A Reward Driven Connectionist Model of Cognitive Development

Lorna Peters, Neil Davey
{l.peters, n.davey}@herts.ac.uk
Department of Computer Science

Pam Smith, David Messer
{p.m.smith,d.j.messer}@herts.ac.uk
Department of Psychology

University of Hertfordshire
College Lane
Hatfield AL10 9AB, UK
+44 1707 284000

ABSTRACT
Children learn many skills under self-supervision where exemplars of target responses are not available. Connectionist models which rely on supervised learning are therefore not appropriate for modelling all forms of cognitive development. A task in this class, for which considerable data has been gathered in relationship to Karmiloff-Smith's Model of Representational Redescription (RR) (Karmiloff-Smith, 1973, 1992); is one in which children learn through trial and error to balance objects. Data from these studies have been used to derive a training set and a new approach to modelling cognitive development has been taken in which learning through a dual backpropagation network (Munro, 1987) is reward-driven. Results have shown that the model can successfully learn and simulate aspects of children's behaviour without explicit training information being defined. This approach however is incapable of modelling all levels of the RR Model.

Keywords
Connectionism, cognitive development

INTRODUCTION
Children learn many skills without the environment explicitly providing an example of the target response. One class of such skills is those involving balance of some kind (e.g. learning to walk, ride a bicycle, handle and balance objects). These present a challenge for connectionist modelling and it is one of these skills that we address here.

The task: Behaviour
First investigated by Karmiloff-Smith and Inhelder (Karmiloff-Smith 1973) and subsequently explored in our laboratory and elsewhere, the task presents young children with the challenge of balancing variously weighted wooden beams across a fulcrum (Messer, 1993, 1996; Pine, 1995). For some beams (symmetrical) geometric and gravitational centres coincide; for others (asymmetrical) weight configurations ensure that they do not coincide and that gravitational centres differ within the set.

Observed patterns of behaviour over the age range three to seven have been have been fitted, with considerable success, to Karmiloff-Smith's description of skill acquisition as one of endogenously driven redescription of representations from implicit and encapsulated, to explicitly available, to other operations but not conscious, to conscious, and then, finally, verbalisable. It is the early phases with which we are concerned in this study. Initially very young children engage in unsystematic exploratory behaviour, although when they appear to be trying to balance the beams there is a tendancy to place them on the fulcrum around their geometric centre. This largely unsuccessful behaviour is followed by successful balancing of both symmetrical and asymmetrical beams by children moving each beam (with one or both hands) across the fulcrum until it balances. The inference is that children are responding to proprioceptive feedback (i.e. from receptors in muscles, tendons, joints etc). Asked why a beam balances in that specific position, the chilren cannot explain verbally, although they may sketchily gesture their hand movements. Karmiloff-Smith interprets this behaviour as building implicit, procedural (Implicit Level) representations. It is this stage of chilren's developing skill we address here, although we relate this to the later stages in the discussion.

The task: Modelling
Two aspects of the task present particular challenges to the modeller.

Firstly the environment does not provide a correct position until the child has succeeded in balancing the

beam, but it does provide (proprioceptively) graded information of the 'not near', 'nearer', type.

Secondly the task has two components for the child to learn: s/he has to identify the significance of the proprioceptive feedback and then use that to guide movements. To meet the first issue we decided to use a learning procedure where output (beam placement) would elicit a graded reward value. To meet the second we decided to use a dual network where the reward value associated with a placement would be learned first and that would then be used to teach the desired movement of a beam from an initial position.

THE MODEL

Inspiration for an appropriate learning procedure approach was sought in the area of dynamic systems control, where learning occurs under uncertainty, noise, and without explicit instructional information. The approach taken is based on Munro's dual back-propagation scheme (Munro, 1987) where a single scalar, representing 'goodness of fit' is used as a teaching signal. Since the signal indicates the reward received for a particular choice of action (this can be 0 or 1, or a range of values between) and not what the correct action should have been, this form of teaching is psychologically more plausible for modelling a situation where a child is learning through trial and error.

The model, Figure 1, consists of two networks, a Teacher network and an Action network. Using the generalised delta rule (Rumelhart, 1986), the Teacher network is trained on a variety of beam configurations and positions in relation to the fulcrum. (The configurations were based on actual beams given to children for balancing.) The positioning of each beam across the fulcrum elicits a defined reward response within the range of 0 - 1 so that the network is taught to output high responses when gravitational centres are encountered and, in a graded fashion, lower responses as the gravitational centre is moved away from.

Thus the network is taught to sense proximity to gravitational centres, and so model proprioceptive input from the child's internal environment. The model of proprioception then becomes the teacher for the Action network. The Action network is trained to produce, for each beam presented, an action which represents a fulcrum positioning that will produce the highest level of activation (reward) in the Teacher network's output unit (i.e. a position that will cause the beam to balance).

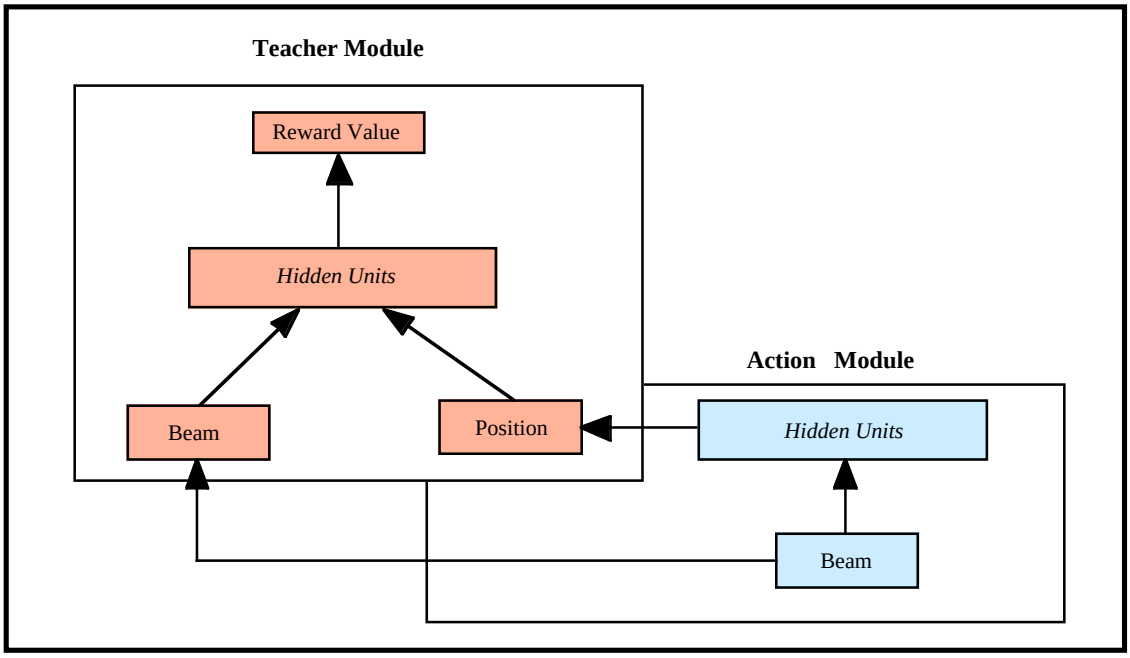

Figure 1: *Schematic diagram of the dual model of beam balancing*

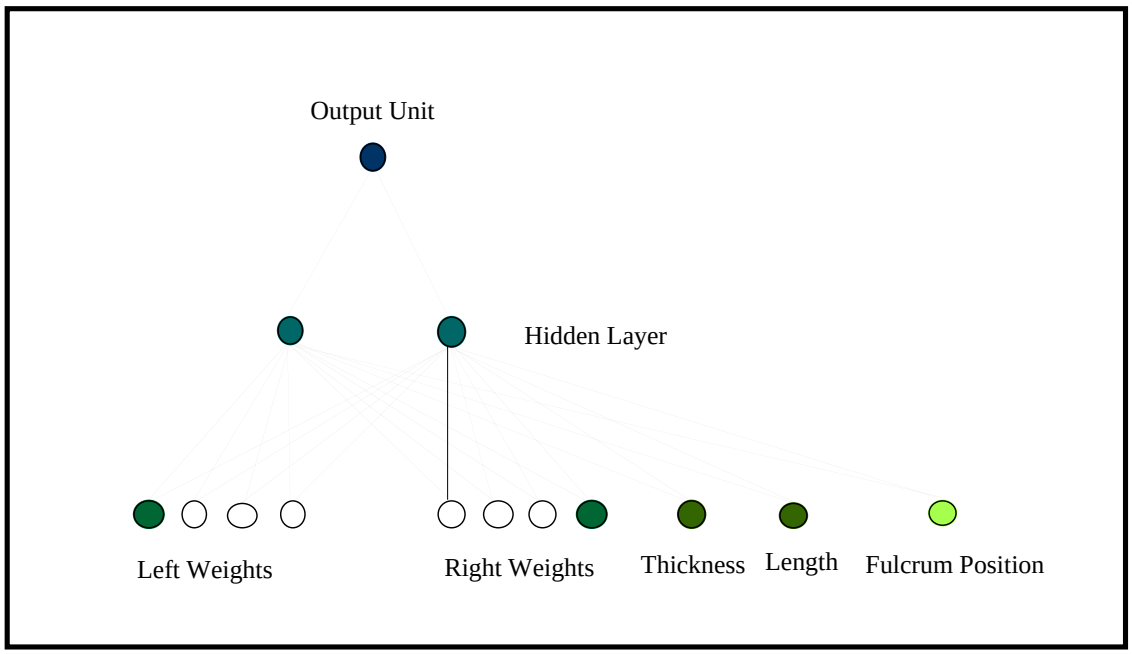

Figure 2: *The Teacher Network*

Training

First the Teacher network is trained using 75 beams in six or seven differing positions, with appropriate reward values. This gives a training set of 470 vectors. After training (using backpropagation of errors) this network successfully generalises over 25 unseen beams, tested at the 49 possible fulcrum points (1225 vectors).

At this point the Action Network is trained by the Teacher network. The training set consisted of the 75 beams in the training set of the Teacher network. Training proceeds as follows:

1. A beam is presented to the inputs of the Action network, and activation is propagated forwards to the outputs of this module, giving a prediction of the beams balancing position.

2. The Teacher Network assesses the quality of the proposed balance position by propagating the position and beam through to the reward output.

3. Gradient ascent is applied to the reward, so that the weights in the Action Network are updated to maximise the reward. The weights in the Teacher Network are never changed in this phase.

4. 1 to 3 are repeated until the overall mean reward is judged sufficient.

RESULTS

During training of the Teacher network, it was found that an early generalisation was that the centre of the beam was the area predicted to achieve highest reward. This is consistent with young children's behaviour where initial beam placements were around the beam's centre (Karmiloff-Smith, 1973; Peters, 1999). With further experience the network developed good generalisation of gravitational centres and this also corresponded well with observed behaviour of some children who developed good initial beam placement without any conscious recognition of their ability (Peters 1999a, 1999b).

The RMS error graph of a typical run is shown in Figure 3. As can be seen the performance over the test set reaches a minimum at about 510 epochs, and this was taken to be a fully trained teacher.

The Action network was subsequently trained from this teacher. The reward signal generated by the complete dual network, during training is shown in Figure 4.

As can be seen learning was fairly rapid, and by epoch 100 average an average reward of 0.6 was being produced.

The dual network was tested by presenting the Action module with novel beam configurations and recording both the action produced by the Action module and the Reward produced in the Teacher module. Testing was repeated at Epoch 30, Epoch 50 and Epoch 100 of Action module training to discover how learning had progressed over the first 100 epochs. After epoch 30, fulcrum positions which are close to the geometric centres of the beams are being generated but some movement towards the positions that the Teacher module represents as gravitational centres, can be observed.

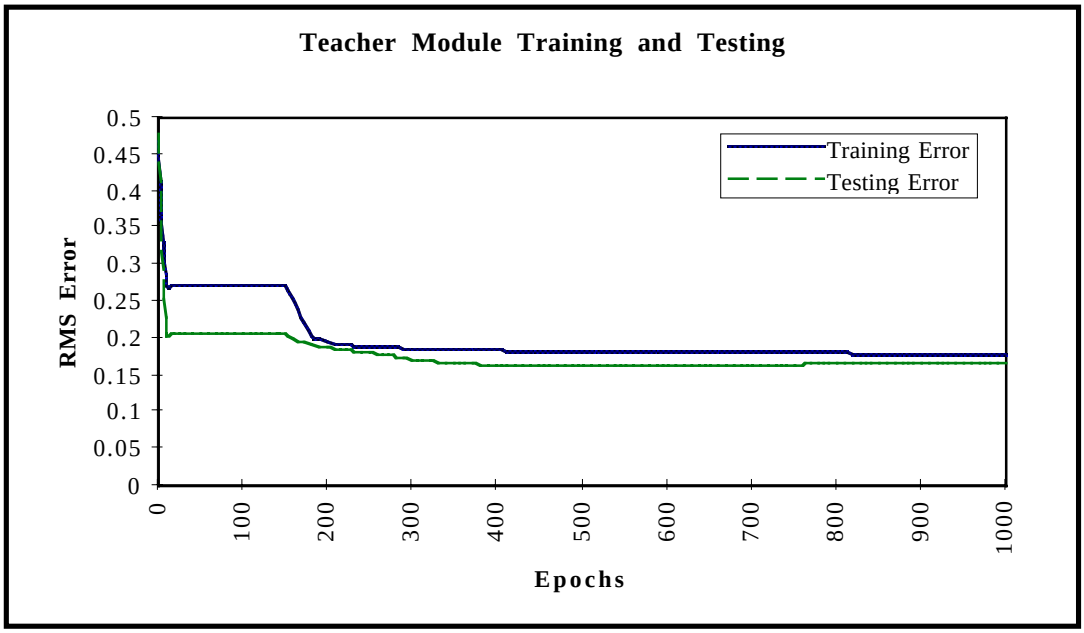

Figure: RMS errors for a typical run as the teacher module is trained

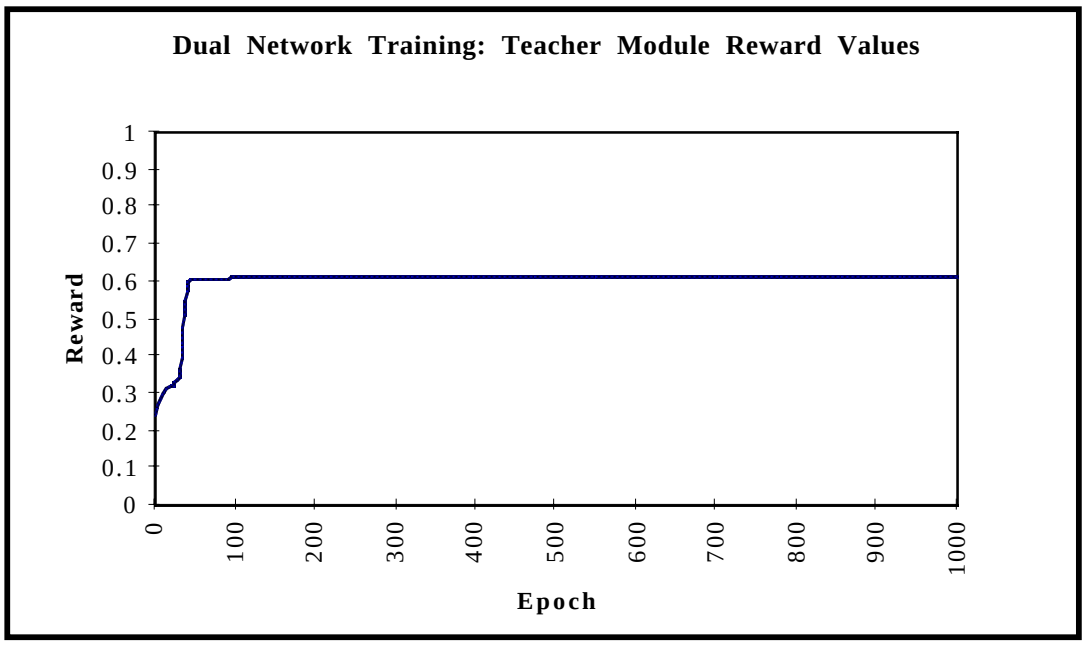

Figure: Reward signal generated by the complete dual network, during training, using 75

After epoch 50, the Action module outputs fulcrum positions which are very close to or exactly at the gravitational positions specified by the Teacher module.

At epoch 100 there is little change in either accuracy of the Action module or the level of reward generated in the Teacher module. Thus training appears to be fairly complete by epoch 50.

DISCUSSION

The simulation presented here models aspects of the children's behaviour in the balance beam task: an early

bias in the Proprioception net mirrors children's early tendency to place beams near the geometric centre, while the trained Proprioception net is effective in guiding learning of the appropriate movement of a beam towards its gravitational centre. We argue that this model provides a plausible associationist account of the early stage of development of balancing objects. The success of this unique approach to cognitive modelling is encouraging. Reward driven models could be applicable to other areas of skill acquisition.

The model is limited, however. A challenge which has not yet been met lies in the development which takes place beyond the Implicit Level. The RR Model briefly theorises about a mechanism which causes the information contained in the Implicit Level representations to be abstracted and new representations to be formed (Explicit Level E1). Behaviourally this level is demonstrated in the balance beam task by children forming a 'centre theory' which leads them to expect all beams to balance at the geometric centre. Their failure with asymmetric beams at this stage frustrate and puzzle them. A definition of this mechanism is yet to be attempted but putting this difficulty aside for the moment, it seems possible that a model such as ours could provide input to an 'abstracting' net where data compression leads to a generalisation about the most frequent single balancing point (the geometric centre) before fine tuning leads to retention of this for symmetric beams and a 'torque' generalisation for asymmetric beams. Further than this however, and connectionism is set to fail. Beyond the first level of abstraction lies consciousness (Level E2/3), an as yet insoluble problem for connectionist modelling techniques. This ultimately sets a limit on the extent to which the RR Model can be researched within a connectionist framework.

Moreover microanalysis of balance beam behaviour (Peters, 1999a) shows individual developmental trajectories which are difficult to reconcile with the postulated RR Levels and more importantly, an interaction between declarative knowledge provided by structured tuition and the developmental progress of children holding representations at all levels. Modelling the development of trial and error learning without explicit correct exemplars was challenging: modelling interaction of trial and error learning intertwined with verbal tuition and self explanation is a real challenge for connectionism.

REFERENCES

Haykin, S. (1994). *Neural Networks.* NY: Macmillan.

Karmiloff-Smith, A., Inhelder, B. (1973). If You Want To Get Ahead Get a Theory. Cognition, 1973, 3, 3: 195-212.

Karmiloff-Smith, A. (1992). Beyond Modularity. Mit Press, London.

Messer, D., Joiner, R., Loveridge, N., Light, P., Littleton, K. (1993). Influences on the effectiveness of peer interaction: children's level of cognitive development and the relative ability of partners. Social Development, 1993 2, 3: 270-294.

Messer, D., Norgate, S., Joiner, R., Littleton, K, Light, P. Development Without Learning?. (1996). Educational Psychology, 1996, 16, 1: 5-19.

Munro, P. (1987). A Dual Back-Propagation Scheme for Scalar Reward Learning. Proceedings of the Ninth Annual Conference of the Cognitive Science Society. Erlbaum, Hillsdale, NJ, 1987.

Peters, L., Messer, D., Smith, P., Davey, N. (1999a). An Investigation into Karmiloff-Smith's RR Model: The Effects of Structured Tuition. British Journal of Developmental Psychology.,

Peters, L. (1999b). Children's Early learning About Object Balancing: Behavioural and Connectionist Studies. Thesis.

Pine, K., Messer, D. (1995). Children's Changing Representations of a Balance beam Task: A Quasi-Longitudinal Study. Paper presented at the British Psychological Society London Conference, December, 1995.

Rumelhart, D. E., McClelland, J. L. (1986). On Learning the Past Tenses of English Verbs. In J. L. McClelland, D. E. Rumelhart and The PDP Research Group (Eds.) Parallel Distributed Processing Vol 2. , MIT Press, London, 1986.