth associated with the larger reward even though in other circumstances they show knowledge that rewards in such configurations are unobtainable. Consequently, the cloth-pulling task provides further information for the motivation-integration model described as needed for the object-retrieval task.

The Invisible-Displacement 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. In contrast, if the same tube is set up along the horizontal plane, tamarins generate the correct expectations about the invisibly displaced target location. In the vertical case, any learning seems to be blocked by a strong prior expectation — either genetic or experiential — for the effect of gravity. Apparently this bias is so strong that, unlike the comparable situation in the object-retrieval task, in this case the tamarins are unable to use the knowledge from the horizontal condition when they return to attempting the vertical one.

Similar results have been shown in very different research contexts. For example, free-ranging rhesus macaques search below a table rather than on it when they see fruit drop from above the table but are prevented from seeing where it lands. Further, for both experimental conditions, looking-time experiments indicate that primates seem to *expect* the right thing to happen (Santos & Hauser, 2002). For example, if a macaque is constrained from searching but shown both the condition *and* the result (e.g. first witness the food drop as in the earlier experiment, then be shown where it landed), they look longer (demonstrating surprise) if the food is on the ground (where they would search) than if it is on the table. Thus at some level the macaques seem to know what should actually happen, yet they do not appear to have access to this information when planning their own search. Similar results, both for gravitational biases and for contradictions between action and looking time, have been found in human development (e.g. Hood, Carey, & Prasada, 2000).

It is this last task, invisible-displacement, that we have chosen to model first. Like the object-retrieval task, it provides a test of when behavior is integrated as well as how it is learned. In the following sections we describe briefly how we intend to model these tasks using specialized, modular learning, and how these modular specializations can in turn be unified into coherent behavior for a single agent.

Safety, Specialization and Modularity in AI

In the previous section, we described evidence from nature that learning is not unequivocally a good thing. We showed that natural systems tend to limit learning, favoring associations that are likely to be useful, and providing specialized

representations to facilitate the learning of things that cannot be encoded genetically. In this section, we will consider the problems and solutions of learning from a more computational perspective.

Without some sort of constraint (also known as *bias* or *focus*) reliable learning and even action selection are impossible (Wolpert, 1996; Culberson, 1998; Chapman, 1987). This is because the attempt to learn a useful behavior is a form of search, the central problem of agent intelligence (Russell & Norvig, 1995). An agent must be able to 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 an agent with limited time or memory (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 generally provided by the programmer, though here too it may one day be supplemented by social learning (see e.g. Wei & Sen, 1996; Schaal, 1999).

Further, as argued in the introduction, safety and reliability issues for learned behavior are an inescapable consequence of learned behavior's novelty for the agent. In artificial agents, reliable behavior is guaranteed by long processes of verification, whether done formally or by experiment (Gordon, 2000; Bryson, Lowe, & Stein, 2000). The only way novel changes in behavior introduced by individual learning can be guaranteed by these methods is if every possible (or at least likely) 'new' behavior the agent might acquire has also been subject to these verifications. Such a case is most plausible if variation is constrained into a limited range.

One mechanism for providing constraint or bias in an artificial system is to provide specialized representations that support a particular learning strategy or input space. Another is to use a modular decomposition for the system, with each module designed to focus its attention to a particular sort of problem, perception or input. These strategies have been unified in Behavior-Oriented Design (BOD) (Bryson & Stein, 2001b; Bryson, 2001). BOD is the approach we are using as a starting point for our models.

Behavior Oriented Design (BOD)

Behavior-Oriented Design is a development methodology for creating modular intelligence for complete, complex agents — that is, for self-contained agents with multiple, possibly conflicting goals and multiple, possibly mutually exclusive, means for achieving those goals. To date BOD has been applied to the problem of developing intelligence for VR characters, mobile robots, and artificial life simulations.

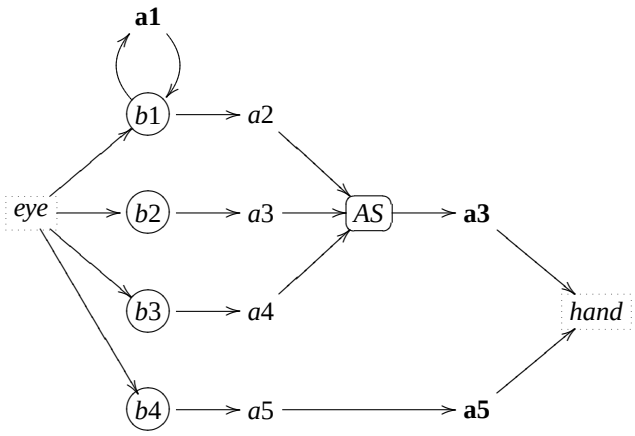

Figure 1: The architecture of a BOD agent. *Behaviors* ($b_1 \ldots$) generate *actions* ($a_1 \ldots$) based on their own perception (derived from *sensing*, the eye icon). Actions which affect state external to their behavior (e.g. *expressed actions*, the hand icon), may be subject to arbitration by *action selection* (AS) if they are mutually exclusive (e.g. sitting and walking). In this diagram, three actions 1, 3 and 5 are taking place concurrently, while 2 and 4 have been inhibited by action selection. Learning takes place *within* behaviors supported by specialized representations. (Bryson & Stein, 2001c)

BOD exploits Object-Oriented Design (OOD) (e.g Parnas, Clements, & Weiss, 1985; Coad, North, & Mayfield, 1997) to improve the power and development ease of Behavior-Based AI (e.g. Brooks, 1991; Arkin, 1998). Contemporary OOD theory puts representations at the center of software modular decomposition, and BOD does the same for AI (Bryson & Stein, 2001b; Bryson, 2000b). In BOD, all of the agent's actions are performed by behavior modules. A behavior module is responsible for:

- the generation of a set of actions,
- any perception necessary for this set of actions, and
- any memory necessary for any learning required to support these perception or actions.

Memory is the permanent variable state which changes when the agent learns something. Under BOD, its amount and the structure of its representation is specialized to task. For example, if we were building a grasping robot using BOD, one behavior would contain state representing visual attention, which would change regularly and rapidly, while another might hold parameters for controlling the robots arm motors, which might be learned only once, or change only slowly as the mechanisms in the robot's arm wear. If the robot also needs to navigate, another behavior might hold a learned map of its environment, and a set of associations between locations and their probable contents.

Modularity and specialized learning simplify the problem of design because they reduce it to a set of simpler subproblems. However, modularity also generates a new set of problems. As with most behavior-based AI, BOD allows all

of its modules to operate in parallel. It is therefore possible that more than one module might recommend an action at the same time. Further, those actions might be contradictory. For example, with the navigating and grasping robot above, the robot's grasping module may want to hold the robot still while it executes a grasp, while its navigation module wants it to move and explore a new region of its environment.

To arbitrate between competing behaviors, BOD uses a special module containing explicit, hierarchically-represented reactive plans. BOD differs from other architectures combining reactive plans and behavior-based AI (e.g. Georgeff & Lansky, 1987; Malcolm & Smithers, 1990; Bonasso *et al.*, 1997; Konolige & Myers, 1998) first by maintaining most of the autonomy for the behavior modules, which play a significant role in action selection, and second, by emphasizing the role of specialized learning by situating all learning within the behavior modules with purpose-built representations (see Figure 1).

Adapting BOD to Primate Learning

We expect that BOD will have to be modified in order to support modeling of the primate learning described above. BOD is intended primarily to support software engineering, and to create efficient, reliable artificial agents. As it currently stands, BOD does not allow for either the addition of new behaviors (and representations) during the "lifetime" of the agent, nor for a change in prioritization between behaviors such as seems apparent in the primate experiments described above. However, we do expect the changes to be relatively minor. For example, the module containing the reactive plans could be made more like the other behavior modules, incorporating learning routines to change the prioritization and contextual criteria that determine when a behavior becomes applied.

Further, the sorts of learning performed by the monkeys are not radical departures from their existing behavior repertoire. The monkeys do not start manipulating objects with their tails or using language or telekinesis. Rather, they operate off a set of primitives already existing with relatively minor modifications. Thus a new module might be learned by cloning a copy of an existing one when it is determined that different parameter sets are useful in different contexts (c.f. Bryson & Stein, 2001c; Demiris, 2001).

Research Program

We have now described the primate research we intend to model (e.g. the invisible-displacement task), the AI methodology we have adopted (BOD) and some of the adaptations we expect to make to it. In this section we describe our particular research focus, our criteria of success, and the expected benefits of our research.

Debugging Monkeys

One of the standard approaches to understanding the underlying mechanism producing a behavior is to look for the limits of that behavior, in other words at the places it fails and the places where it begins to work. The primate learning examples we gave above are interesting precisely because the tamarins can perform them correctly *only in certain circumstances*. We are particularly interested then in the circumstances when the tamarin *fails* to perform a learning task, given that we know that the fundamental task itself is within the animal's capacity.

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

1. it might not know (be able to produce) the appropriate behavior, so this behavior is not truly an option for the animal, 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. Mitchell, 1997) 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.

Criteria for Success in Modeling

There are two criteria for success in modeling natural intelligence; we would of course like to achieve both. One is for the performance profile of the average software agent to be within tolerance of the composite average performance of the monkey subjects. The other is to be able to replicate each individual's learning history. Statistically, the latter is slightly easier to demonstrate because there are significantly more datapoints in terms of trials per individual than there are numbers of individuals. It also allows for the possibility that the monkey subjects are not all exploiting the same learning strategy. In this case, one could postulate both a number of strategies and their distribution across subjects, and with this both account for and replicate the composite results. This strategy has been well illustrated by Harris & McGonigle (1994).

Expected Benefits of this Research

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.

Secondary benefits are expected to be realized in several domains:

- We expect the models will help us test current hypotheses for monkey learning. Indeed, the architectural idioms will be evaluated on their predictive value. This will of course also serve as an advance in primatology.
- We intend to deliver results not only in terms of detailed program code, but also in general terms for adaptation into a variety architectures: this is what we mean by 'architectural idioms'. We have already demonstrated this technique (Bryson & Stein, 2001a). We also hope to further refine our innovations into design patterns (Gamma *et al.*, 1995), which would then generalize the applicability further to any object-oriented software design.
- We will also be attempting to develop specialized tools usable by non-programmers, specifically primatologists. These tools will be useful not only within the laboratory, but also in the classroom. Constructive models have high explanatory and pedagogical power because they are open to thorough examination.

Discussion

As we discussed in the introduction, reactive planning has been one technological response of the AI community to the problem of robust, reliable behavior in real-time agents. Reactive artificial intelligence is analogous to genetically-determined behavior in animals. At a first approximation, both systems are considered to rely on instructions *provided* to the agent as part of its fundamental makeup, and to be independent of any learning or deliberation by the agent itself.

Research has shown, however, that purely reactive intelligence is very limited. Variable state and learning are ubiquitous in natural intelligence and generally necessary or at least extremely useful in AI (Hexmoor, Horswill, & Kortenkamp, 1997; Kortenkamp, Bonasso, & Murphy, 1998; Bryson, 2000a). Even in reactive AI, well-ordered behavior in complex agents (e.g. those capable of pursuing multiple, potentially-conflicting goals), generally incorporates stored state recording recent decisions to focus attention on a subset of possible actions (Newell, 1990; Gat, 1998).

The term *learning* is usually applied not to such transient changes in the state, but to mechanisms that have lasting impact on agent behavior. Nevertheless, the temporal extent of this 'lasting' in natural systems varies widely. One can easily talk about learning something on one day (or even at one minute) and having forgotten it the next. Learning in nature is not only constrained temporally: it is also dependent on context, complexity (both of stimulus and action), previous experience and other individual factors. One explanation for these constraints is that they allow animals to safely and reliably incorporate experience-based variation into their behavior repertoire, yet remain immune to possibly distracting experiences that could alter the response repertoire for the worse.

We would like to highlight the applicability of our research to the concerns of the safe-learning community as spelled out in the aims of this symposium. We have suggested that modular architectures are well-suited to providing the sort of constrained, specialized learning that nature seems to utilize. This is true only if learning is itself modularized. This has not been true in BDI architectures such as PRS (Georgeff & Lansky, 1987), which use a single database for learning, and is unusual[3] in implementations of the subsumption architecture (Brooks, 1991) due to proscriptions against learning. However, object-oriented design supports this system, and and multi-agent learning systems might also be conducive.

We have mentioned briefly here and discussed at length elsewhere (e.g. Bryson & Stein, 2001b) the interaction between such modular learning and reactive planning. Gat (1998) also describes a compatible though not identical vision of this relationship. We also mentioned briefly the implications of societies of agents to safe learning. This is an area we have not yet explored experimentally, but may over the course of our research. For social animals, it may be useful for individuals to have different strategies or biases particularly for exploration and learning. This is particularly true in the case where the animals also posses social learning, since an entire family group can learn from one animal's success.

Summary

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. In this paper, we have reviewed evidence from both natural and artificial intelligence that the only solution is to provide constraints to learning.

We have presented a research program which is approaching the problem of safe learning by working from an animal model. At least to begin, 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.

Acknowledgments

We would like to thank Will Lowe and Laurie R. Santos for their assistance with this document. This work was funded by National Science Foundation (NSF) grant EIA-0132707.

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.

[3] But not absent, (see e.g. Matarić, 1997).

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.

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.

Brainard, M. S., and Knudsen, E. I. 1998. Sensitive periods for visual calibration of the auditory space map in the barn owl optic tectum. *Journal of Neuroscience* 18:3929–3942.

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

Bryson, J., and Stein, L. A. 2001a. Architectures and idioms: Making progress in agent design. In Castelfranchi, C., and Lespérance, Y., eds., *The Seventh International Workshop on Agent Theories, Architectures, and Languages (ATAL2000)*. Springer.

Bryson, J., and Stein, L. A. 2001b. 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. 2001c. 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.; Lowe, W.; and Stein, L. A. 2000. Hypothesis testing for complex agents. In *NIST Workshop on Performance Metrics for Intelligent Systems*.

Bryson, J. 2000a. Cross-paradigm analysis of autonomous agent architecture. *Journal of Experimental and Theoretical Artificial Intelligence* 12(2):165–190.

Bryson, J. 2000b. Making modularity work: Combining memory systems and intelligent processes in a dialog agent. In Sloman, A., ed., *AISB'00 Symposium on Designing a Functioning Mind*, 21–30.

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.

Cosmides, L., and Tooby, J. 1992. Cognitive adaptations for social exchange. In Barkow, J. H.; Cosmides, L.; and Tooby, J., eds., *The Adapted Mind*. Oxford University press. 163–228.

Culberson, J. C. 1998. On the futility of blind search: An algorithmic view of "no free lunch". *Evolutionary Computation* 6(2):109–128.

Demiris, Y. 2001. Imitation as a dual-route process featuring predictive and learning components: a biologically-plausible computational model. In Dautenhahn, K., and Nehaniv, C., eds., *Imitation in Animals and Artifacts*. Cambridge, MA: mitpress. chapter 13.

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.

Fisher, R. A. 1930. *The Genetical Theory of Natural Selection.* Oxford-University-Press.

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.

Gamma, E.; Helm, R.; Johnson, R.; and Vlissides, J. 1995. *Design Patterns*. Reading, MA: Addison Wesley.

Garcia, J., and Koelling, R. A. 1966. The relation of cue to consequence in avoidance learning. *Psychonomic Science* 4:123–124.

Gat, E. 1998. Three-layer architectures. In Kortenkamp, D.; Bonasso, R. P.; and Murphy, R., eds., *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. Cambridge, MA: MIT Press. chapter 8, 195–210.

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.

Gordon, D. 2000. APT agents: Agents that are adaptive, predictable, and timely. In *Proceedings of the First Goddard Workshop on Formal Approaches to Agent-Based Systems (FAABS'00)*.

Harris, M. R., and McGonigle, B. O. 1994. A modle of transitive choice. *The Quarterly Journal of Experimental Psychology* 47B(3):319–348.

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

Hertz, J.; Krogh, A.; and Palmer, R. G. 1991. *Introduction to the Theory of Neural Computation*. Redwood City, CA: Addison-Wesley Publishing Company.

Hexmoor, H.; Horswill, I.; and Kortenkamp, D. 1997. Special issue: Software architectures for hardware agents. *Journal of Experimental & Theoretical Artificial Intelligence* 9(2/3).

Hineline, P. N., and Rachlin, H. 1969. Escape and avoidance of shock by pigeons pecking a key. *Journal of Experimental Analysis of Behavior* 12:533–538.

Hinton, G. E., and Nowlan, S. J. 1987. How learning can guide evolution. *Complex Systems* 1:495–502.

Hood, B.; Carey, S.; and Prasada, S. 2000. Predicting the outcomes of physical events: Two-year-olds fail to reveal knowledge of solidity and support. *Child Development* 71(6):1540–1554.

Konolige, K., and Myers, K. 1998. The saphira architecture for autonomous mobile robots. In Kortenkamp, D.; Bonasso, R. P.; and Murphy, R., eds., *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. Cambridge, MA: MIT Press. chapter 9, 211–242.

Kortenkamp, D.; Bonasso, R. P.; and Murphy, R., eds. 1998. *Artificial Intelligence and Mobile Robots: Case Studies of Successful Robot Systems*. Cambridge, MA: MIT Press.

Malcolm, C., and Smithers, T. 1990. Symbol grounding via a hybrid architecture in an autonomous assembly system. In Maes, P., ed., *Designing Autonomous Agents: Theory and Practice from Biology to Engineering and Back*. Cambridge, MA: MIT Press. 123–144.

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

Matarić, M. J. 1997. Behavior-based control: Examples from navigation, learning, and group behavior. *Journal of Experimental & Theoretical Artificial Intelligence* 9(2/3):323–336.

Meyer, J.-A., and Wilson, S., eds. 1991. *From Animals to Animats: Proceedings of the First International Conference on Simulation of Adaptive Behavior*. Cambridge, MA: MIT Press.

Mitchell, T. 1997. *Machine Learning*. McGraw Hill.

Newell, A. 1990. *Unified Theories of Cognition*. Cambridge, Massachusetts: Harvard University Press.

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.

Russell, S. J., and Norvig, P. 1995. *Artificial Intelligence : A Modern Approach*. Englewood Cliffs, NJ: Prentice Hall.

Santos, L. R., and Hauser, M. D. 2002. A non-human primate's understanding of solidity: Dissociations between seeing and acting. *Developmental Science* 5. in press.

Schaal, S. 1999. Is imitation learning the route to humanoid robots? *Trends in Cognitive Sciences* 3(6):233–242.

Schneider, J. G. 1997. Exploiting model uncertainty estimates for safe dynamic control learning. In Mozer, M. C.; Jordan, M. I.; and Petsche, T., eds., *Advances in Neural Information Processing Systems*, volume 9, 1047. The MIT Press.

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

von der Malsburg, C., and Singer, W. 1988. Principles of cortical network organization. In Rakic, P., and Singer, W., eds., *Neurobiology of Neocortex*. John Wiley. 69–99.

Wei, G., and Sen, S., eds. 1996. *Adaptation and Learning in Multi-Agent Systems*. Springer.

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

Wolpert, D. H. 1996. The lack of A priori distinctions between learning algorithms. *Neural Computation* 8(7):1341–1390.

Zahavi, A., and Zahavi, A. 1997. *The Handicap Principle: A missing piece of Darwin's puzzle*. Oxford: Oxford University Press.