Anim Cogn (2007) 10: 1–15
DOI 10.1007/s10071-006-0024-9

ORIGINAL ARTICLE

Joanna J. Bryson · Jonathan C. S. Leong

Primate errors in transitive 'inference': a two-tier learning model

Received: 30 May 2005 / Revised: 1 May 2006 / Accepted: 4 May 2006 / Published online: 30 June 2006

Abstract Transitive performance (TP) is a learning-based behaviour exhibited by a wide range of species, where if a subject has been taught to prefer *A* when presented with the pair *AB* but to prefer *B* when presented with the pair *BC*, then the subject will also prefer *A* when presented with the novel pair *AC*. Most explanations of TP assume that subjects recognize and learn an underlying sequence from observing the training pairs. However, data from squirrel monkeys (*Saimiri sciureus*) and young children contradict this, showing that when three *different* items (a triad) are drawn from the sequence, subjects' performance degrades systematically (McGonigle and Chalmers, Nature 267:694–696, 1977; Chalmers and McGonigle, Journal of Experimental Child Psychology 37:355–377, 1984; Harris and McGonigle, The Quarterly Journal of Experimental Psychology 47B:319–348, 1994). We present here the *two-tier model*, the first learning model of TP which accounts for this systematic performance degradation. Our model assumes primate TP is based on a general-purpose task learning system rather than a special-purpose sequence-learning system. It supports the hypothesis of Heckers et al. (Hippocampus 14:153–162, 2004) that TP is an expression of two separate general learning elements: one for associating actions and contexts, another for prioritising associations when more than one context is present. The two-tier model also provides explanations for why phased training is important for helping subjects learn the initial training pairs and why some subjects fail to do so. It also supports the Harris and McGonigle (The Quarterly Journal of Experimental Psychology 47B:319–348, 1994) explanation of why, once the training pairs have been acquired, subjects perform transitive choice automatically on two-item diads, but not when exposed to triads from the same sequence.

Keywords Transitive inference, choice or performance · Task learning · Hippocampal learning · Modelling

J. J. Bryson (✉)
Artificial Models of Natural Intelligence, University of Bath,
Bath BA2 7AY, UK
e-mail: J.J.Bryson@bath.ac.uk
Tel.: +44-1225-383934
Fax: +44-1225-383493

J. C. S. Leong
Max Planck Institute for Neurobiology,
Cellular Systems Neurobiology,
Martinsried, Germany
e-mail: leong@neuro.mpg.de

Introduction

Transitive inference (TI) is the process of reasoning whereby one determines for some quality that if $A > B$ and $B > C$, then $A > C$. In some domains, such as integers or heights, this property holds for any A, B or C. For other domains, such as primate dominance hierarchies, the property does not necessarily hold (Wright 2001). Transitive *performance* (TP) refers to the same determination as TI, but makes no claim about what cognitive processes underlie the observed behaviour.

TP has become a significant benchmark task for psychologists of both animal and human cognition and has also prompted many modelling attempts. While it is well known that mistakes often tell us more about the nature of cognitive processes underlying behaviour than does flawless performance, few models of TP account for failures to learn the task. There are two types of mistakes to account for. First, both human and animal subjects often fail to meet criterion on acquiring the initial pair-relations (e.g. $A > B, B > C, \ldots$) despite careful training, though if these pairs are acquired, subjects reliably display TP for pairs of items arbitrarily drawn from the series. Second, a set of data originally due to McGonigle and Chalmers (1977) shows that both young children and monkeys systematically fail to generalise their ability to perform transitive 'inference' from the context of two different items to that of three though if two of the three, items are the same the new three-item context presents no problems (McGonigle and Chalmers 1992).

Here we present a model of performance on the transitive task that explains both types of mistakes. In doing so, we reveal that the problem with previous artificial intelligence

models of transitive inference is that they learn *too well* to accurately model the target behaviour. The machine learning techniques used too easily solve the problem of finding local minima—finding a solution that while better than the most similar alternatives is not the best solution overall. We suggest that primates have more trouble ignoring attractive locally-optimal solutions than some previous models account for.

Our research supports suggestions of Heckers et al. (2004) and others that TP relies on two separate learning processes: one to associate each stimulus with an action, and another to prioritise which stimulus-action association is most salient in a context where more than one association is relevant. We call our model the two-tier model because we dedicate a tier of associative learning to each of the two learning problems. While multi-layer models have already been attempted, previous models used backpropagation, which ensured globally optimal learning (e.g. De Lillo et al. 2001; Frank et al. 2003). By contrast, our model keeps its two tiers of learning independent and so errs appropriately where other systems have learned to perform flawlessly. This sort of system has been demonstrated as useful for general-purpose task learning and performance (Anderson 1993; Bryson and Stein 2001). As such, our model supplies a simple solution to the question as to why animals would need a special capacity for TP: they would not.

In the next section we describe TP training, results and effects, including a review of the McGonigle and Chalmers (1977, 1992) triad data. We also describe a model due to Harris (1988) which explains the triad performance by fully trained individuals, but does not demonstrate learning. In the following sections we describe our own model, experiments and results. We discuss the implications of our model and how it relates to other existing models. Our model's results also produce testable predictions, which are discussed at the end of the paper.

The task

Transitive inference and performance

Piaget (1954) described TI as an example of concrete operational thought. That is, children become capable of TI only when they become capable of mentally performing the physical manipulations they would otherwise use to determine the correct answer, normally at the age of about 6 years. For example, TI involves ordering the objects into a sequence using the rules $A > B$ and $B > C$, and then observing the relation between A and C.

Yet Piaget was also aware of an 'automatic' variety of TI in younger children, distinguished from true TI by the subject's inability to explain their performance (Piaget 1928; Wright 2001). Since the 1970s, TP has been demonstrated in young, pre-concrete-operational children (Bryant and Trabasso 1971) and a variety of animals—e.g. monkeys (McGonigle and Chalmers 1977), rats (Dusek and Eichenbaum 1997) and pigeons (Fersen et al. 1991)—not normally ascribed with concrete operational abilities. Further, Siemann and Delius (1993) have shown that for adult humans who learned to choose between pairs of doors during an exploration-type computer game, there was no difference in the performance of individuals who formed explicit comparison models and those who did not ($N = 8$ vs. 7 respectively; the presence of explicit models were checked by verbal report at the end of testing). All of these results cast doubt upon the belief that all TP reflects true inference by the subject.

Characteristic TP effects

Besides the 'inference' itself, TP is characterised by a number of behavioural phenomena which have been taken to indicate something about the cognitive processes underlying the behaviour (Bryant and Trabasso 1971). Some researchers have questioned the significance of these effects, particularly the temporal aspect of the symbolic distance effect (McGonigle and Chalmers 1992; Rapp et al. 1996). Nevertheless, the following effects[1] have been shown broadly across experimental subjects, including children, monkeys, rats, pigeons, and adults (Wynne 1998).

- *The end anchor effect (EAE)*: subjects make an evaluation faster and more accurately when a test pair contains one of the end stimuli.
- *The serial position effect (SPE)*: even taking into account the EAE, subjects do more poorly the closer the stimuli displayed are to the middle of the sequence.
- *The symbolic distance effect (SDE)*: even compensating for the EAE, the further apart on the series two stimuli are, the faster the subject makes the evaluation. This effect is generally taken to contradict any step-wise chaining model of transitive inference (i.e. Piaget's concrete operations), since distant stimuli would require more steps and therefore a longer reaction time (RT).

Training subjects for TP

Training a subject to perform TP is not trivial. Subjects train on ordered pairs, typically in batches. Because of the EAE, there must be at least five items $(A, \ldots, E)$ to demonstrate transitivity on just one untrained pair (BD). Seven or more items would give further information, but successful training is notoriously difficult to achieve and even children who can master five items often cannot master seven. This is true even for simple sorting of ordered items such as posts of different lengths (McGonigle and Chalmers 1996). Normally, though, stimuli are labelled in a non-ordinal way, such as by colour 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*).

[1] This is a subset of all reported TI effects, see further Wynne (1998) or Shultz and Vogel (2004).

Table 1 Phases of training and testing used for children, taken from Chalmers and McGonigle (1984, pp. 359–360)[a]

Phase	Training and criteria
P1	Each pair in order (*ED*, *DC*, *CB*, *BA*) repeated until 9 of the 10 most recent trials correct. Reject if requires over 200 trials total.
P2a	4 of each pair in order. Criterion: 32 consecutive trials correct Reject if requires over 200 trials total.
P2b	2 of each pair in order. Criterion: 16 consecutive trials correct Reject if requires over 200 trials total.
P2c	1 of each pair in order. Criterion: 30 consecutive trials correct No rejection criteria.
P3	1 of each pair randomly ordered. Criterion: 24 consecutive trials correct. Reject if requires over 200 trials total.
T1	Diad 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	Triad tests: 6 sets of 10 triads in random order, reward for all
T3	Extended version of T2

[a]Methodology for monkeys is similar

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.[2] When subjects reach criterion, they are trained on *CD*. After all pairs are successfully trained, there is usually 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 phases P1–P3 in Table 1).

Once subjects are trained to criterion, they are exposed to test pairs. In order to ensure that there is no training of the test pairs, they are meant to be "nondiscriminatively rewarded" (Bryant and Trabasso 1971). McGonigle and Chalmers (1977) rewarded either choice on test pairs. This is presumably the least disruptive non-discriminative reward schedule, because whatever item is chosen is the one the subject is most likely to have expected to be rewarded for, and since learning tends to occur when expectations are violated, it is less disruptive to meet those expectations. Training pairs are often interspersed with test pairs during the testing phase, with the training pairs still being differentially rewarded. This has been found necessary to maintain performance on the original training pairs.

Triad data sets

Table 1 finishes with a set of triad testing. These tests are to date unique to the work of McGonigle and colleagues. A triad test presents three rather than two stimuli drawn from the set of stimuli for which subjects reached training criterion. Trigram tests were originally designed to test McGonigle and Chalmers's Binary Sampling Theory of diadic TP (see Appendix A). Most subjects exhibit systematically degraded TP on trigram tests.

Triadic testing has been criticised on the grounds that the sudden presence of three items might confuse the subjects and so degrade their performance. This criticism was addressed by McGonigle and Chalmers (1992) when they repeated their 1977 experiments to gather more data on RTs. In 1992 they also tested their subjects on pseudo-trigrams, in which one of the stimuli is presented in duplicate (e.g. *A*, *A*, *C*; *B*, *D*, *D*). Subjects showed no significant performance degradation in this case. The quality of the data set is further supported by the fact that it was accounted for extremely well by the model of Harris and McGonigle (1994), which we describe next.

The production-rule-stack model

Our two-tier model was inspired by the best previous model of the triad data set. This model is due to Harris (1988). The production-rule-stack model is a static, non-learning model of fully-trained subjects which accounts for the McGonigle and Chalmers (1977) triad data, both in aggregate and as an explanation of individual performances. This work helped motivate the McGonigle and Chalmers (1992) study, and was ultimately published by Harris and McGonigle (1994).

The Harris model is based on a production-rule stack. *Production rules* come from artificial intelligence. They are representations which tightly associate particular contexts or *sensory preconditions* with particular actions. Preconditions indicate when a context is appropriate for an individual to express the associated action. A *stack* is a common representation from computer science. As the name suggests, it is a set of objects which have to be visited in order: beginning with the first item at the 'top' of the stack, the *n*th item must always be examined before the $n + 1$th item. With a production-rule stack, production rules are checked in order beginning from the top of the stack. If, when checked, a production's precondition is met by the environment then the associated action is expressed. For example, a precondition might be *able to see A* and an action might be *grab the item holding your visual attention*.

The Harris production-rule-stack model requires the following assumptions:

I. The subject knows a set of rules of the nature "if *A* is present, select *A*" or "if *D* is present, avoid *D*".
II. The subject has a prioritisation of these rules.

For an example of the model, consider a subject with the stack:

1. (*A* present) ⇒ select *A*
2. (*E* present) ⇒ avoid *E*
3. (*D* present) ⇒ avoid *D*
4. (*B* present) ⇒ 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 in the test pair. However, rule

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

3 indicates the subject should avoid D, so consequently it selects C.

Harris and McGonigle make one more critical assumption:

III. When there are more than two items present (as in the triad test cases), an *avoid* rule results in random selection between the items not currently attended to.

For example, consider the situation where there are three blocks available B, C, E. If the subject is applying the rule 2 above, and has found and attended to a block E, the 'avoid' action means that it is equally likely to grasp either B or C. This assumption explains the performance degradation of children and monkeys shown in the triad data.

The Harris and McGonigle model statistically accounts for the aggregate monkey data. For example, over all possible triads, the production-rule-stack hypothesis predicts a distribution of 0, 25 and 75% for the lowest, middle and highest items respectively. True inference of course predicts 0, 0 and 100% respectively. The squirrel monkeys (*Saimiri sciureus*) in McGonigle and Chalmers (1977) showed 1, 22 and 78% respectively. Further, Harris (1988) was able to match the individual performance of most monkeys to a particular stack.

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

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

always selects B from the triad BCD by using rule $2'$, while the previous stack selects B 50% of the time and C 50% because it bases its decision on rule 3.

There are eight discernible correct rule stacks of three rules each, all of which reach criterion in training and exhibit TP on all test pairs. There are actually 16 correct stacks of four rules, but triad experiments cannot discriminate whether the lowest priority rule selects or avoids (Harris and McGonigle 1994, p. 325).

Model

Our model has two tiers of learning: one to associate actions with stimuli (d in Fig. 1), and another to prioritise association 'rules' (b in Fig. 1). For both learning tasks, priority is represented with a weight, the value of which is learned through reinforcement. Larger weights correspond to higher priorities. Which 'rule' is fired is determined first by which stimulus present is associated with the largest weight, and then by which actions in that stimulus' associated action list has the largest weight.

Our goal is *not* to improve on the Harris and McGonigle (1994) outcome for modelling trained monkeys, since that fit is already strong. Rather, our interest is in modelling how that performance develops in live subjects.

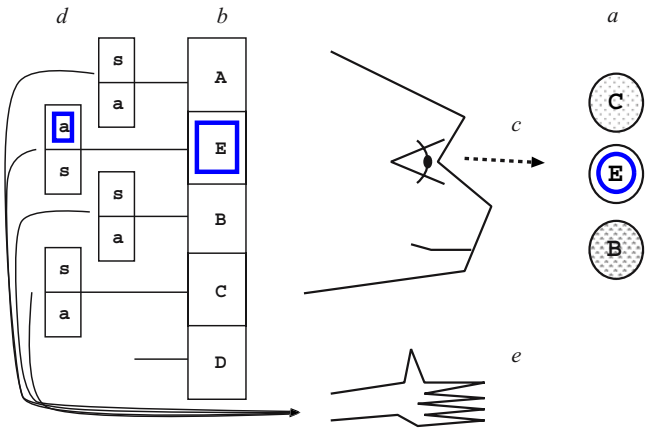

Fig. 1 The two-tier model. When the subject observes a set of stimuli a a weight vector (b, the first tier) determines which item present is most salient. This attracts visual attention c and determines which rule vector (d, the second tier) selects the appropriate action (select or avoid). This determines what item the subject grasps e. The two vectors that were most recently active (b and one of d) are then updated in response to the reward as per Eq. (1)

The two-tier model comprises a list of perceptual categories corresponding to the different stimuli seen. Half of the subject's task is to prioritise this list. The other half of the task (the second tier of the model) is prioritising the actions associated with each stimulus. The same learning rule is applied to both tasks (see Eq. (1) below).

Figure 1 illustrates how a subject chooses a stimulus under the two-tier model. At the beginning of a trial, the subject is presented with a number of stimuli, generally two or three. If any of the stimuli are novel, they are added to the first tier by a process described below. Next, the subject attends to the stimuli present with the highest priority. The subject then examines the second tier of action prioritisation to express the highest-priority associated action. The subject either selects the object it is attending to, or 'avoids' that object by grasping another object. If there is more than one other object present and the subject is 'avoiding,' then the object that is to be grasped is randomly chosen. For both tiers, whenever more than one eligible tier element has the same priority, one of these tier elements is selected at random.

After a stimulus is selected, the subject is either rewarded or not as appropriate. Weights for both tiers are updated independently after every trial. We use a simple step function for weight adjustment which roughly approximates known conditioning models (Waelti et al. 2001; Rescorla and Wagner 1972).

As mentioned earlier, the same learning rule is used for both tiers. All of the weights in a single list are normalised, so that they always sum to 1. When a new stimulus is seen, it receives the weight $1/N$, where N is the current number of distinct stimuli categories so far seen. New items in the stimulus list are further associated with a new action list, which is initialised with two actions, *select* and *avoid*, both of which are given a starting weight of 0.5.

The weights for any particular list (the stimuli list in the first tier or one of the action lists associated with a single

stimulus in the second tier) are represented as a vector $\mathbf{w}$. Consider the pair XY, where X is the list element the subject attended to and Y is a near alternative,[3] then $\mathbf{w}_X$ and $\mathbf{w}_Y$ are the weights associated with X and Y respectively. The update rule for these weights is:

$$\text{If } X \text{ is correct and } (\mathbf{w}_X - \mathbf{w}_Y < \tau), \text{ add } \delta \text{ to } \mathbf{w}_X;$$
$$\text{else, if } X \text{ is incorrect, add } \delta \text{ to } \mathbf{w}_Y. \quad (1)$$

where τ and δ are free parameters, held constant for any particular subject, but varied across subjects for the experiments. τ is a threshold, over which reward is so expected that it no longer prompts learning (Waelti et al. 2001). δ is the amount a weight is changed by a single bout of learning. If weight change occurs, $\mathbf{w}$ is subsequently renormalised.

Methods

Model testing through artificial life simulation

The experiments in this paper were performed using artificial life (ALife) simulations. An ALife model can be thought of as a conventional (though meticulously specified) hypothesis. ALife experiments operate by running simulations, then performing standard hypothesis testing to see whether the simulated results are a good match to the original data. For clarity, we will follow the convention of referring to the artificial subjects in these experiments as *agents*.[4]

Once a model has been built, the process of simulation 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. If desired, a relatively sparse set of these predictions, perhaps those that are most surprising or vary most between different versions of the hypothesis, can then be tested against experiments with living subjects. 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. Further technical details of the simulation can be found in Appendix B.

Experimental overview

To fully motivate the two-tier model, we run a preliminary experiment with a single tier. This roughly corresponds to existing models (e.g. Wynne 1998; Delius and Siemann 1998; Frank et al. 2003; Shultz and Vogel 2004) which order the stimuli rather than productions. More elaborate analysis and comparisons with these models are made later in the Discussion section. Our second experiment shows how the two-tier model works without being exposed to the sort of phased learning required by primates, while Experiment 3 shows the two-tier model *with* such training. The final experiment shows that our model also accounts for children's performance on learning non-overlapping pairs of stimuli ($A > B, C > D$).

Experiments

Experiment 1

Procedure

In the first experiment, we did not use the full two-tier architecture, but rather tested the learning algorithm shown in Eq. (1) in a single-tier model. There was only a single vector with each element corresponding to a stimulus. The agent chooses the stimulus corresponding to the vector element with the highest-priority weight.

For training we exposed the artificial subjects to all training pairs in a random order for approximately 400 trials. This training regime is sometimes used with rats (Wynne 1998).

Results and discussion

The single-tier system always learns to order the stimuli perfectly, provided that τ is small enough. The last error by these systems is normally made by the 100th trial and weights stop fluctuating (or *stabilise*[5]) about 50 trials later.

The ability to pass criterion consistently without a training regime is very unlike primates, which normally need a training regime and which do not always pass criterion. Also unlike primates, this model performs perfectly on triads, since once the ordering is learned it will always select the highest priority stimulus.

Figure 2 shows a typical result for when $\sum_{n=1}^{N} n\,\tau \leq 1$. If $\tau > 0.1$, then a stable solution for five items cannot be reached. This is because there is no way that five weights can be more than 0.1 different from each other and still sum to 1 (see Eq. (1) and the discussion of normalisation). If $\tau = 0.1$, then it is possible that the weights can be [0, 0.1, 0.2, 0.3, 0.4] which sums to 1. For any value of $\tau < 0.1$ there are many possible stable solutions for learning the ordering of five items.

If learning cannot stabilise, the model's behaviour is open to a 'hot hands'-like phenomenon (Gilovich et al. 1985), where a solution that has recently been very successful may get more weight than it deserves. When there is a chance reiteration of one particular pair, the higher element of that

[3] If X is a stimulus, Y is one of the other stimuli present chosen randomly, unless the rule was *avoid* in which case it is the item actually grasped. If X is an action then Y is the second-highest priority action for that stimulus.

[4] The term *agent* is actually intended to refer to any actor, artificial or not, but it has become associated with AI software systems.

[5] Normally in AI, agents are considered to have fully learned a task only when their weights have stabilised (stopped changing). This of course can not map directly to the animal research, where learning must be judged by expressed behaviour.

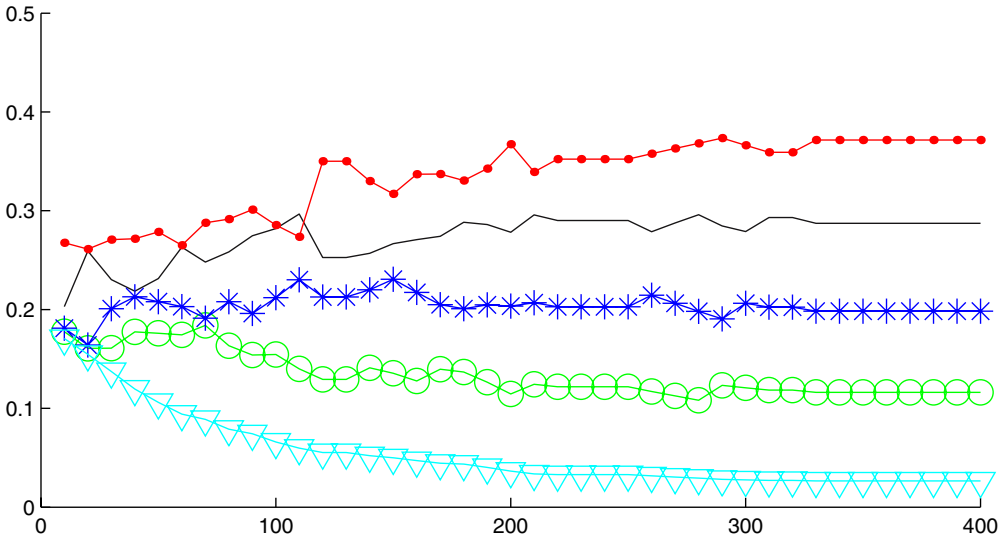

Fig. 2 A typical result for a one-tier learning agent. X-axis: trial number; Y-axis: weights of the vector element corresponding to each stimuli (which sum to one). Free variables are set to parameters: $\tau = 0.08$, $\delta = 0.02$. The key for the lines (in order from top down as of trial 400) are A (•), B (–), C ($*$), D (◯), E (▽)

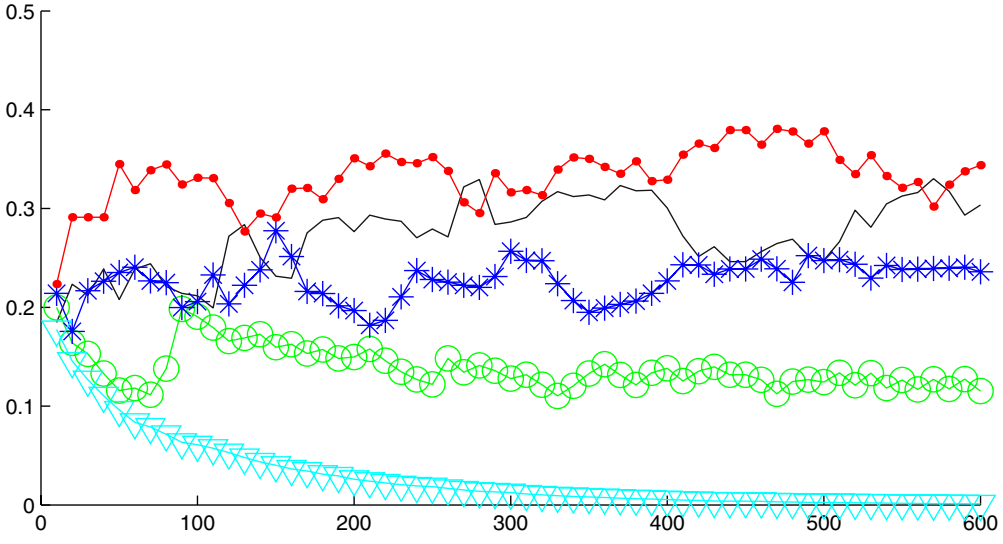

Fig. 3 One-tier learning in an agent with a 'stupider' parameter set: $\tau = 0.12$, $\delta = 0.02$. This cannot find a stable solution (see text) thus occasionally gives wrong responses. Key (top down for trial 600): A (•), B (–), C ($*$), D (◯), E (▽)

pair can accumulate so much reinforcement that its weight surpasses the element that should be above it. Thus the agent over-estimates the value of an item because of its recent 'winning streak'. This is illustrated in Fig. 3. Even so, however, these agents very rarely make mistakes and so would easily pass training criterion.

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.

Experiment 2

Procedure

In the second experiment we used the two-tier model, but still trained it by presenting training pairs in random order. We tested learning in the two-tier model across a range of parameter values: every combination of τ drawn from $\{0.08, 0.1, 0.12, 0.14\}$ and δ drawn from $\{0.01, 0.02, 0.04, 0.08, 0.12, 0.16\}$. Twelve subjects were run with each possible parameter combination for a total of 288 subjects.

Results and discussion

Only about one-fifth of two-tier agents learn the training pairs entirely successfully (56 of 288, see Table 2, column 2 below). We do not know how this corresponds to primate subjects without a training regime as such results have not been reported, but we can assume that primates also fail frequently, as this would justify the use of elaborate training procedures.

Agents that learn the training pairs successfully perform on triad testing exactly as described by Harris (1988) because a snapshot instance (that is, one with learning frozen) of a successfully trained two-tier model is logically equivalent to a production-rule stack.

Table 2 Production-rule-stack equivalents to solutions by *Saimiri sciureus* subjects (last column) and by two-tier AI subjects undergoing various forms of training[a]

		Regime starting ED		Regime starting AB		Starting AB
	No regime	After training	After testing	After training	After testing	McGonigle and Chalmers (1992)
$s(A)s(B)s(C)$	8	51	41	–	–	–
$s(A)s(B)a(E)$	12	68	26	–	–	–
$s(A)a(E)a(D)$	3	–	1	4	2	2
$s(A)a(E)s(B)$	7	4	16	3	1	2
$a(E)a(D)s(A)$	9	–	1	57	50	–
$a(E)a(D)a(C)$	8	–	–	59	47	1
$a(E)s(A)a(D)$	7	3	–	4	11	–
$a(E)s(A)s(B)$	2	1	13	–	3	–
Total correct	56	127	98	127	114	5
Total	288	144	144	144	144	7

[a]The distribution of solutions for two-tier agents is strongly determined by the order training pairs are presented. The analysis of the live monkeys' correlated stacks reported in the last column was performed by Harris and McGonigle (1994)

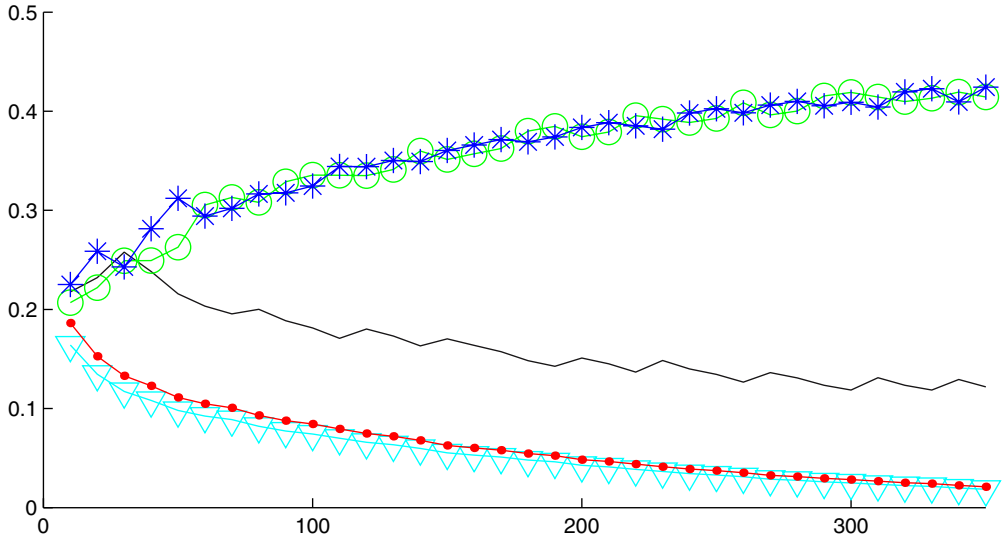

Fig. 4 Two-tier learning with no training regime. Rules learned in descending order of priority: select D (◯), avoid C ($*$), avoid B (–). The rules of the bottom two stimuli are rendered insignificant, see text. The system cannot stabilise because it is far from a complete solution, but it behaves correctly for every training pair except CD. Parameters: $\tau = 0.08$, $\delta = 0.02$

Figure 4 illustrates a typical, failing result. Rule selection is made very early in training and remains stable once established, so is not depicted in these figures but rather only in the accompanying legends. Nevertheless, the added complication of rule learning defeats the simple training regime used in the one-tier experiments. With this training regime, the two-tier model generally learns either the solution shown in Fig. 4 or a symmetric one with the select B and avoid C rules fighting for top priority.

There are only a limited number of accurate solutions the two-tier model can have, corresponding to the correct rule stacks enumerated by Harris (1988). A correct solution must be either an ordered sequence of selects $[s(A)s(B)s(C)]$, a reverse order sequence of avoids $[a(E)a(D)a(C)]$ or an ordered cross of these (e.g. $[s(A)a(E)s(B)]$). See Table 2, column 1 for a complete list.

Although the rules learned by the typical, failing two-tier agents have a very different order than just described, they still perform well on the training task. For each failing agent, only one training pair is incorrect: that containing the two top-priority stimuli. For example, in Fig. 4, the only training pair which cannot be handled is CD. Notice that although the weights in Fig. 4 have not stabilised, the behaviour has. Whether $a(C)$ or $s(D)$ is highest priority, the agent will incorrectly grasp D when presented with C, D. Such agents display something like the SPE by confusing only central pairs. Taken in aggregate, the agents display the full SPE: the most frequent errors involve the two central pairs, but there are occasionally errors involving other elements.

All the agents show the EAE. Agents quickly learn rules which avoid making errors involving the two end stimuli. For example, the agent in Fig. 4 may appear to neglect the two end stimuli, since the weights of the stimulus-rule pairs associated with those stimuli are very low. But in fact, this agent gets the end pairs (AB and DE) correct 100% of the time by associating the rule avoid with B and select with D. By reducing the priority of the A and E rules, the agent learned the 'correct' behaviour in the end-term cases. Associating this knowledge with the inner member of the end pair protects the agent from a possible incorrect rule

associated with the outer member. However, this strategy leaves such agents with no possible means to correctly learn both middle pairs. The learning system fixates on trying to resolve an impasse in the middle of the sequence, but the learning algorithm, based on gradual change, cannot solve that quandary.

Nineteen percent of the time two-tier agents without a training regime learn a correct solution. If successful learning were the simple consequence of the agents being at chance for learning a rule about either the inner or the outer element of the two end pairs, we would expect that agents would learn both ends correctly 25% of the time, only one end 50% of the time, and neither end 25% of the time. We can dismiss this as the full explanation for the agent's failure: $\chi^2(3, N = 288) = 35.68$, $p < 0.001$. The fact that the inner end-pair stimuli, B and D occur in twice as many pairs as the end stimuli leads the two-tier model to the case of using both inner rules 40% of the time (166 in 288), not 25%. When correct solutions are learned, they come evenly from all parameter values, and seem evenly distributed across all possible correct solutions (Table 2, column 2).

We would obviously like to compare these results with the outcomes of live primate subjects who fail to meet criterion on the initial training for transitive learning. Although no triad results were reported for monkeys or children that missed criterion, one monkey subject, Roger, passed criterion but still showed a consistent error between the third and fourth item (Harris and McGonigle 1994, p. 332). Roger's errors are in keeping with the results of this model.

Experiment 3

Procedure

In the third experiment, we trained two-tier models using the training regime in Table 1. We used the same range of parameters as for the previous experiment again with 12 instances of each for a total of 288 AI subjects.

Results and discussion

Summarised in Table 2, our results are that 88% (254 of 288) of the AI agents successfully learn the training pairs and therefore the TP task. This is slightly better than the live, monkey subjects did, though not significantly given the small number of monkeys ($\chi^2(1, N = 7) = 0.42$). Nearly all successful software agents converge quickly, and the ones that fail to meet criterion fail early, usually by Phase 2a.

Successful learning for agents with phased training is highly dependent on δ; when δ was large (values in 0.08, 0.12, 0.16), one in four agents failed, which is actually a tighter fit to the monkey failure rate than our full aggregate result. Otherwise there were only a very few failures (3), all of which had the lowest tested δ, $\delta = 0.01$ ($N_{\delta=0.01} = 48$). Since δ determines the rate of weight change after training, it is unsurprising that a very low δ results in a slow learner. Even when such agents pass criterion, they may never learn a stable solution (see e.g. Fig. 6). Interestingly, agents do sometimes learn when $\delta > \tau$, which means that for any trial on which learning occurs, the attended item will change places in the priority stack.

One advantage of the two-tier model over a single learning tier can be seen in the fact that during initial training stable representations are learned that involve rules with nearly the same priority. This is because some rules will never be compared. For example, items E and A do not occur together in any of the training pairs, so there is no pressure to differentiate their weights prior to TI testing (see Fig. 5, phases P2b–P3). This is significant if neural systems have limited capacities to discriminate different stable orderings (see discussion for Experiment 2, and Cowan 2001; Bryson and Lowe 1994; Gallistel et al.1991). For tasks involving stimuli which never co-occur, the rule representation allows for stable learning with either more items or larger values of τ. A more natural example of such a task than TI would be navigation, where some landmark features might never occur in the same place.

Another thing to notice in the phased learning results is that significant learning occurs during testing. This phenomenon was also reported with monkeys (Harris and McGonigle 1994). Learning occurs because rules that were never compared (e.g. those triggered by any two non-adjacent items) previously are now compared. If their weights do not already happen to be at least τ apart, learning is triggered, regardless of whether they were correct or how they are reinforced. This explains the utility of continuing to differentially reinforce training pairs, a common procedure during the testing phase of the TI task.

Experiment 4

Procedure

De Lillo et al. (2001) model an experiment by de Boysson-Bardies and O'Regan (1973) related to TP which gives information on how children represent training pairs. In their Experiment 4, de Boysson-Bardies and O'Regan presented children with two non-overlapping pairs, AB, CD.

de Boysson-Bardies and O'Regan were testing an explanation of children's TP which they call the *labelling strategy*. This first assumes the children associate the labels 'big' and 'small' with items in pairs, and then posits a set of transformations to explain performance on interior items. They test their theory by training children *only* on the pairs $A > B$ and $C > D$, then testing them on all possible combinations of stimuli (see the first two columns of Table 3). Their labelling-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 $C > D > A > B$).

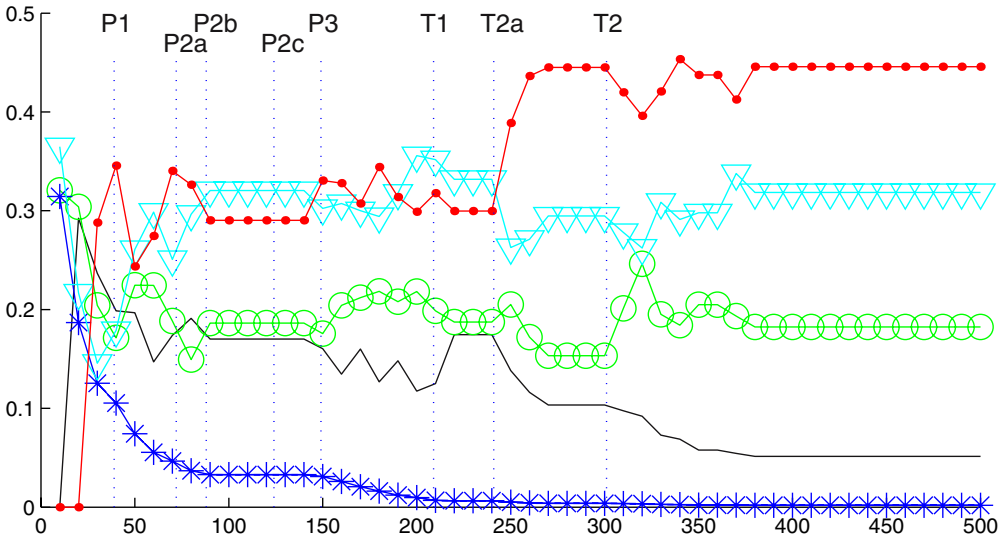

Fig. 5 Rule Learning with Phased Training. Labelled lines indicate the *end* of training and testing phases (see Table 1). This agent arrives at different stable solutions at different points, but they are all correct. This solution is an example of select (A, •), avoid (E, ▽), avoid (D, ◯), see Table 2. It also selects (B, –) and no rule is learned for **C** ($*$). This agent succeeds with very 'stupid' parameters: $\tau = 0.12$, $\delta = 0.06$

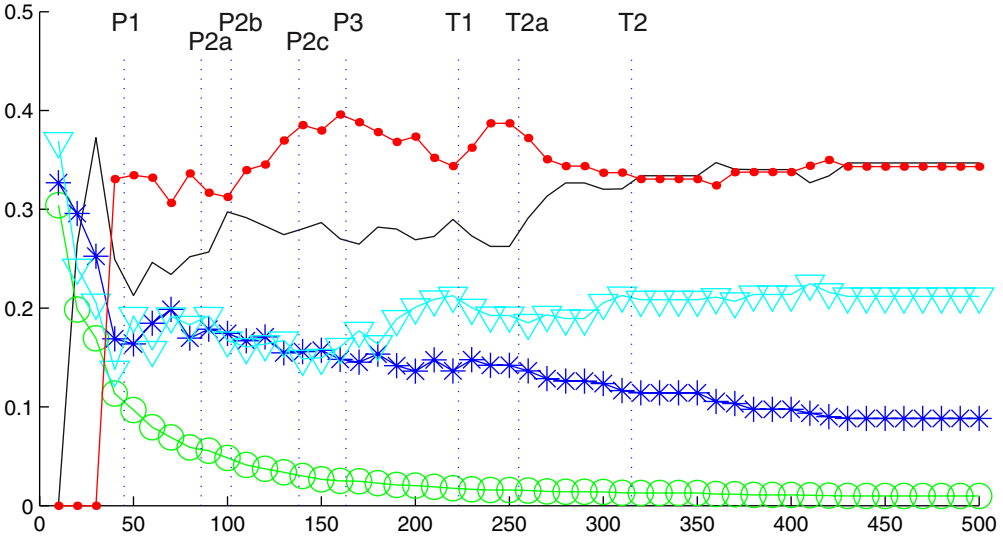

Fig. 6 Phased Training where learning slips during triad testing. Rules: select A (•), select B (–), and either avoid E (▽) or select C ($*$). Parameters: $\tau = 0.12$, $\delta = 0.02$

Table 3 Comparison of results from de Boysson-Bardies and O'Regan (1973) Experiment 4, their labelling hypothesis, and the two-tier model[a]

Pair	Subject scores	Labelling hypothesis	Two-tier hypothesis
A > **B**	88	*100*	*100*
$A > C$	52	50	50
$A > D$	85	100	87.5
$B > C$	27	0	12.5
$B > D$	50	50	50
C > **D**	87	*100*	*100*

[a]The subjects were deliberately not trained to full performance on the training pairs (AB and CD) but the models do not take this into account. See Table 4 for explanation of the two-tier values

Modelling this experiment required no actual runs of the two-tier model, because the results can be determined analytically, see Table 4.

Results and discussion

The two-tier model on this task predicts that *for each pair* the children learn

1. to focus on one item, and
2. to either select or avoid that item.

No rule will be learned for the other item, because the weights learned from the first training session will determine the rest of the outcomes. Given this model, and assuming that all rules are equally likely to be learned across all items, aggregate data for the two-tier model actually matches the de Boysson-Bardies and O'Regan 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 25% of the children will have no applicable rule.

Table 4 Explanation of two-tier predictions from Table 3[a]

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	
	None	50	
$A > D$	$s(A), a(D)$	100	**87.5**
	$s(A)$	100	
	$a(D)$	100	
	None	50	
$B > C$	$s(B), a(C)$	0	**12.5**
	$s(B)$	0	
	$a(C)$	0	
	None	50	
$B > D$	$a(B), a(D)$	50	**50.0**
	$a(B)$	0	
	$a(D)$	100	
	None	50	
C > D	$s(C)$	100	**100.0**
	$a(D)$	100	

[a]For each training pair, we assume each subject is at chance for learning one rule (either a select or an avoid) to solve it. Since the two rules are learned in isolation from each other, we also assume that when both happen to be applicable, which rule has higher priority is at chance. Predictions are for the percent chance of selecting as if the rule shown under 'pair' holds. Individual predictions are based on possible rule sets an individual might have acquired in training; aggregate predictions assumes an even distribution of individuals

General discussion

We have achieved our goal of showing that a system like that of Harris and McGonigle can be learned, and with a simple, biologically-plausible learning algorithm.

Moreover, in doing so we have a produced a model that displays the End Anchor and Serial Position effects, and requires the same phased training that children and squirrel monkeys require to have a similar number of subjects pass criterion. That these features of the model were unintended consequences of the two-tier structure further validates both our model and the work of Harris and McGonigle (1994).

The parsimony of the two-tier model

The two-tier model, while apparently more baroque than some other models (e.g. Wynne 1998; Delius and Sieman 1998), is actually parsimonious in that it also addresses how animals know whether or not to represent stimuli as elements of a sequence. In the two-tier model, they don't need to. The sequencing is of behavioural priorities, not of items, and ordering behaviour priorities *is* a natural and obligatory aspect of a multi-step task. It is important that steps nearer the goal of the task have higher priority so that an individual is able to take advantage of opportunities. Thus the consummatory actions should always have the highest priority (Tyrrell 1993; Bryson and Stein 2001). Previous research has shown that primates are capable of taking advantage of such opportunities, whereas pigeons cannot (Terrace and McGonigle 1994).

Nevertheless, the quality of the fit still seems surprising. First, why would subjects focus only on a single item as the precondition of an action? Why would they not instead learn the appropriate action for each pair of stimuli when this would give a more accurate result? Second, why would an 'avoid' action which non-deterministically chooses between any other options be one of only two possible actions?

We propose a possible explanation based on considering the full individual history of the acquisition of the TP task:

1. In the earliest stages 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. When subjects are first presented with pairs, they are initially trained on extended blocks of trials using just a single pair. Subjects discover that they may select only one object. We propose that at this stage subjects learn two things: to discriminate stimuli, and also to inhibit the *select* action in some contexts. This inhibition results in the *avoid* action being learned and associated with some objects that, for whatever reason of individual history or preference, are highly salient (hold the subject's attention).
3. Finally, as they are exposed to more pairs of stimuli, subjects must learn (or adjust) prioritisation between neighbouring inhibition rules. This is the stage of learning that the two-tier model models.

The origin of the single-item cue is the original learned behaviour pattern: the grasping of a single item. The origin of the two actions expressed by the subjects is also accounted for.

This explanation assumes that animals are innately conservative in what they try to learn about, adding no more features than are strictly necessary. Such a strategy is parsimonious, and also reduces the degrees of freedom of the problem, thus increasing the probability that the animal will learn successfully.[6]

Our model is also parsimonious in that it provides for nearly the entire set of TP effects described in Characteristic TP effects section, except for the SDE. Accounting for the SDE as well requires a further assumption: that the process of selecting between two rules takes longer when their priorities are less than τ apart, and that this effect is exaggerated the closer the priorities are to each other. For example, the 'decision' to act on one rule may require a build-up 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 both take longer and be more arbitrary. This is essentially an elabora-

[6] We use the term *strategy* here to mean an innate, evolved solution, not something intentionally selected by the subjects.

tion of a common assumption expressed in existing simple associative models (Wynne 1998), and in other temporal models based on neural activation (Glasspool 1995; McDonald and Lowe 1998). Shultz and Vogel (2004) present a particularly nice model of this extra assumption as an additional layer to their TP learning system. The small additional network they use accounts not only for the SDE, but also for the contiguity effect—that humans respond more quickly when the question (e.g. "which stick is *longer*") is correctly answered by an end-anchor item present in the current pair (in this case, the longest stick seen in training).

Such a layer for solving the SDE could easily be added onto the two-tier model. In the case of the two-tier model, the result of this process would not produce the typical SDE for every individual subject, because which pairs are ambiguously close to each other depends on which of the possible solutions the individual subject has learned, and how well. However, the SDE appears to be an aggregate rather than an individual effect (McGonigle and Chalmers 1992). We believe that an aggregate SDE would emerge from our models since they display the serial position effect, but we have not modelled this yet. We intend to explore the SDE issue further in future work.

Related research

There is a vast cognitive modelling literature that uses production rules. This literature is generated primarily by two communities that use related technologies for creating their models: Soar (Newell 1990) and ACT-R (Anderson 1993). Of these, ACT-R is more similar to the two-tier model since it also provides a single representational means of ordering applicable productions. However, the learning system that prioritises rules for ACT-R is not geared for creating a total ordering of applicable rules with respect to each other, but rather learns a *utility value* for each individual rule which depends on its probability of success. For the TI task, every item except the two end items will have the same utility, thus the 'rule stack' learned is quite different from what is learned by our system, and essentially only has two variants depending on whether *select* or *avoid* rules wind up dominating.

Wood et al. (2004) have examined ACT-R on the TP task including triad testing. We found that two of the five monkeys described in (Harris and McGonigle 1994) matched one of the possible ACT-R solutions closer than a fully ordered production rule stack such as Harris originally postulated, while the other three did seem to learn fully ordered production stacks. The representations underlying both the two-tier model and ACT-R could in theory represent either solution. However, the basic theory behind ACT-R and its actual learning system mandates the utility-based solution only. The learning rule for the two-tier model will only stabilise (that is, stop learning) if the weights attain a strong ordering. However, as Fig. 4 shows, other sorts of equilibria can also be found by this system which might approximate the ACT-R response.

De Lillo et al. (2001) present a backpropagation model which displays the SDE (see Characteristic TP effects section), but does not account for the McGonigle and Chalmers (1977) triad results. Backpropagation 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, De Lillo et al. (2001) use a two-layered approach. The difference in outcomes between these models shows the importance of not only the basic architecture but also the learning algorithm and primitives.

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 De Lillo et al. model were connected to *select* and *avoid* instead of ordinal positions on the board (*left* and *right*), their model would be fairly similar to ours, and might even learn the task. However, for the two-tier model 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 triad case. Further, the main purpose of backpropagation as an algorithm is to avoid the kind of mistakes that the two-tier system made when not trained with phased training. This is done by sharing learned information across the layers of the system. Our research gives evidence that in real primates, there are two discrete learning systems involved in TP. As mentioned in the introduction, this theory has been strongly supported by the neuroscience literature, including both fMRI and lesion work (Heckers et al. 2004; Alvarado and Bachevalier 2000).

Shultz and Vogel (2004) present a recent model of *all* known TP effects. Like ours, their model is feed forward with no propagation, and split into two different layers. However, all of the pair learning and TP is generated in the first layer which is a simple single-layer network.[7] As such, the Shultz and Vogel (2004) model can neither exhibit any failures to reach criteria, nor replicate the triad data. The second layer of their network is dedicated to representing a competitive response network (choosing the 'left' or 'right' item when asked for 'longer' or 'shorter' element). This is the layer which accounts for the SDE and contiguity effect, as reviewed earlier. If this second layer of the Shultz and Vogel (2004) model were added as a *third* tier of our model, the new three-tier model might also display CE and SDE.

O'Reilly and Rudy (2001) proposed a *Coordination Account* model of transitive inference which, like all the models reviewed here, is not based on logical inference. However, their model is more like the Binary Sampling Theory, hypothesising that intervening items were effectively 'imagined' by the hippocampus. Also like the Binary

[7] Shultz and Vogel refer to this layer as a Cascade Correlation (CC) network (Fahlman and Lebiere 1990). However, the single-layer TI problem is so simple (as we demonstrate in Experiment 1) that the CC algorithm never adds any hidden units.

Sampling Theory, O'Reilly and Rudy (2001) violate the serial position effect and the SDE.

When testing their theory on live rats, Van Elzakker et al. (2003) rediscovered both the SDE and the contiguity effect.[8] These results inspired a theory similar to the value transfer theory (von Fersen et al. 1991; Zentall and Sherburne 1994), where each item elicits an excitation roughly monotonic with its place in the sequence. This excitation results not from a sequential representation, but from simple association. Firstly, the end items are strongly reinforced (either positively or negatively), while all intermediate items are reinforced roughly equivalently, as per simple conditioning. Second, the items *near* the end items have their weights shifted perceptibly up or down, due to frequent association with either a very positive or a very negative item. This association leads to successful TP on items that are sufficiently far apart, but has little effect on central pairs (particularly on long lists such as 6-pair, 7-item sequences).

Frank et al. (2003) then built a second model of TI reflecting these results. Similar to De Lillo et al. (2001), both the 2003 and the 2001 models used backpropagation for their learning regime, though in a far more elaborate (and modular) architecture intended to replicate brain structure. The 2003 model extended the 2001 to include an additional differential learning aspect for the stimuli which "swamped" (p. 352) the part of the model (still embedded) which had previously run counter to the SDE. However, again, this model relies on simple ordering of the stimuli so cannot replicate Harris' match to individual performance on triads.

The Frank et al. (2003) model could presumably be extended again to learn orderings of performance rules instead of stimuli. However, at least for pigeons, there is strong evidence of value-transfer behaviour (Siemann et al. 1996). It is also well established that primates and pigeons represent sequences differently (Terrace and McGonigle 1994), so differences in their TP are perhaps less surprising than a difference between primates and rats. However, unlike testing for both human and non-human primates, standard procedure for testing and training rats on TP requires using olfactory cues to dig directly for food. This may trigger a different, possibly older, olfactory-lobe learning system (c.f. Hurliman et al. 2005). Certainly some olfactory learning of food cues in rats is amygdalic (Bermúdez-Rattoni et al. 1983), a fact overlooked in some early hippocampal modelling work (McClelland et al. 1995).

We now turn to existing criticism of the Harris (1988) model, which underlies our own. Van Elzakker et al. (2003) assert incorrectly that Harris' model depends on another McGonigle and Chalmers (1992) theory, the 'symbolic distance view'. This assertion is incorrect; Harris' model clearly in no way assumes an underlying sequential representation. Van Elzakker et al. also mistakenly claim that normalisation cannot account for the sort of variation in performance between series of different lengths. If the normalisation is combined with a stabilising factor (e.g. our 'significant difference', τ) their reported results (p. 339) are accounted for—the ease of learning the next inner pair from the end anchors is approximately the same regardless of series length. This is significant with respect to our model only due to our claim about the SDE above; for Van Elzakker et al. (2003) nowhere address a model actually like Harris'. Van Elzakker et al. principle objection is to those who think that the hippocampus is performing 'inference-like' computations on learned stimuli. Neither tier of our model actually performs inference-like computations. In terms of neurological correlates for our model, Heckers et al. (2004) suggests original pair learning may occur in the parahippocampal gyrus, while many researchers suggest that relational processes like prioritisation *between* rule pairs takes place in the hippocampus (c.f. Alvarado and Bachevalier 2000; Baxter and Murray 2001; Heckers et al. 2004; Buckmaster et al. 2004).

Delius and Siemann (1998) also offer a critique of Harris. While accurately describing Harris' model (p. 128), they inaccurately claim that the model would have no special problem with circular series ($A > B, B > C, C > A$). In fact, while the model would learn and perform in this context, no more than two pairs would be correctly represented at a time, and any correct answers on the third would be the result of either uncertainty or of an instability in the learned pairs. This accurately reflects how monkeys perform on this particular task: for a large number of initial trials there is one pair they seldom get right, though eventually they suddenly solve the entire problem (Alvarado and Bachevalier 2000). We assume this performance shift reflects a recruitment of new resources beyond those modelled here—possibly the identification of one pair as a separate, disjoint task. Delius and Siemann (1998) also assert that the Harris model cannot represent and compare two items from two disjoint series (e.g. $A > B > C$; $a > b > c$), but this probably false since it is a simple extension of our Experiment 4. Delius and Siemann (1998) also claim that Harris' model cannot easily be converted to a neural one, an issue our model has addressed.

Conclusions and predictions

We have presented a new model, the two-tier model, which accounts for the learning of TP in both squirrel monkeys and human children younger than 6. This is the first learning model which accounts for the systematic degradation in primate performance when subjects are presented with *three* stimuli rather than just two (McGonigle and Chalmers 1977; Chalmers and McGonigle 1984; Harris and McGonigle 1994). We provide a novel explanation for subjects' failures to pass criterion when being trained for transitive inference, which is already supported by errors observed by one monkey that passed criterion. We have also provided a better explanation for the results of de Boysson-Bardies and O'Regan (1973) on children's performance when trained on two non-overlapping pairs.

[8] In animals, the contiguity effect can only be expressed in terms of favouring the rewarded end of a series.

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

1. 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.
2. In general, RTs and visual scanning behaviour should be discernibly different for select and avoid rules.
3. If subjects who fail to pass criterion on training pairs are given triad testing, most should show a misordered priority stack with high priority rules for neighbouring pairs of non-endpoint stimuli.
4. For individual subjects, the ordering of a newly presented item (as in de Boysson-Bardies and O'Regan, Experiment 3) should be determined by the existing rule stack. For example, if the rule stack is all selects as in Eq. 2.5, a new item would be positioned last or second to last, if they were all avoids it would be positioned first.

Testing these predictions on primates requires running triad experiments after TP pair training in order to discriminate which rules were learned by individual subjects.

Appendix A: the binary sampling model

In one of the earliest responses to Bryant and Trabasso (1971), McGonigle and Chalmers (1977) not only demonstrated non-human animal learning of TP, 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 stimulus, 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 can be viewed as a naive probabilistic model—it incorporates the concept of expectation, but not in a full-fledged probabilistic framework. It proved controversial both because of lack of parsimony (or explanation) for the 'imagining' the extra items, and for apparently contradicting the SDE, since further-apart pairs may require more operations (though see McGonigle and Chalmers 1992). What is significant to the present model is that it motivated McGonigle and Chalmers to generate a data set showing the results of testing monkeys (and later children Chalmers and McGonigle 1984) on *triads* of three items. The binary-sampling theory predicts that for the triad 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). Any model using a fully sequential representation, or indeed true TP, would of course predict 0, 0 and 100%. In fact, the monkeys showed 3, 36 and 61%, respectively. Six-year-old human children showed a similar pattern on triad results (Chalmers and McGonigle 1984). While this is a fairly good match, the Harris (1988) production-rule model provides a significantly better one.

Appendix B: details of the simulation

The ALife agents were written in the Common Lisp Object System (CLOS) (Steele 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 2001), and implemented on top of a graphical software development environment for CLOS called LispWorks (Xanalys 2001).[9]

The artificial intelligence (AI) program that runs the simulations controls not only the learning 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 agents' knowledge. That is, although the testing-apparatus modules contain knowledge about the correct solution of the experiment, the test subject has no direct access to this knowledge except as evidenced by the reward.

On each trial, the testing agent (the apparatus) generates an n-gram (either diad or triad) as appropriate to the current phase of the experiment and places its element in the test-board. The learning agent (the subject) then selects one of the options by transferring it from the test-board to its hand. The apparatus determines whether the subject has chosen correctly and provides reinforcement with either a 'peanut or a 'buzzer token. The apparatus also records the trial results, and ends the trial by clearing the testing area.

[9] A personal edition of LispWorks (which runs on all platforms) is available for free download from its manufacturers, and all software used in this paper is available from the authors.

When the subject is presented with the test-board it selects an object as described (see Fig. 1). When the subject receives reinforcement, it applies the learning rule in Eq. (1) based on its current variable state (its expectations), its current context (the contents of the test board and its hand, and the rules to which it is attending) and the presence or absence of the 'peanut reward.'

In Experiments 1 and 2, an indefinite number of trials were run until the simulation was terminated by the human experimenter. In Experiment 3, the apparatus was enhanced to run the training and testing procedure shown in Table 1.

Acknowledgements We would like to thank Will Lowe, Mark Baxter, Juan Delius, Brendan McGonigle, Lynn Andrea Stein, Olin Shivers, John Mann, Emily Korvin, Mark Wood, Marc Hauser and the denizens of the Harvard Primate Cognitive Neuroscience Lab, particularly Roian Egnor. We would also like to thank our anonymous reviewers for many helpful comments and criticisms. All of the subjects used for the novel results in this article were computer programs and as such are not subject to any experimental ethics regulation either in the UK or elsewhere. However, the validity of AI models is entirely dependent on data from live subjects, and we fully support our colleagues involved in responsible animal research.

References

Alvarado MC, Bachevalier J (2000) Revisiting the maturation of medial temporal lobe memory functions in primates. Learn Mem 7:244–256

Anderson JR (1993) Rules of the mind. Lawrence Erlbaum Associates, Hillsdale, NJ

Baxter MG, Murray EA (2001) Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys. Hippocampus 11:61–71

Bermúdez-Rattoni F, Rusiniak KW, Garcia J (1983) Flavor-illness aversions: potentiation of odor by taste is disrupted by application of novocaine into amygdala. Behav Neural Biol 37:61–75

Bryant PE, Trabasso T (1971) Transitive inferences and memory in young children. Nature 232:456–458

Bryson JJ (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 JJ, Lowe W (1997) Cognition without representational redescription. Behav Brain Sci 20:743–744. Commentary on Ballard et al., Deictic codes for the embodiment of cognition

Bryson JJ, Stein LA (2001) Architectures and idioms: making progress in agent design. In: Castelfranchi C, Lespérance Y (eds) The seventh international workshop on agent theories, architectures, and languages (ATAL2000). Springer, Berlin Heidelberg New York

Buckmaster CA, Eichenbaum H, Amaral DG, Suzuki WA, Rapp PR (2004) Entorhinal cortex lesions disrupt the relational organization of memory in monkeys. J Neurosci 24:9811–9825

Chalmers M, McGonigle BO (1984) Are children any more logical than monkeys on the five term series problem? J Exp Child Psychol 37:355–377

Cowan N (2001) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. Brain Behav Sci 24:87–114

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

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

Delius JD, Siemann M (1998) Transitive responding in animals and humans: exaptation rather than adaptation? Behav Processes 42:107–137

Dusek JA, Eichenbaum H (1997) The hippocampus and memory for orderly stimuls relations. Proc Nat Acad Sci USA 94:7109–7114

Fahlman SE, Lebiere C (1990) The cascade-correlation learning architecture. In: Touretzky DS (ed) Advances in neural information processing systems 2 (NIPS'90). Morgan-Kaufmann, San Mateo, CA, pp 524–532

Fersen L, Wynne CDL, Delius J, Staddon JER (1991) Transitive inference formation in pigeons. J Exp Psychol Anim Behav Processes 17:334–341

Frank MJ, Rudy JW, O'Reilly RC (2003) Transitivity, flexibility, conjunctive representations, and the hippocampus: II. A computational analysis. Hippocampus 13:341–354

Gallistel C, Brown AL, Carey S, Gelman R, Keil FC (1991) Lessons from animal learning for the study of cognitive development. In: Carey S, Gelman R (eds) The epigenesis of mind. Lawrence Erlbaum Hillsdale, NJ, pp 3–36

Gilovich T, Vallone R, Tversky A (1985) The hot hand in basketball: on the misperception of random sequences. Cognit Psychol 17:295–314

Glasspool DW (1995) Competitive queuing and the articulatory loop. In: Levy J, Bairaktaris D, Bullinaria J, Cairns P (eds) Connectionist models of memory and language. UCL Press, London

Harris MR (1988) Computational modelling of transitive inference: a micro analysis of a simple form of reasoning. PhD thesis, University of Edinburgh

Harris MR, McGonigle BO (1994) A model of transitive choice. Q J Exp Psychol 47B:319–348

Heckers S, Zalesak M, Weiss AP, Ditman T, Titone D (2004) Hippocampal activation during transitive inference in humans. Hippocampus 14:153–162

Hertz J, Krogh A, Palmer RG (1991) Introduction to the theory of neural computation. Addison-Wesley, Redwood City, CA

Hurliman E, Nagode JC, Pardo JV (2005) Double dissociation of exteroceptive and interoceptive feedback systems in the orbital and ventromedial prefrontal cortex of humans. J Neurosci 25:4641–4648

McClelland JL, McNaughton BL, O'Reilly RC (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. Psychol Rev 102:419–457

McDonald S, Lowe W (1998) Modelling functional priming and the associative boost. In: Gernsbacher MA, Derry SD (eds) Proceedings of the 20th annual meeting of the cognitive science society. Lawrence Erlbaum Associates, New Jersey, pp 675–680

McGonigle BO Chalmers M (1977) Are monkeys logical? Nature 267:694–696

McGonigle BO, Chalmers M (1992) Monkeys are rational! Q J Exp Psychol 45B:189–228

McGonigle BO, Chalmers M (1996) The ontology of order. In: Smith L (ed) Critical readings on piaget. Routledge London, chapter 14

Newell A (1990) Unified theories of cognition. Harvard University Press, Cambridge, MA

O'Reilly RC, Rudy JW (2001) Conjunctive representations in learning and memory: principles of cortical and hippocampal function. Psychol Rev 108:311–345

Piaget J (1928) Judgment and reasoning in the child. Routledge and Kegan Paul, London

Piaget J (1954) The construction of reality in the child. Basic Books, New York

Rapp PR, Kansky MT, Eichenbaum H (1996) Learning and memory for hierarchical relationships in the monkey: effects of aging. Behav Neurosci 110:887–897

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Black AH, Prokasy WF (eds) Classical conditioning, vol II. Appleton, New York, chapter 3, pp 64–99

Shultz TR, Vogel A (2004) A connectionist model of the development of transitivity. In: The 26th annual meeting of the cognitive science society (CogSci 2004). Lawrence Erlbaum Associates, Chicago, pp 1243–1248

Siemann M, Delius JD (1993) Implicit deductive reasoning in humans. Naturwissenshaften 80:364–366

Siemann M, Delius JD, Dombrowski D, Daniel S (1996) Value transfer in discriminative conditioning with pigeons. Psychol Record 46:707–728

Steele, GL, Jr (1990) Common Lisp: the language, 2nd edn. Digital Press, Bedford, MA

Terrace HS, McGonigle BO (1994) Memory and representation of serial order by children, monkeys and pigeons. Curr Dir Psychol Sci 3:180–185

Tyrrell T (1993) Computational mechanisms for action selection. PhD thesis, University of Edinburgh. Centre for Cognitive Science

von Fersen L, Wynne CDL, Delius JD, Staddon JER (1991) Transitive inference formation in pigeons. J Exp Psychol: Anim Behav Processes 17:334–341

Van Elzakker M, O'Reilly RC, Rudy JW (2003) Transitivity, flexibility, conjunctive representations, and the hippocampus: an empirical analysis. Hippocampus 13:334–340

Waelti P, Dickinson A, Schultz W (2001) Dopamine reponses comply with basic assumptions of formal learning theory. Nature 412:43–48

Wolpert DH (1996) The lack of a priori distinctions between learning algorithms. Neural Comput 8:1341–1390

Wood MA, Leong JCS, Bryson JJ (2004) ACT-R is *almost* a model of primate task learning: experiments in modelling transitive inference. In: The 26th annual meeting of the cognitive science society (CogSci 2004). Lawrence Erlbaum Associates, Chicago, pp 1470–1475

Wright BC (2001) Reconceptualizing the transitive inference ability: a framework for existing and future research. Dev Rev 21:375–422

Wynne CDL (1998) A minimal model of transitive inference. In: Wynne CDL, Staddon JER (eds) Models of action. Lawrence Erlbaum Associates Mahwah, NJ, pp 269–307

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

Zentall TR, Sherburne LM (1994) Transfer of value from S+ to S− in a simultaneous discrimination. J Exp Psychol: Anim Behav Processes 20:176–183