Primate Errors in Transitive Inference

Joanna J. Bryson and Jonathan C. S. Leong

Harvard Primate Cognitive Neuroscience Laboratory
Cambridge, MA 02138
jbryson@wjh.harvard.edu, jleong@fas.harvard.edu

August 27, 2002

Abstract

This paper provides a new model of primate transitive inference learning. We begin from the production-rule model of Harris and McGonigle (1994), which was previously the best model at accounting for the performance of squirrel monkeys (*Saimiri sciureus*) and human children trained on transitive inference (e.g. given $A > B$ and $B > C$, then $A > C$) when presented with *three* items from the training set. We have extended this model by creating a version that learns, and in so doing have accounted for more of the data, particularly subject's failures to learn transitive inference. We demonstrate a two-tier task-learning model which, in accounting for learning the training pairs, generates 'transitive inference' in pairs (but not triads) for free. In addition to replicating this aspect of the Harris and McGonigle (1994) model in a learning system, we provide a novel explanation for the difficulty in acquiring the pairs in the first place and the utility of the training regimes commonly used. We discuss the biological plausability of our model and its relationship to other popular models of the transitive inference task.

1 Introduction

Transitive inference (TI) is the process of reasoning whereby one deduces that if, for some quality, $A > B$ and $B > C$, then $A > C$. In some domains, such as integers, this property will hold for any A, B or C. For other domains, such as sporting competitions and primate dominance hierarchies, the property does not necessarily hold. For example, international tennis rankings, while in general being a good predictor of the outcome of tennis matches, may systematically fail to predict the outcome of two particular players.

Many animals demonstrate TI without special training once they learn the ordered set of adjacent pairs (e.g. $A > B$, $B > C$). There have been many attempts to explain this phenomena via modeling (Wynne, 1998, for a review). However, very few models account for the performance of subjects presented with *trigrams*, the choice between three rather than just two items from the trained pairs (e.g. A, B, C). TI performance in this case is systematically degraded both from the two-item case and from what standard TI models predict. Of the models that do account for this data, that of Harris and McGonigle (1994) provide the current best fit to data. The Harris and McGonigle (1994) model is in the form of a standard artificial intelligence learning representation called a *production-rule stack*. If an agent learns a rule stack that works correctly on all pairs, then TI for two items is expressed as a simple consequence of the underlying pair representation. Additionally, the Harris and McGonigle (1994) model accounts for degraded three-item performance data, providing matches for both aggregate data and individual performance.

In this paper we present a new model which extends Harris and McGonigle (1994) to include learning. To model the data accounted for by Harris and McGonigle (1994), we split the task of TI into two problems: learning an association between stimuli and actions; and prioritizing stimuli. This two-tier learning mechanism is able to perfectly learn the Harris and McGonigle (1994) TI model. In addition, the two-tier model demonstrates three interesting and unintended correlates to real subjects:

- Human children and monkeys require training in order to reliably learn TI. The two-tier model does not reliably learn TI without a similar type of training.
- The performance of the two-tier model was more robust than that of the standard single-tier model of TI. That is, TI learning in the two-tier model is stable over a larger set of training parameters.
- When the two-tier model fails to learn TI, it makes errors similar to those observed in primates.

In the following sections we summarize the current findings on transitive inference, in particular the Harris and McGonigle (1994) model. We then detail the two-tier model and present our experimental methods and results. Finally we discuss the implications of the two-tier model, including neural correlates and implications for more general-purpose task learning.

1.1 Transitive Inference (TI)

Piaget described TI as an example of concrete operational thought (Piaget, 1954). That is, children become capable of doing TI when they become capable of mentally performing the physical manipulations they would use to determine the correct answer. In the case of TI, this manipulation involves ordering the objects into a sequence using the rules $A > B$ and $B > C$, and then observing the relative location of A and C.

Since the 1970's, however, demonstrations of TI in children and a variety of animals — apes, monkeys, rats and pigeons (Wynne, 1998, for a review) — have suggested that TI is a basic animal competence, not a skill involving sequential operations or explicit logic. Further, Siemann and Delius (1993) have shown that in adults who learned to choose between pairs of doors during an exploration-type computer game, there was no performance difference between individuals who formed explicit comparison models and those who did not ($N = 8$ vs. 7 respectively).

1.2 Characteristic TI Effects

The following effects hold across experimental subjects, whether they are children, monkeys, rats, pigeons, or adult video-game players. Once a subject has learned the ordering for a set of pairs (e.g. AB, BC, CD...), they tend to show significantly above-chance results for transitive inference. They also show the following effects (see Wynne, 1998, for review):

- *The End Anchor Effect*: subjects make an evaluation faster and more accurately when a test pair contains one of the ends. This is usually explained by the fact they have learned only one behavior with regard to the end points (e.g. nothing is $> A$).

- *The Serial Position Effect*: even taking into account the end anchor effect, subjects do more poorly the closer to the middle of the sequence they are.

- *Symbolic Distance Effect*: even compensating for the end anchor effect, the further apart on the series two items are, the faster and more accurately the subject makes the evaluation. This effect contradicts any step-wise chaining model of transitive inference (i.e. Piaget's concrete operations) since distant items would require *more* steps and therefore a longer reaction time (RT).

1.3 Training Subjects for TI

Training a subject to perform transitive inference is not trivial. Subjects are trained on a number of ordered pairs, typically in batches. Because of the end anchor effect, there must be at least five items ($A \ldots E$) to clearly demonstrate transitivity on just one untrained pair (BD). Obviously, seven or more items would give further information, but training for transitivity is notoriously difficult. Even children who can master five items often cannot master seven. This is true even for simple sorting of conspicuously ordered items such as posts of different lengths (McGonigle and Chalmers, 1996). Individual items are generally labeled in a deliberately non-ordinal way, such as by color or pattern, and controlled by varying the assignment of rank by subject (e.g. one subject may learn $blue < green < brown$ while another $brown < blue < green$).

The subjects are first taught to use the testing apparatus; they are presented with an object and rewarded for selecting it. Next, they are trained on the first pair DE, where only one element, D is rewarded[1] When the subjects reach criterion, they are trained on CD. After all pairs are trained, there is generally a phase of ordered repeated training on all the pairs, but with fewer exposures per pair, which is then followed by a period of random presentations of training pairs (See Figure 1).

Once a subject has been trained to criterion, they are exposed to testing pairs. In testing, either choice is rewarded in an effort to minimize the effects of further training. However, training (adjacent) pairs are often interspersed with testing pairs during the testing phase, with the training pairs still being differentially rewarded.

1.4 Standard Models of TI

The currently dominant way to model transitive inference is in terms of the probability that an element will get chosen. Probability is determined by a weight associated with each element. During training, the weight is updated using standard reinforcement learning rules. Normalization of weights across rules is necessary to get results invariant on the order of training (whether AB is presented first or DE). Taking into account some pair-specific context information has proven useful for accuracy, and necessary to explain the fact that animals can also learn a looping end pair (e.g. $E > A$ for 5 items). Wynne (1998) provides the following as a summary model:

[1] The psychological literature is not consistent about whether A or E is the 'higher' (rewarded) end. This paper uses A as high.

P1	Each pair in order (ED, DC, CB, BA) repeated until 9 of 10 most recent trials are correct. Reject if requires over 200 trials total
P2a	4 of each pair in order. Criteria: 32 consecutive trials correct. Reject if requires over 200 trials total
P2b	2 of each pair in order. Criteria: 16 consecutive trials correct. Reject if requires over 200 trials total
P2c	1 of each pair in order. Criteria: 30 consecutive trials correct. No rejection criteria.
P3	1 of each pair randomly ordered. Criteria: 24 consecutive trials correct. Reject if requires over 200 trials total
T1	Bigram tests: 6 sets of 10 pairs in random order. Reward unless failed training pair.
T2a	As in P3 for 32 trials. Unless 90% correct, redo P3.
T2	Trigram tests: 6 sets of 10 trigrams in random order, reward for all.
T3	Extended version of T2.

Table 1: Phases of training and testing used for children (Chalmers and McGonigle, 1984). Monkeys are actually easier to train. Trigram testing is described in Section 1.5

$$p(X|XY) = \frac{1}{1+e^{-\alpha(2r-1)}} \quad (1)$$

$$r = \frac{V(X)+V(Z)+\gamma V(\langle X|XY\rangle)}{V(X)+V(Y)+2V(Z)+\gamma V(\langle X|XY\rangle)+\gamma V(\langle X|XY\rangle)}$$

where $V(X)$ is the value of stimulus X, $V(\langle X|XY\rangle)$ indicates extra value for X given also the context XY, Z is a normalizing term, and α and γ are free parameters. Such models assume that reaction time somehow co-varies with probability of correct response.

1.5 The Binary Sampling Model

In one of the earliest of the non-deductive TI papers, McGonigle and Chalmers (1977) not only demonstrated animal learning of TI, but also proposed a model to account for the errors the animals made. Their subjects were squirrel monkeys (*Saimiri sciureus*). Like human children, these monkeys tend to score only around 90% on the pair BD. To explain this, McGonigle and Chalmers proposed the *binary-sampling theory*. This theory assumes that:

- Subjects consider not only the items visible, but also items that might be *expected* to be visible. That is, they take into account elements associated with the current stimuli, especially intervening stimuli associated with both.
- Subjects consider only two of these possible elements, choosing the pair at random. If they were trained on that pair, they perform as trained; otherwise they perform at chance, unless one of the items is an end item, A or E, in which case they perform by selecting or avoiding the item, respectively.
- If the subject chooses an item that is only expected, not actually present, it obviously cannot act on that selection (e.g. grasp the item). However, selection reinforces consideration of that item, which makes it likely the next pair the animal considers includes one of the higher-valued of the items displayed.

Thus for the pair BD, this model assumes an equal chance the monkey will focus on BC, CD, or BD. Either established training pair results in the monkey selecting B, while the pair BD results in an even (therefore 17%) chance of either element being chosen. This predicts that the subjects would select B about 83% of the time, which is near to the average actual performance of 85%.

The binary-sampling theory is something of a naïve probabilistic model: it incorporates the concept of expectation, but not in a full-fledged probabilistic framework. More importantly, it contradicts the symbolic distance effect, since again further-apart pairs may require more operations. However, it motivated McGonigle and Chalmers to generate a data set showing the results of testing monkeys (and later children (Chalmers and McGonigle, 1984)) on *trigrams* of three items. The binary-sampling theory predicts that for the trigram BCD there is a 17% chance D will be chosen (half of the times BD is attended to), a 33% chance C will be chosen (all of the times CD is attended to) and a 50% chance of B (all of BC plus half of BD). True TI would of course predict 0%, 0% and 100%. In fact, the monkeys showed 3%, 36%, and 61%, respectively.

1.6 The Production-Rule Stack Model

The trigram data has since been used by Harris and McGonigle (1994) to create a better-fitting model, which is based on the concept of a stack of production rules. This model matches the performance of the monkeys very well whether they are modeled as a group or as individuals.

The production-rule stack model requires the following assumptions:

1. The subject learns a set of rules of the nature "if A is present, select A" or "if D is present, avoid D".
2. The subject learns a prioritization of these rules.

This process results in a *rule stack*, where the first rule is applied if its trigger finds the context appropriate. If not, the second, and so on.

For an example, consider a subject that has learned a stack like the following:

1. (A present) $\Rightarrow$ select A
2. (E present) $\Rightarrow$ avoid E
3. (D present) $\Rightarrow$ avoid D
4. (B present) $\Rightarrow$ select B

Here the top item (1) is assumed to have the highest priority. If the subject is presented with a pair CD it begins working down its rule stack. Rules 1 and 2 do not apply, since neither A nor E is present. However, rule 3 indicates the subject should avoid D, so consequently it selects C. Priority is critical. For example, for the pair DE, rules 2 and 3 give different results. However, since rule 2 has higher priority, D will be selected.

Adding the assumption that in the trigram test cases, an 'avoid' rule results in random selection between the two remaining items, Harris and McGonigle (1994) model the conglomerate monkey data so well that there is no significant difference between the model and the data. For example, over all possible trigrams, the stack hypothesis predicts a distribution of 0, 25% and 75% for the lowest, middle and highest items. Binary sampling predicts 3%, 35% and 63%, and logic of course 0%, 0% and 100%. The monkeys showed 1%, 22% and 78%. Further, individual performance of most monkeys were matched to a particular stack.

Without trigram data, there would be no way to discriminate which rule set the monkeys use. However, with trigram data, the stacks are distinguishable because of their errors. For example, a stack like

1′. (A present) $\Rightarrow$ select A
2′. (B present) $\Rightarrow$ select B
3′. (C present) $\Rightarrow$ select C
4′. (D present) $\Rightarrow$ select D

would always select B from the trigram BCD by using rule 2′, while the previous stack would select B 50% of the time and C 50% because it would base its decision on rule 3.

There are 8 discernible correct rule stacks of three rules each which will solve the initial training task. There are actually 16 correct stacks of four rules, but trigram experiments cannot discriminate whether the fourth rule selects or avoids.

2 Model

Although Harris and McGonigle (1994) account amazingly well for the trigram data, their production-rule model is somewhat unsatisfying, because its components seem artificial. Why would subjects prioritize productions rather than items, particularly if there are the same number of each? Why would a rule be triggered by only a single stimulus rather than by both items of a trained pair? Why are their only two possible actions, select and avoid, one of which (avoid) has an arbitrary outcome? In order to better understand the model and to further test it for biological plausibility, we have built a learning version of their system.

Consistent with the production-rule hypothesis, we built a two tier system: one tier to learn the appropriate rule associated with each item, and one to learn a prioritization between items.

The Harris and McGonigle (1994) model provides for only one rule per item and only two possible rules. Our model consists of a list of items seen, each associated with a weight and with a unique 2-item list of possible rules (select or avoid) also associated with weights. When the agent sees a new item, it adds it to its list of stimuli. The biological plausibility of these assumptions is addressed in the discussion.

All of the weights in a single list are normalized (they sum to 1) — new items receive the weight $1/N$ (where N is the current number of items). New items in the stimulus list are also associated with a rule-list where both *select* and *avoid* have weight 0.5. provided. When the agent is given a trial, it examines its list for the rules associated with the stimuli present. For that set of rules, the highest-priority stimulus determines the focus of the agent's attention. Then the highest priority rule in the rule-list associated with that stimuli determines whether the agent selects or avoids the object it has fixated on. In the case of a tie, outcome is arbitrary.

Weights are updated after every trial, where a trial is a presentation of a pair or trigram, a selection by the agent, and a reward. In contrast to the complex update rule in eq. 1, we use a simple step function. The subject learns a weight vector $\mathbf{v}$. For the pair XY, where X is selected by the subject and $\mathbf{v}_X$ and $\mathbf{v}_Y$ are the weights associated with X and Y respectively:

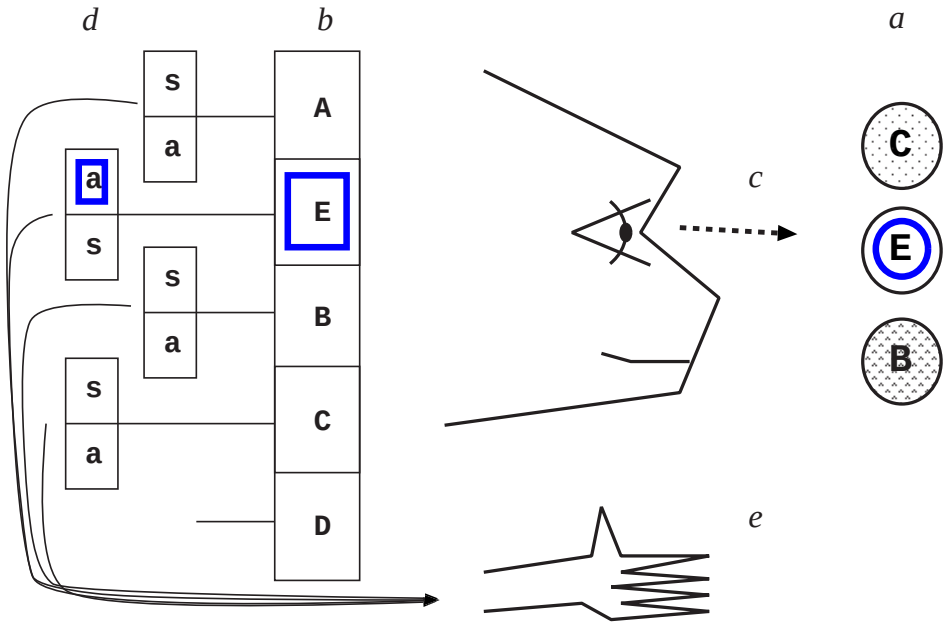

Figure 1: The two-tier model. When the agent observes a set of stimuli (a), a weight vector (b, the first tier) determines which item present is most salient. This attracts visual attention (c) and determines which rule vector (d, the second tier) selects the appropriate action (**s**elect or **a**void) which controls the monkeys grasp (e). The two vectors that were most recently active (a and one of d) are then updated as determined by the result and the learning rule (Equation 2).

$$\begin{aligned} &\text{If X is correct and } (\mathbf{v}_X - \mathbf{v}_Y < \Xi), \text{ add } \delta \text{ to } \mathbf{v}_X; \\ &\qquad \text{else, if X is incorrect, subtract } \delta \text{ from } \mathbf{v}_X. \end{aligned} \quad (2)$$

where Ξ and δ are free parameters. Ξ is a threshold over which reward is so expected that it no longer prompts learning. δ is the amount a weight is changed by a single bout of learning. If a weight-change occurs, $\mathbf{v}$ is subsequently renormalized.

3 Methods

3.1 Model Testing through Artificial Life Simulation

The experiments in this paper were performed in artificial life (ALife) simulations. ALife allows a researcher to evaluate algorithms that are too complex to analyze formally. ALife evaluations operate by running simulations, then performing standard hypothesis testing to see whether the simulated results are a good match to the original data.

Once a model has been built, the process of simulation also allows one to search broad parameter spaces that would be relatively difficult or expensive to test in the laboratory. These runs then serve as predictions from the model, and a relatively sparse set of them can be tested against real-world experiments to provide further validation.

Further validation of an ALife model occurs when simulation results unexpectedly converge with previously-observed, real-world phenomena. This is the case for the experiments described below. The models were built to test whether a biologically-plausible extension of the Harris and McGonigle (1994) production-rule model could learn the productions and their prioritization. That the model not only learns but also unexpectedly reproduces two aspects of error-production by real animals is, we feel, a significant result.

3.2 Details of the Simulation

The ALife agents were written in the Common Lisp Object System (CLOS) (Steele Jr., 1990), a standard programming language frequently used for artificial intelligence. The code was developed under the Behavior Oriented Design methodology, developed by one of the authors (Bryson and Stein, 2001; Bryson, 2001), and implemented on top of a graphical software development environment for CLOS called LispWorks (Xanalys, 2001). A personal edition of LispWorks is available for free download from its manufacturers, and all software used in this paper is available from the authors by email. Develop-

ment took place in a Linux environment (Red Hat, 2001) and experiments were run under both Linux and Windows (Microsoft, 2001).

The artificial intelligence (AI) program that runs the simulations controls not only the agents, but also the operation of the test apparatus including the recording of results. The program is modular, and the knowledge in the modules representing different real-world agents (the subjects, apparatus and operator) is kept isolated from the other agent's modules.

On each trial, the training apparatus generates an n-gram (either bigram or trigram) as appropriate to the current phase of the experiment and places its element in the test-board. The learning agent (or subject) then selects one of the options by transferring it from the test-board to its hand. The apparatus determines whether the subject chose correctly and provides reinforcement with either a 'peanut or a 'buzzer token. It also records the results, and ends the test by clearing the testing area. When the subject is presented with the test-board it selects an object as illustrated in Figure 1. When it is given reinforcement, it applies the learning rule from Section 2 based on its current variable state (its expectations) and the presence or absence of the peanut reward.

In the first two conditions described below, an indefinite number of trials were run until the simulation was terminated by the human experimenter. In the third condition, the apparatus part of the agent was enhanced to run the training and testing procedure shown in Table 1.

Details of the construction of the ALife simulation are described by Bryson (2001). In order to support the analysis in this paper we have extended the simulation somewhat, but not in ways significant to the model.

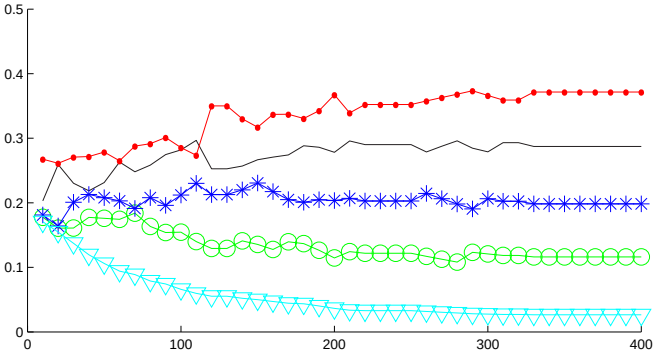

Figure 2: Typical results for single-vector learning (only one rule — select). X-axis: trial number; Y-axis: weights of stimuli vector (sum to one). Free variables are set to Parameters: $\Xi = .08$, $\delta = .02$. Key: $A\bullet$, $B-$, $C*$,$D\bigcirc$, $E\bigtriangledown$.

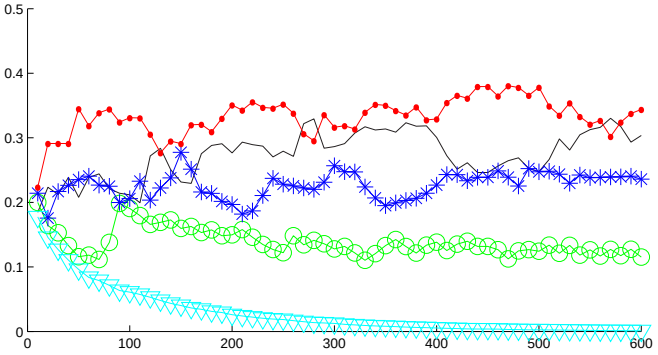

Figure 3: Single vector with 'stupider' parameters: $\Xi = .12$, $\delta = .02$. This fails to find a stable solution thus occasionally giving wrong responses. Key: $A\bullet$, $B-$, $C*$,$D\bigcirc$, $E\bigtriangledown$.

4 Results

4.1 Single-Vector (Standard Model) Results

Before completely implementing the two-tier model (see Section 2), we first tested our simulation and learning rule with a single-tier model. This provides a rough replication of the standard, single-vector approach (see Section 1.4). Figure 2 shows a typical result for standard transitive inference modeling, when pairs have been presented in random order. Unlike real animals (including children), these models perform well regardless of training regime. Learning on this task is twice as fast and more reliable if the agent is initially presented only with training pairs than if it is trained on every possible pair, presumably because the training pairs provide more information on the border conditions.

If $\Xi > 0.1$, then a stable solution for five items cannot be reached (see Equation 2 above). Because learning doesn't stabilize, the model is open to a 'hot hands'-like phenomena (Gilovich et al., 1985). When there is a chance reiteration of one particular pair, the higher element of that pair can accumulate so much reinforcement that its weight surpasses the element above it. Thus the agent over-estimates the value of an item because of its recent "winning streak". This is illustrated in Figure 3. These results provide one possible explanation for individual differences in transitive task performance. Individual differences in stable discriminations between priorities can affect the number of items that can be reliably ordered.

On the other hand, in the single-vector model the unstable competing weights often represent the two top-priority stimuli. This violates the End Anchor Effect, so is not biologically plausible. However, the other end of the vector is reliably well-discriminated. Consequently, the model might be fixed if instead of using a single vec-

tor to represent the learned priority sequence we used a dual-vector representation such as the Start-End model proposed by Henson (1998) to reliably anchor each end. Even so, this model would not explain performance on trigram testing as well as Harris and McGonigle (1994) have.

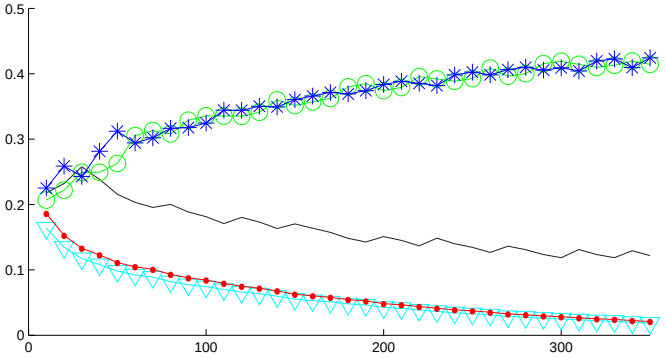

Figure 4: Rule learning with no training regime. Rules learned in descending order of priority: select $D\bigcirc$, avoid $C*$, avoid $B-$. The system cannot stabilize because it is far from a complete solution, but it behaves correctly for every training pair except CD. Parameters: $\Xi = .08$, $\delta = .02$.

4.2 Rule Learning Without Phased Training

The two-tier model is an extension of the single-tier model which learns not just priorities for stimuli, but also connections between stimuli and actions. Figures 4 and 5 illustrate outcomes for the system with additional vectors provided for representing rules, as described in Section 2. Rule selection is made very early in training and remains stable once established, so is not represented in most figures except in the captions. Nevertheless, the added complication of rule learning defeats the simple learning regime used in the earlier experiments. Without a training regime, the multi-vector model generally learns either the solution shown in Figure 4 or a symmetric one with the select (B) and avoid(C) fighting for top priority. These agents visually neglect the two endpoints, but get the end pairs (AB and DE) correct 100% of the time by associating the rule *avoid* with B and *select* with D. However, this leaves them with no possible solution for correctly learning both middle pairs. The learning system fixates on trying to solve an impasse in the middle of the sequence, always failing to learn one pair.

About one time in four agents without a training regime do learn a correct solution. We tested learning across a range of parameter values: every combination of Ξ drawn from $.08, .1, .12, .14$ and δ drawn from $.01, .02, .04, .08, .12, .16$. Correct solutions came evenly from all parameter values, and seemed evenly distributed across all possible correct solutions enumerated by Harris and McGonigle (1994) (Table 2).

A correct stack must be either in sequence $[s(A)s(B)s(C)]$ or $[a(E)a(D)a(C)]$ or an ordered cross of these (e.g. $[s(A)a(E)s(B)]$). See Table 2 for a complete list. Although the rules learned by the average (failing) agents without a training regime have a very different order, they still perform fairly well. Only one training pair is incorrect: that of the two top priority stimuli. For example, in Figure 4, the only training pair which cannot be handled is BC. Fig. 4 *does* display the End Anchor Effect. Agents quickly choose rules which avoid making errors on the two end points, but then are in a configuration where they cannot learn a rule to complete the solution. Whether +B or -C is highest priority, the outcome is exactly the same. These solutions also display something of the Serial Position Effect by confusing only central pairs.

It would be interesting to compare these results with the outcomes of the real subjects who fail to meet criterion in learning transitive inference. Although no trigram results were reported for monkeys or children that missed criterion, one monkey subject, Blue, passed criterion but still showed a consistent error between the 3^{rd} and 4^{th} item Harris and McGonigle (1994, p. 332). Blue's errors are in keeping with the results of this model.

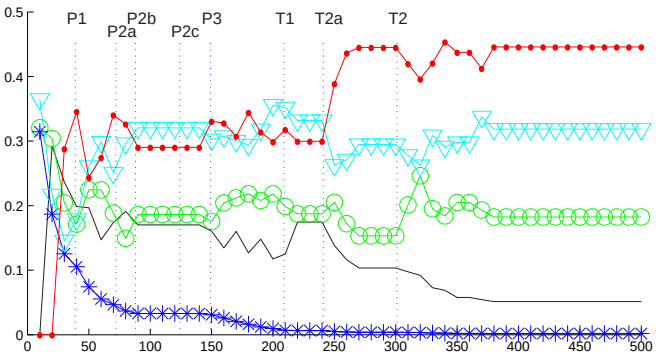

Figure 5: Rule Learning with Phased Training. Labeled lines indicate the *end* of training and testing phases (see Table 1). This agent arrives at different stable solutions at different points, but they are all correct. Here the rules are: select $A\bullet$, avoid $E\bigtriangledown$, avoid $D\bigcirc$. The agent succeeds with very 'stupid' parameters: $\Xi = .12$, $\delta = .06$.

4.3 Rule Learning with Phased Training

Using the training regime for real children (Table 1), the full model can learn far more reliably (Table 2, example in Figure 5).

Correct Stacks	No Regime (random)	Regime starting w/ E, D after training	Regime starting w/ E, D after testing	Regime starting w/ A, B after training	Regime starting w/ A, B after testing	Harris & McGonigle
s(A) s(B) s(C)	4	51	41	–	–	–
s(A) s(B) a(E)	8	68	26	–	–	–
s(A) a(E) a(D)	3	–	1	4	2	2
s(A) a(E) s(B)	4	4	16	3	1	2
a(E) a(D) s(A)	5	–	1	57	50	–
a(E) a(D) a(C)	4	–	–	59	47	1
a(E) s(A) a(D)	5	3	–	4	11	–
a(E) s(A) s(B)	1	1	13	–	3	–
Total Correct	34	127	98	127	114	5
Total	144	144	144	144	144	7

Table 2: Production-rule stack equivalents to solutions by monkey subjects and by two-tier models undergoing various forms of training. The distribution of solutions is strongly determined by the order training pairs are presented.

The multi-vector model is somewhat more like the squirrel monkeys than the children in that it more reliably learns TI than children given the same training regime. Nearly all successful agents converge quickly, and the ones that fail to learn fail early, usually during Phase 2a. The rule representation also allows for stable learning with either more items or larger values of Ξ. This is because some rules will never be compared. Notice for example that during training, rules involving non-adjacent items do not need to discriminate their weights since they are never compared (Figure 5, P2b–P3).

Large values of δ result in less systematic convergence, and provide a greater diversity of rule sets, though a few more failures to reach criterion. Using the same array of parameters as in Section 4.2, shows approximately 7 in 8 agents learning successfully. Convergence in this case is highly dependent on δ; when δ was in $.08, .12, .16$, one in four agents failed, otherwise there were only a very few failures (3), all of which had $\delta = .01$ ($N_{\delta=.01} = 48$). A very low δ results in a slow learners. Even when such agents make criterion, they may never find a stable solution (see e.g. Figure 6).

Note that significant learning occurs during testing — another phenomena also reported with real monkeys (Harris and McGonigle, 1994). Learning occurs because rules that previously were never compared (e.g. those triggered by any two non-adjacent items) will be now. If their weights do not already happen to be at least Ξ apart, learning is triggered, regardless of the reinforcement scheme. This explains the utility of continuing to reinforce training pairs, a common procedure during the testing phase of the TI task.

Harris and McGonigle (1994) report improved performance during trigram testing despite the fact the animals were rewarded regardless of choice. This indicates the monkeys were also still learning. However, some two-tier agents showed performance degradation during trigram training (see "after training" columns in Table 2). Figure 6 shows one cause for this failure: a slow-learning agent that fails to find a stable solution has two priorities converge, which can lead to incorrect choices if these are both selects or both avoids. Agents with larger values of δ generally find a new stable though incorrect solution during trigram testing, when *any* solution is reinforced (during bigram testing, training pairs are still differentially rewarded, giving the agent some useful feedback). These phenomena were not reported by Harris and McGonigle (1994). If they never occur, this would be a point of inconsistency with the model.

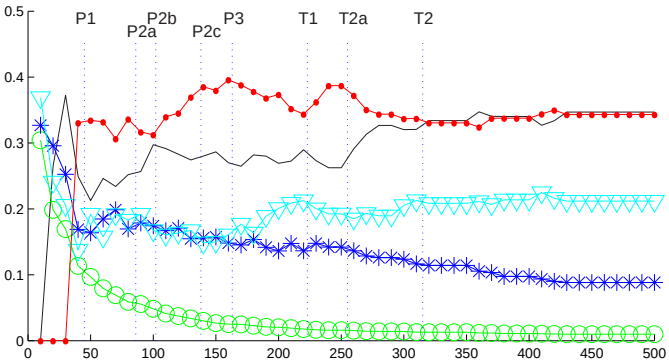

Figure 6: Phased Training where learning slips during trigram testing. Rules are: select $A\bullet$, select $B-$, and either avoid $E\triangledown$ or select $C*$. Parameters: $\Xi = .12$, $\delta = .02$

5 Discussion

The results of our model have not only met but exceeded the goals of our simulation. We have achieved our goal by showing that a system which performs like that of Harris and McGonigle (1994) can be learned. Further, it can be learned with a very simple learning mechanism. Unlike backpropogation (a popular AI neural-network learning algorithm often applied to psychological modeling Lillo et al., 2001; McClelland et al., 1995), the learning rule we used is uncontroversial in its neurological plausibility. It is a simplified version of 'delta learning' (Widrow and Hoff, Jr., 1960). The fact the model unexpectedly generates errors in learning similar to those of real subjects further validates both our two-tier model and the work by Harris and McGonigle (1994).

5.1 Production Rules as Natural Representations

Intuitively, the quality of the fit of the Harris and McGonigle (1994) model is surprising. Why would subjects focus only on a single item as the precondition of an action? Why would a non-deterministic 'avoid' be one of only two possible actions? Given the demonstrated quality of both the Harris and McGonigle (1994) model and of the present two-tier model, these questions have been transformed into open research questions.

We have come up with two possible explanations, which are not necessarily mutually exclusive. First, focusing on a single item as the precondition matches the precondition to the eventual goal. Grasping a single item is tightly associated with receiving the reward. There is some evidence that the representation underlying task learning in primates is at least analogous, and perhaps identical, to the machine-learning concept of a generative model (e.g. Hinton and Ghahramani, 1997; Bishop et al., 1998). That is, the same model which is attempting to learn ways of classifying the world by observation can also be driven forward to make predictions, or in the case of task learning, to drive actions. This is a parsimonious way to combine learning and control in the same system. Some support for this hypothesis comes from neurophysiological evidence that some individual neurons in the premotor cortex are associated with both perception and action (Miller, 2000; Rizzolatti et al., 2000). Since subjects learn to grasp only one item, there may be a neurophysiological bias towards having the perceptual indicator or releaser for that action also be based on a single item.

The second possible explanation requires considering the ontogenetic acquisition of the two-tier model. We propose this account:

1. In the earliest phase of training, when the subjects learn to grasp an object for a reward, they learn the basic rule structure associating any stimulus and the response *select*.

2. In the large blocks of ordered pair training, subjects discover that they may select only one object. In this phase they learn two things: to discriminate stimuli, and also to inhibit the *select* stimulus in some contexts. This inhibition results in the *avoid* rule being learned and associated with some objects that, for whatever reason, are highly salient (hold the subject's attention.)

3. As the blocks of stimuli change, subjects must learn (or adjust) prioritization between neighboring inhibition rules.

In this account, the origin of the single-item cue is the original learned behavior pattern: the grasping of a single item.

This account nearly provides for the entire set of TI effects described in Section 1.2, except for the symbolic distance effect. Accounting for this requires further assuming that the process of selecting between two rules takes longer the more similar the two priorities are. For example, the 'decision' to act on one rule may require a buildup of activation that is inhibited by competing rules in proportion to their priority. Thus when two competing rules are similarly weighted, this process is likely to be both longer and more arbitrary. This is essentially an elaboration of the common assumption of the sequence-learning models (e.g. Wynne, 1998).

5.2 Other Models of Transitive Inference

We have already addressed the single-weight-vector modeling approach that currently dominates this area (Wynne, 1998), and McGonigle and Chalmers (1977) binary sampling theory. In this section we will address a few other models.

Lillo et al. (2001) recently presented a backpropogation model which displays the Symbolic Distance Effect (see Section 1.2), but does not account for the McGonigle and Chalmers (1977) trigram results. Backpropogation is a machine-learning technique which allows weight changes to be determined across layers of vectors simultaneously (Hertz et al., 1991). Similar to the two-tier model, Lillo et al. (2001) use a two-layered approach. The difference in outcomes between these models shows the importance not only of modularity (in this case, layers) within a network, but also of configuration.

No learning system can operate successfully in a reasonable amount of time without being provided with bias (Wolpert, 1996). The very structure of a network, its connectivity and its connections to sensing and action, serves as that bias. If the two output units of the Lillo

pair	subject scores	labelling hypothesis	two-tier hypothesis
A > B	88	*100*	*100*
$A > C$	52	50	50
$A > D$	85	100	88
$B > C$	27	0	13
$B > D$	50	50	50
C > D	87	*100*	*100*

Table 3: Comparison of results from de Boysson-Bardies and O'Regan (1973) Experiment 4, their labeling hypothesis, and the two-tier model. The subjects were deliberately not trained to full performance on the training pairs (AB and BD) but the models do not take this into account. See Table 4 for explanation of the two-tier values.

et al. (2001) model were connected to *select* and *avoid* instead of ordinal positions on the board, their model would be fairly similar to ours, and might even learn the task. However, notice that for our simulation these 'layers' were separated in different modules, the results of each of which were necessary for the action sequence. The output of the first 'layer' determines not only which rule is chosen, but also which item is being visually fixated when that rule is chosen. This is critical to the operation of the *avoid* rule, particularly in the trigram case. Also, the learning rule we used functions more quickly and with more biological plausibility than backpropagation.

Lillo et al. (2001) also compare the results of their model with a study by de Boysson-Bardies and O'Regan (1973). In this study, both children and adults were presented the TI task with no phased learning. Rather, they were given either pairs presented in random order or iterations of ordered presentations (i.e. $AB, BC, CD, DE, AB, BC \ldots$). Adults learned the condition with the ordered presentations better, but children learned both conditions at the same rate. Both children and the Lillo et al. (2001) model sorted a newly presented element to be in the middle of their sequence.

Our single-vector model showed in Section 4.1 would also sort a new item in the middle, for the trivial reason that this is how we chose to set prior expectation for a newly found item — to be the mean of all the other weights. On the other hand, not every production rule stack generated by our full model would necessarily do this, though all possible rule sets might well do it in aggregate.

de Boysson-Bardies and O'Regan (1973) do not give much information on the procedure for training the adults, but given their perfect accuracy levels it seems fair to guess that the adults formed a full, explicit model of the task in the case of ordered presentations and were performing classic, logic-based transitive inference to solve

pair	applicable rules	individual predictions	aggregate prediction
A>B	s(A)	100	**100.0**
	a(B)	100	.
$A > C$	s(A), s(C)	50	**50.0**
	s(A)	100	.
	s(C)	0	.
		50	.
$A > D$	s(A), a(D)	100	**87.5**
	s(A)	100	.
	a(D)	100	.
		50	.
$B > C$	s(B), a(C)	0	**12.5**
	s(B)	0	.
	a(C)	0	.
		50	.
$B > D$	a(B), a(D)	50	**50.0**
	a(B)	0	.
	s(D)	100	.
		50	.
C>D	s(C)	100	**100.0**
	a(D)	100	.

Table 4: Explanation of two-tier predictions from Table 3. For each training pair, we assume each agent is at chance for learning one rule (either a select or an avoid) to solve it. We also assume that when two rules with different outcomes may be applied, which rule has higher priority is at chance. 'Individual' predictions are based on possible rule sets an individual might have acquired in training; 'aggregate' assumes an even distribution of individuals.

the problems. On the other hand, they may not have received enough training to form the implicit rule-based model the children formed in either condition.

de Boysson-Bardies and O'Regan (1973) have their own hypothesis for explaining the children's strategies, which they call the "labeling strategy". This is somewhat like the binary sampling theory, in that it first assumes the children associate the labels "big" and "small" with items in pairs, and then posits a complicated set of transformations to explain performance on the middle items. They test their theory in their fourth experiment by training children *only* on the pairs $A > B$ and $C > D$, then testing them on all possible combinations of stimuli. Their labeling-strategy model successfully predicts that most children consider that $A > D$ and $B < C$, rather than coming up with a complete ordering of the pairs (e.g. $A > B > C > D$ or $C > D > A > B$).

The two-tier model gives the related but not identical prediction that for each pair the children learn 1) to focus on an item and 2) to either select or avoid that item. We

assume they learn no rule for the other item. Given this model, and assuming that all rules are equally likely to be learned across all items, aggregate data for two-tier model actually matches the de Boysson-Bardies and O'Regan (1973) data better than their own model does on the two pairs where the predictions differ (see Table 3). Note though that for *individuals*, our model predicts 75% of children will consistently choose $A > D$ and $B < C$, while 25% will be completely at chance for the first presentation of these pairs, because they will have no applicable rule.

Finally, another model related to the labeling strategy is the value-transfer model (Siemann et al., 1996). This model proposes that the agents learn through simple reinforcement that A is good, E (or whatever the final element is) is bad, and all the other elements are neutral, since they are equally likely to be rewarded or punished. However, the value associated with A leaks to its neighbor B, which is seen in common association with A (in fact, in every training pair containing A.) Siemann et al. (1996) have demonstrated conclusively that value transfer happens between frequently associated pairs of stimuli. However, again this model fails to explain the trigram data. It is worth noting that most value-transfer models have been fitted to pigeon TI data. We have no evidence to date that pigeon TI would be well accounted for by the two-tier model; pigeons certainly learn sequences very differently from primates (Terrace and McGonigle, 1994).

5.3 Possible Neurological Correlates

The clearest aspect of our model is that it splits the problem of task-learning into two parts:

1. learning to associate actions with environmental cues
2. learning to prioritize these cues when more than one set holds.

Because the Harris and McGonigle (1994) model indicates that the subjects never learn multiple rules for the same cue contexts, determining a prioritization for the cues / items is identical with determining a prioritization for the rules.

Neuroscience indicates that long-term rule learning, including categorization for action selection, seems to be dependent on both the prefrontal cortex (Freedman et al., 2001; Wallis et al., 2001) and the hippocampus. In particular, the entorhinal cortex seems necessary for forming episodic or event memory necessary for such tasks as paired association, while the hippocampal formation proper is necessary for learning more complex relations (Alvarado and Bachevalier, 2000; Baxter and Murray, 2001).

The hippocampus has long been implicated in memory consolidation. Its very sparse, redundant coding system seems ideally suited to recording memories in real time which can then be consolidated — saved to long term storage in an integrated representation — using a slower learning mechanism (McClelland et al., 1995; Louie and Wilson, 2001). Thus the following might be a reasonable neurobiological explanation of the model presented in this paper. The basic association between stimuli and behaviors would be performed by the prefrontal cortex generalizing from associations learned by the entorhinal cortex. The hippocampal formation would learn prioritizations between these associations.

5.4 Relationship to AI and Cognitive Science

We should point out that our model is hardly the first nor the most wide-ranging documentation of the efficacy of production-rules as cognitive models. There is an enormous literature produced by the Soar community (Newell, 1990) and by the ACT-R community (Anderson, 1993) on this subject.

Our model is also relevant to the problem of action selection. It has been suggested that robust, reactive agent control requires, among other things, focused action-selection competencies based on prioritized lists of subgoals, each guarded by context-dependent triggers (Nilsson, 1994; Bryson and Stein, 2001). This research shows the difficulty even clever learners such as monkeys and human children have acquiring such rules through uninformed search. This would seem to emphasize the importance of providing plans for intelligent agents, whether such plans are innate (in AI, preprogrammed) or acquired socially through imitation or instruction (Schaal, 1999).

6 Conclusions and Predictions

This paper has presented a model of TI learning in monkeys and humans. The model presented here combines several prior models (Wynne, 1998; McGonigle and Chalmers, 1977; Harris and McGonigle, 1994) to provide a novel extension in the form of a two-tiered learning system.

By showing that the model of Harris and McGonigle (1994) can be learned in a biologically plausible way, we have further strengthened the validity of their model, which was already by far the best-fitting model of trigram TI performance. By finding further correlations between our new model and real animal learning, particularly in the error conditions, we believe we have provided further insight into general task learning in primates. We have also suggested a neurologically-based explanation of why production rule models are so often successfully used for cognitive modeling in humans and other animals.

Our model makes a number of testable predictions. For example:

- Visual attention should settle on the item associated with a rule just before the grasp is made — in the case of an *avoid* rule, this would not be the same item as the one selected.
- In general, reaction times and visual scanning behavior should be different for select and avoid rules.
- For individuals, the ordering of a new item (as in de Boysson-Bardies and O'Regan (1973) experiment 3) should be determined by the existing rule stack. For example, if the rule stack is all selects as in Eq. 1.6, a new item would be positioned last or second to last, if they were all avoids it would be positioned first.

Testing these predictions requires running trigram experiments after TI pair training in order to discriminate which rules were learned by individual subjects.

In our own future work, we are attempting to generalize the two-tier model to explain primate performance on more types of specialized task learning, such as the invisible-displacement task (Hood et al., 2000; Hauser et al., 2001) or the object-retrieval task (Diamond, 1990; Santos et al., 1999). We hope the insight from this model will help explain both abilities and deficits in learning performance on these sorts of tasks.

Acknowledgments

We would like to thank Marc Hauser, Will Lowe, Mark Baxter, Juan Delius, Lynn Andrea Stein, Olin Shivers, and the denizens of the Harvard Primate Cognitive Neuroscience Lab (particularly Roian Egnor) for their comments on earlier drafts.

References

Alvarado, M. C. and Bachevalier, J. (2000). Revisiting the maturation of medial temporal lobe memory functions in primates. *Learning and Memory*, 7(5):244–256.

Anderson, J. R. (1993). *Rules of the Mind*. Lawrence Erlbaum Associates, Hillsdale, NJ.

Baxter, M. G. and Murray, E. A. (2001). Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys. *Hippocampus*, 11(1):61–71.

Bishop, C. M., Svensén, M., and Williams, C. K. I. (1998). GTM: The generative topographic mapping. *Neural Computation*, 10(1):215–234.

Bryson, J. J. (2001). *Intelligence by Design: Principles of Modularity and Coordination for Engineering Complex Adaptive Agents*. PhD thesis, MIT, Department of EECS, Cambridge, MA. AI Technical Report 2001-003.

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

Chalmers, M. and McGonigle, B. (1984). Are children any more logical than monkeys on the five term series problem? *Journal of Experimental Child Psychology*, 37:355–377.

de Boysson-Bardies, B. and O'Regan, K. (1973). What children do in spite of adults' hypotheses. *Nature*, 246:531–534.

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.

Freedman, D. J., Riesenhuber, M., Poggio, T., and Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291:312–316.

Gilovich, T., Vallone, R., and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology*, 17:295–314.

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., Williams, T., Kralik, J. D., and Moskovitz, D. (2001). What guides a search for food that has disappeared? experiments on cotton-top tamarins (saguinus oedipus). *Journal of Comparative Psychology*, 115(2):140–151.

Henson, R. N. A. (1998). Short-term memory for serial order: the start-end model. *Cognitive Psychology*, 36:73–137.

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

Hinton, G. E. and Ghahramani, Z. (1997). Generative models for discovering sparse distributed representations. *Philosophical Transactions of the Royal Society B*, 352:1177–1190.

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.

Lillo, C. D., Floreano, D., and Antinucci, F. (2001). Transitive choices by a simple, fully connected, backpropagation neural network: implications for the comparative study of transitive inference. *Animal Cognition*, 4(1):61–68.

Louie, K. and Wilson, M. A. (2001). Temporally structured replay of awake hippocampal ensemble activity during rapid eye movement sleep. *Neuron*, 29(1):145–156.

McClelland, J. L., McNaughton, B. L., and O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review*, 102(3):419–457.

McGonigle, B. and Chalmers, M. (1977). Are monkeys logical? *Nature*, 267:694–696.

McGonigle, B. and Chalmers, M. (1996). The ontology of order. In Smith, L., editor, *Critical Readings on Piaget*, chapter 14. Routledge, London.

Microsoft (2001). *Windows 2000, Service Pack 2*. Redmont, WA.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience*, 1:59–65.

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

Nilsson, N. J. (1994). Teleo-reactive programs for agent control. *Journal of Artificial Intelligence Research*, 1:139–158.

Piaget, J. (1954). *The Construction of Reality in the Child*. Basic Books, New York.

Red Hat (2001). *Linux 7.1*. Raleigh, NC.

Rizzolatti, G., Fogassi, L., and Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In Gazzaniga, M. S., editor, *The New Cognitive Neurosciences*, chapter 38, pages 538–552. MIT Press, Cambridge, MA, 2 edition.

Santos, L. R., Ericson, B., and Hauser, M. D. (1999). Constraints on problem solving and inhibition: object retrieval in cotton-top tamarins. *Journal of Comparative Psychology*, 113:1–8.

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

Siemann, M. and Delius, J. D. (1993). Implicit deductive reasoning in humans. *Naturwissenshaften*, 80:364–366.

Siemann, M., Delius, J. D., Dombrowski, D., and Daniel, S. (1996). Value transfer in discriminative conditioning with pigeons. *The Psychological Record*, 46:707–728.

Steele Jr., G. L. (1990). *Common Lisp: The Language*. Digital Press, Bedford, MA, second edition.

Terrace, H. S. and McGonigle, B. (1994). Memory and representation of serial order by children, monkeys and pigeons. *Current Directions in Psychological Science*, 3(6):180–185.

Wallis, J. D., Anderson, K. C., and Miller, E. K. (2001). Single neurons in the prefrontal cortex encode abstract rules. *Nature*, 411:953–956.

Widrow, B. and Hoff, Jr., M. E. (1960). Adaptive switching circuits. *IRE WESCON Convention Record*, 4:96–104.

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

Wynne, C. D. L. (1998). A minimal model of transitive inference. In Wynne, C. D. L. and Staddon, J. E. R., editors, *Models of Action*, pages 269–307. Lawrence Erlbaum Associates, Mahwah, N.J.

Xanalys (2001). *Lispworks Professional Edition 4.1.20*. Waltham, MA. (formerly Harlequin).