Monkey Errors in Transitive Inference

Joanna J. Bryson
Harvard Primate Cognitive Neuroscience Laboratory
Cambridge, MA 02138
jbryson@wjh.harvard.edu

Abstract

The ability to make transitive inference, $(A > B) \wedge (B > C) \vdash A > C$, occurs in many mammal species, given adequate training. Further, errors made by monkeys are quite similar to those made by 5-year-old children. This paper presents a model that learns transitive inference by acquiring and prioritizing sets of behavior rules linked to contexts. When correctly learned, the resulting system is identical to the model of Harris and McGonigle (1994) which closely matches primate performance. The extension presented here explains further both errors and individual variation in monkey performance.

1. Introduction

Transitive inference 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 real numbers, 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.

Piaget described transitive inference as an example of concrete operational thought (Piaget, 1954). That is, children become capable of doing transitive inference when they become capable of mentally performing the physical manipulations they would use to determine the correct answer. In the case of transitive inference, this manipulation would be 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 transitivity with preoperational children and a variety of animals — apes, monkeys, rats and even pigeons (see for review Wynne, 1998) — has lead some to the conclusion that transitive inference is a basic animal competence, not a skill of either operations or logic. In fact Siemann and Delius (1993) (reported in Wynne, 1998) have shown that in adults trained in the context of an exploration-type computer game, there was no performance difference between the individuals who formed explicit models of the comparisons and those who did not ($N = 8$ vs. 7 respectively).

This paper describes results from a model of transitive inference learning that applies to monkeys and children. The model assumes that animals are actually learning two layers of priorities: a prioritization of rules to apply in a given context, and a prioritization of contexts to attend to when more than one set of cues is presented simultaneously. This model explains several common learning errors present in even highly intelligent subjects such as monkeys and children.

2. Transitive Inference

2.1 Characteristic Transitive Inference 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 do better (more correct and faster reaction times) when a test pair contains one of the ends. 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*: again, 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 is seen as contradicting any step-wise chaining model of transitive inference (such as Piaget's concrete operations) since distant items would require *more* steps and therefore a longer reaction time (RT).

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	Binary 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	6 sets of 10 trigrams in random order, reward for all.
T3	Extended version of T2.

Figure 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 3.

2.2 Training Subjects for Transitive-Inference

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, the chain must have a length of at least five items ($A \ldots E$) in order 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[1]. 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. For example, one subject may learn $blue < green < brown$ while another $brown < blue < green$.

The subjects are first taught the use of the testing apparatus; they are presented with and rewarded for selecting one option. For the experiments reviewed in this paper, subjects learned to look under colored tins for a reward. Next, they are trained on the first pair DE, where only one element, D is rewarded[2] When the subject has learned this to some criteria, they are trained on CD. After all pairs are trained in this manner, there is often a phase of repeated ordered training with fewer exposures (e.g. 4 in a row of each pair) followed by a period of random presentations of training pairs. Once a subject has been trained to criteria, they are exposed to the testing data. In testing, either choice is rewarded, though differentially rewarded training pairs may be interspersed with testing pairs.

2.3 Standard Models

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 has been demonstrated necessary in order 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:

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

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

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.

3. Modeling Transitive Inference Errors

It is often the case in cognitive modeling that the *errors* made by subjects are the most telling way to disambiguate between different possible models. This section describes a set of experiments that indicate that there is more to transitive inference than a single-weight system such as the above indicates. It also describes two models describing these errors.

3.1 The Binary Sampling Model

In one of the earliest of the non-cognitive transitivity papers, McGonigle and Chalmers (1977) not only demonstrated animal learning of transitive inference, but also proposed a model to account for the errors the animals

[1] This is true even for simple sorting of conspicuously ordered items such as posts of different lengths (McGonigle and Chalmers, 1996).

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

made. Their subjects were squirrel monkeys (*saimiri sciureus*). Monkeys, like children, 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 the 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 animal 'selects' an item that is not actually present, it cannot act on that selection. Instead, this selection reinforces its consideration of that item, which tends to push the animal toward 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 naive probabilistic model: it incorporates the concept of expectation, but is not very systematic. It also fails to explain the symbolic distance effect. However, it motivated McGonigle and Chalmers to generate a fascinating 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). In fact, the monkeys showed 3%, 36%, and 61%, respectively.

3.2 The Production-Rule Stack Model

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

$$\begin{array}{ll} 1. & (A \text{ present}) \Rightarrow \text{select } A \\ 2. & (E \text{ present}) \Rightarrow \text{avoid } E \\ 3. & (D \text{ present}) \Rightarrow \text{avoid } D \\ 4. & (B \text{ present}) \Rightarrow \text{select } B \end{array}$$

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

$$\begin{array}{ll} 1'. & (A \text{ present}) \Rightarrow \text{select } A \\ 2'. & (B \text{ present}) \Rightarrow \text{select } B \\ 3'. & (C \text{ present}) \Rightarrow \text{select } C \\ 4'. & (D \text{ present}) \Rightarrow \text{select } D \end{array}$$

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.

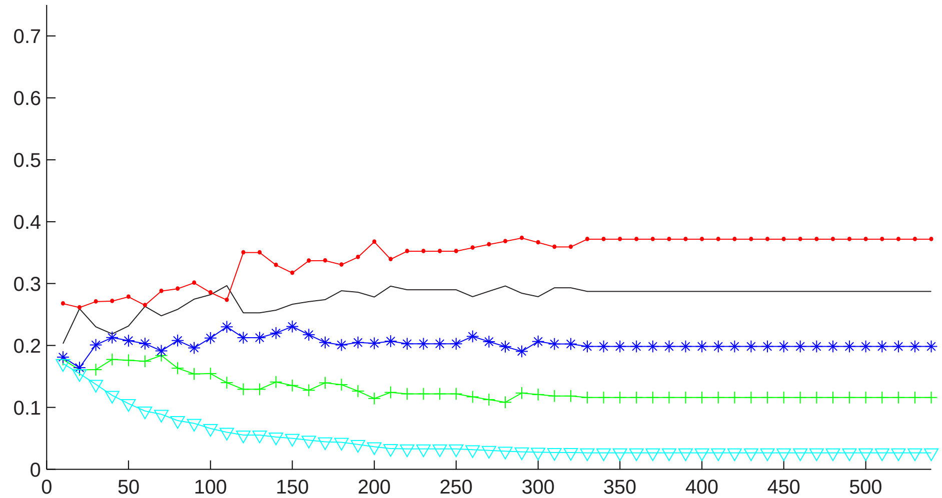

Figure 2: Results for a single vector (only one rule — select). X-axis: trial number ; Y-axis: weights of stimuli vector (sum to one). Free variables are set to: $\Xi = .08$, $\delta = .02$. Key: $A\bullet$, $B-$, $C*$, $D+$, $E\triangledown$.

4. The Proposed Model

The present model is influenced by all three models discussed above. It is also influenced by the neurologically based models of action selection (Redgrave et al., 1999) and reinforcement learning (see Gillies and Arbuthnott, 2000, for a review) in the basal ganglia. The basal ganglia appears to control action-selection primarily through context-based inhibition of activity moving from perceptual and motivational streams toward motor cortices.

4.1 Components of the Model

The model assumes the following:

1. In the earliest phase of training, the subjects learn to grasp any viewed tin in order to get their reward.

2. In the large blocks of ordered pair training, subjects discover that they may only select one item, and that only one is rewarded. Consequently, they must inhibit both applicable grasping behaviors until they have selected one item.

3. As the size of the training blocks is reduced, subjects must also learn prioritization between neighboring inhibition rules. The interactions of these neighbors is generally sufficient to produce a total ordering, although the compartmentalization also allows for learning the additional EA pair.

4. The process of selecting between two rules of similar activation involves probabilistic attentive sampling of the alternative rules. Attention allows for searching for increased evidence and generally increases the activation of the rule. Where two competing rules are similarly weighted, this process is likely to be both longer and more arbitrary.

The remainder of this paper focuses on modeling the third of these proposals. The first two are relatively uncontentious, and the fourth is essentially an elaboration of the common assumption of the sequence-learning models (e.g. Wynne, 1998).

4.2 Learning in the Model

The present model requires several vectors to be learned representing the priorities between contexts and the priorities for the rules as applied to each context.

Each vector $\mathbf{v}$ is composed of normalized weights representing the priority level. The 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. For the update rule, I replace the sigmoid in eq. 1 with a step function. Assume that the pair XY has been presented[3], that X has been selected by the agent, and Ξ and δ are free parameters.

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

For a five item task, there can be up to six vectors: one to represent the different stimuli, and one for each stimuli representing associated rules. In fact, in some of the experimental runs up to two stimuli are never selected for attention. In these runs less vectors are used (see examples below).

For each trial, a subject agent selects first a stimuli to attend to, then a rule to apply, based on the current prioritizations. (Initial weights are even, in the matched case the outcome is arbitrary.) The subject agent also stores its selections in short term working memory. After the subject agent has chosen, the training agent rewards

[3] In the case of a trigram, the subject selects one of the two rejected items at random for the role of Y in equation 2.

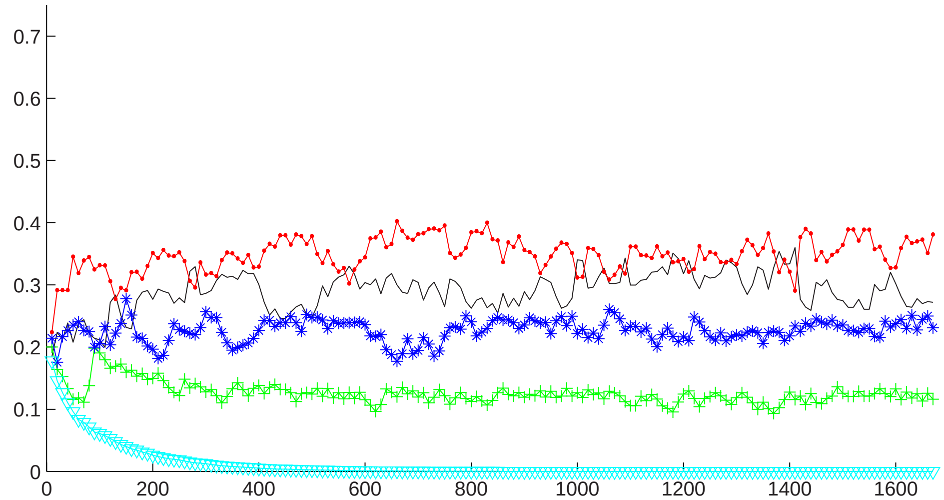

Figure 3: Single Vector but with "stupider" parameters: $\Xi = .12$, $\delta = .02$. Key: $A\bullet$, $B-$, $C*$, $D+$, $E\bigtriangledown$.

or fails to reward the subject. The subject agent updates its vectors based on the reward and its recalled decision.

5. Experiments and Results

The model used for the following results was built under Behavior Oriented Design (Bryson and Stein, 2001), an iterative AI design methodology which supports developing coordinated modular intelligence. The following results are prototypical of approximately 70 experimental conditions.

5.1 Single-Vector (Standard Model) Results

Before implementing the full model described in Section 4., we first implement a simpler model, replicating roughly the single-vector results menitoned in Section 2.3. Figure 2 shows a typical result for standard transitive inference modeling, when pairs have been presented in random order. Unlike real animals or 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). Unless learning tails off, a 'hot hands'-like phenomena causes the weight of an item that has recently occurred in a number of training pairs to grow higher than the next-ranking element. This is illustrated in Figure 3. These results provide one possible explanation for individual difference 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 colors. This violates the End Anchor Effect, so is not biologically plausible. On the other hand, the other end of the vector is reliably well discriminated. Consequently, the model might be fixed if instead of using a single vector 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.

5.2 Rule-Learning Results

Figures 4 and 5 illustrate the system with additional vectors provided for representing rules, as described in Section 4.. Rule selection is made very early in training and remains stable, so is not represented in the 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 constantly learns either the solution shown in Figure 4 or a symmetric one with the select (+) B and avoid (-) C fighting for top priority. Using the training regime for real children (Figure 1), the full model can learn (Figure 5).

A correct stack must be either in sequence [+A +B +C] or [-E -D -C] or an ordered cross of these (e.g. [+A -E +B]). Although the rules learned without the 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

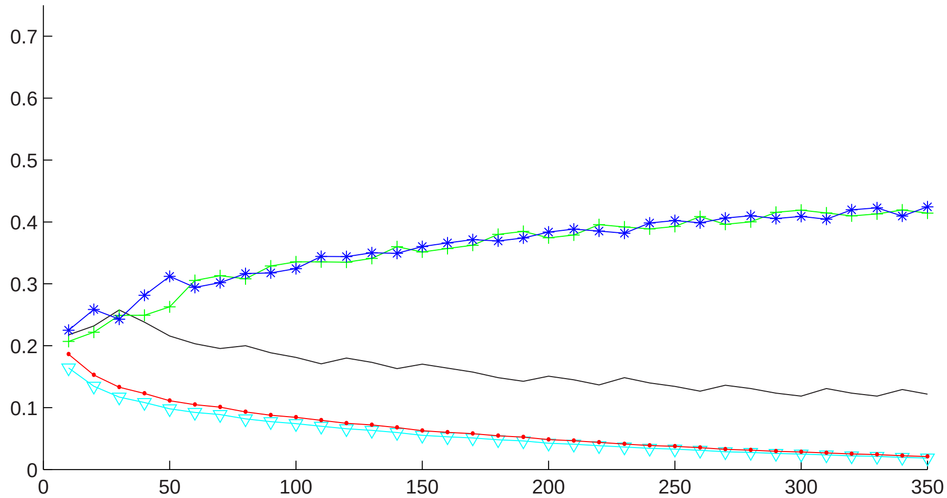

Figure 4: Rule Learning with no training regime. Rules learned in descending order of priority: select $D+$, avoid $C*$, avoid $B-$. Parameters: $\Xi = .08$, $\delta = .02$.

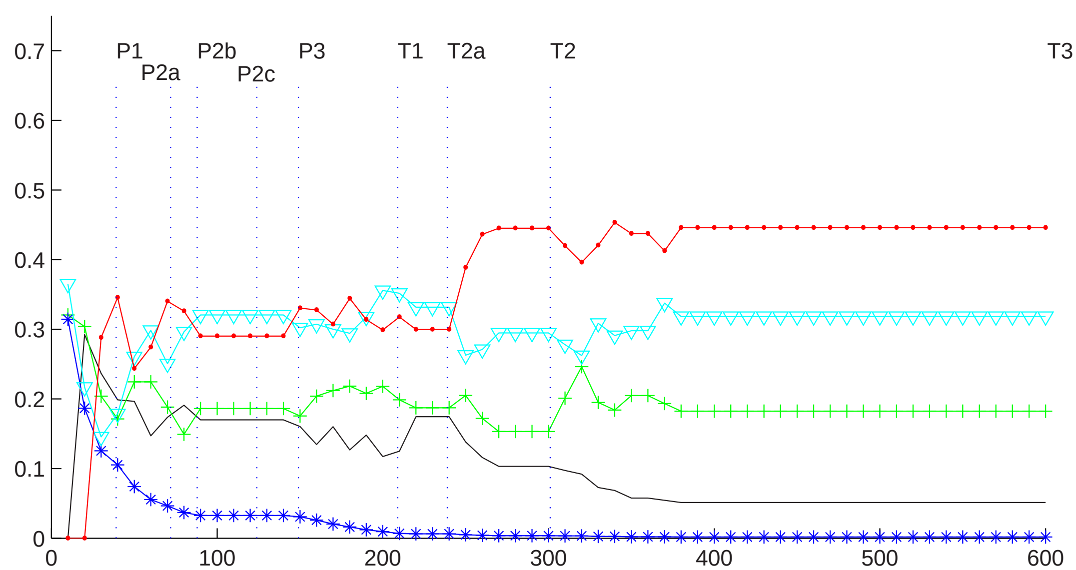

Figure 5: Rule Learning with Phased Training. This regime results in a variety of solutions, all correct. Here the rules are: select $A\bullet$, avoid $E\bigtriangledown$, avoid $D+$. This succeeds even with very "stupid" parameters: $\Xi = .12$, $\delta = .06$.

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 criteria in learning transitive inference. Although no results were reported for monkeys or children that missed criteria, one monkey subject, Blue, passed criteria but still showed a consistent error between the 3^{rd} and 4^{th} item is shown by Harris and McGonigle (1994, p. 332). This is in keeping with the results of this model.

The multi-vector model is also somewhat monkey-like in that the training regiment for children proved more rigorous than necessary for overcoming these problems. Nearly all agents converge quickly, and the ones that fail to learn fail early, usually during Phase 2a. Because some rules will never fire at the same time during training, stable solutions are found even with large values for δ and Ξ. Note that significant learning occurs during bigram testing, showing the utility of continued reinforcement for the training pairs — another feature of experiments with natural subjects. Notice also that further learning happens during undifferentiated training of trigrams. This is also reported by McGonigle. However, for some trials, the top two priorities converged (increasing the distance from the third to Ξ, which leads to incorrect choices if they are both selects or both avoids. This phenomena was not reported by McGonigle. If it never occurred, this would also point towards Henson's dual-vector model of sequence representation.

6. Discussion and Conclusions

This paper has presented a model of transitive inference learning in monkeys and man. The model presented here combines the wisdom of several prior models (Wynne, 1998; McGonigle and Chalmers, 1977; Harris and McGonigle, 1994), providing a novel extension. It also accounts for standard failures in such learning.

Future work might include fully implementing the model proposed in Section 4.. This would require altering the selection between options to be more probabilistic when the model is uncertain, as in the Wynne model (Wynne, 1998). Also, to make the model more biologically plausible, the basic priority representation probably requires two weight vectors instead of one, as suggested in Section 5.1 and by Henson (1998).

The demonstrated model of the transitive-inference task is also interesting with respect to the problem of action selection. It has been suggested that robust, reactive agent control requires, among other things, focussed 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 provided plans for intelligent agents, whether such plans are preprogrammed (in animals, genetic) or acquired socially through imitation or instruction.

Acknowledgments

I would like to thank Lynn Andrea Stein, Olin Shivers, Will Lowe, anonymous reviewers, and the denzines of the Hauser Lab (particularly Roian Egnor) for their comments on an earlier draft. Will Lowe also assisted with the matlab figures.

References

Bryson, 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.

Gillies, A. and Arbuthnott, G. W. (2000). Computational models of the basal ganglia. *Movement Disorders*, 15(5):762–770.

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

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

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

McGonigle, B. and Chalmers, M. (1996). The ontology of order. In Smith, L., (Ed.), *Piaget: A Critical Assessment*. Routledge.

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.

Redgrave, P., Prescott, T. J., and Gurney, K. (1999). The basal ganglia: a vertebrate solution to the selection problem? *Neuroscience*, 89:1009–1023.

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

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