Nonlinear Dynamics, Psychology, and Life Sciences, Vol. 4, No. 1, 2000

Dynamic Models of Simple Judgments: II. Properties of a Self-Organizing PAGAN (Parallel, Adaptive, Generalized Accumulator Network) Model for Multi-Choice Tasks

Douglas Vickers[1,3] **and Michael D. Lee**[2]

This is the second of two papers comparing connectionist and traditional stochastic latency mechanisms with respect to their ability to account for simple judgments. In the first, we reviewed evidence for a self-regulating accumulator module for two- and three-category discrimination. In this paper, we examine established neural network models that have been applied to predicting response time measures, and discuss their representational and adaptational limitations. We go on to describe and evaluate the network implementation of a Parallel Adaptive Generalized Accumulator Network (PAGAN), based on the interconnection of a number of self-regulating, generalized accumulator modules. The enhancement of PAGAN through the incorporation of distributed connectionist representation is briefly discussed.

KEY WORDS: connectionism; stochastic modeling; reaction time; identification; adaptation.

INTRODUCTION

In a previous paper, (Vickers & Lee, 1998), we reviewed evidence for a self-regulating accumulator model for two- and three-category discrimination. We argued that this adaptive stochastic decision process possessed all the essential ingredients of intelligent behavior, and was eminently suited as a basic computing element, or module, in a larger network, that would

[1]Department of Psychology, University of Adelaide.
[2]Department of Psychology, University of Adelaide. Now at: Communications Division, Defence Science and Technology Organisation, PO Box 1500, Salisbury, South Australia, 5108.
[3]Correspondence should be directed to Douglas Vickers, Department of Psychology, University of Adelaide, Adelaide, South Australia, 5005; e-mail: douglas.vickers@adelaide.edu.au.

1090-0578/00/0100-0001$18.00/0 © 2000 Human Sciences Press, Inc.

be capable of carrying out more complex tasks. In this second paper, we begin by examining a range of established neural network models that have been recently applied to the prediction of response time measures in tasks traditionally addressed by stochastic decision models. We argue that none fully utilises the adaptive and representational advantages of the connectionist framework, and go on to describe the neural network implementation of a Parallel Adaptive Generalized Accumulator Network (PAGAN), in which self-regulating generalized accumulator modules are linked into a simple configuration.

Neural Network Models

Neural network models consist of simple representational elements, referred to as 'units' or 'nodes,' which are interlinked by a set of weighted connections. The processing of information is achieved by transferring numerical 'activation values' between these representational elements in a principled manner, mediated by the connection weights. Learning is accomplished through the modification of the connection weights, in accordance with one or more learning rules. Useful treatments of neural network models, adopting several complementary perspectives, are presented by Anderson (1995), Arbib (1995), Haykin (1994), Hertz, Krogh, and Palmer (1991), and Rumelhart (1989).

The attraction of the neural network framework with regard to cognitive modeling is at least four-fold. First, computation within this framework readily implements the 'bounded' or 'constrained' optimization processes which characterize much of human cognitive behavior (e.g., Anderson, 1990; Gigerenzer & Goldstein, 1996; Simon, 1982), as evidenced by the recent spate of neural networks which find reasonable solutions to computationally intractable combinatorial optimization problems such as the Traveling Salesman Problem (for overviews, see Peterson & Söderberg, 1995; Yuille, 1995). Secondly, neural network modeling, through the appropriate assignment of 'sensor' and 'effector' units, can naturally accommodate the view that human cognition can be realized only within an embodied and situated agent (Brooks 1991a, 1991b; Norman 1993; cf. Glenberg, 1997). Thirdly, neural network models are suited to the realization of the fundamental notion that mental representations are active and emergent phenomena, rather than passive data structures overseen by some rule-driven executive (Cussins, 1990; Hofstadter 1985, ch. 26; Smolensky, 1988). Finally, there are a large number of established learning rules, including unsupervised and self-supervised rules, based on principles of self-organization and regulation, and supervised rules, which require some form of external instruction or feedback. Hanson and Burr (1990, Fig. 1) present a useful taxonomy of various neural network learning rules.

Establishing a Temporal Metric

To provide a model of reaction time phenomena, a neural network model must be subject to some form of temporal metric. Unfortunately, cognitive neural network models have predominantly focused on response accuracy, rather than response time, as a measure of performance. This has led to a situation in which large classes of neural networks cannot readily be identified with any such metric. In particular, it is difficult to derive sensible measures of processing time in neural networks with feed-forward architectures. Within these models, information provided at an input layer propagates to generate activation values at all subsequent layers. Since the value of the inputs, coupled with the current state of the network's connection weights and other parameters, fully define activation value across all units in the network, this propagation of information is not measurable in terms of time.

Measures of information processing time are, however, readily derived from neural network models in which the propagation of activation values through the units has some component of recurrence. By this, we mean that there is at least one unit which has architectural interconnections to other units such that the current activation value of this unit will, at some stage during the network's evolution, affect its own subsequent activation value. Therefore, unlike feed-forward networks, the propagation of activation information in recurrent networks does not terminate. While the pattern of activation values across the units of a recurrent network may stabilize, the flow of information underpinning the activation values does not cease. Thus, if a neural network model is recurrent, it is natural to measure the processing time since an input was presented in terms of the number of iterative updates of activation values that have occurred.

Fully Connected Architecture

In surveying previous neural network models of response times, it is helpful to distinguish between two broad classes of architectural recurrence—full recurrence and layered recurrence. As shown in Fig. 1, fully recurrent networks possess a connection from every unit to every other unit, and can profitably be viewed as non-linear dynamical systems which evolve through the iterative updating of the patterns of activation values

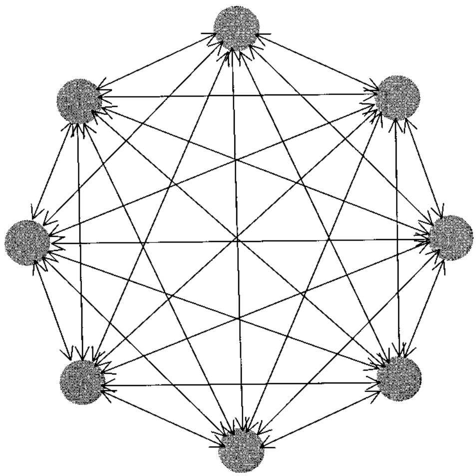

Fig. 1. An example of a fully recurrent neural network architecture.

across the units. Within these networks, it is possible either to pre-determine (typically using the outer-product form of the Hebbian learning rule), or learn (typically through applying gradient descent optimization principles to an error measure) sets of connection weights which implant attractors in the state space of this dynamic system. Furthermore, as noted in McClelland's (1991) discussion of neural network response time modeling, it is also often possible with fully recurrent networks to define a global 'energy' function which can be shown to be (locally) minimized by the network's evolution.

Neural networks with this architecture, therefore, can be applied to the modeling of response times by defining correspondences between patterns of activation values across the units and various responses, and then developing a set of connection weights within the model, either pre-determined or learned, which create point attractors at the system states identified with the responses. Accordingly, the generation of an overt response is triggered whenever the network's state is sufficiently close to that of a response state.

Anderson's (1991, see also Anderson, 1995, ch. 15) application of the 'Brain-State-in-a-Box' (BSB) neural network to the modeling of response times is one example of this approach. Anderson (1991) examines the BSB model's performance on a 'same-different' discrimination task across a set of stimulus vectors with a range of different pair-wise similarities. Primarily, the evaluation of the simulation results takes the form of a qualitative discussion of various features of the response time distributions across different levels of response accuracy and stimulus similarity. Ratcliff and van Zandt's (1996) application of a BSB network model to a signal detection task involves only a minor reinterpretation of the stimulus vectors used by Anderson (1991), and incorporates an evaluation through comparison with empirical data. In essence, the conclusion of Ratcliff and van Zandt (1996) is that "the BSB model does an extremely good job of predicting the experimental data, except for aspects of the data involving error reaction times and individual differences in sequence effects" (p. 31). It is arguable that the second of these deficiences constitutes a serious challenge to almost all reaction time modeling. However, after attempts to redress the first deficiency, Ratcliff and van Zandt (1996) conclude that "we have found no reasonable way to get the BSB model to produce a few extreme errors with fast reaction times" (p. 31), and argue that this inability may well be a fundamental property of the information processing dynamics of BSB models.

Ruppin and Yeshurun (1991) and Chappell and Humphreys (1994) develop other cognitive models, based on fully recurrent networks, which are primarily concerned with atemporal cognitive phenomena, but do consider response time performance. Ruppin and Yeshurun (1991) employ simulated mean response times to evaluate their model in relation to serial position and memory retrieval effects, while Chappel and Humphreys (1994) demonstrate that the latency distributions of their model qualitatively match those empirically found and, more generally, emphasize the advantages of using sparsely distributed representations (cf. Kanerva, 1988; Keeler, 1988).

Layered Architecture

A second form of neural network recurrence involves units being grouped into distinct 'lateral inhibition layers,' which replace the units themselves as the architectural 'atoms' of the network. In a lateral inhibition layer (see Grossberg 1976, 1980) each unit has an inhibitory connection to every other unit, whilst possessing an excitatory connection to itself. Basically, therefore, units in lateral inhibition layers 'compete' for activation values in a manner which is sensitive both to the received input, and the contextual configuration of various alternatives represented by the units.

The lateral inhibition-based 'interactive activation architecture,' shown in Fig. 2, underpins one of the earliest and most impressive cognitive models which considers response time issues, namely, the letter-perception model described by McClelland and Rumelhart (1981; see also Rumelhart &

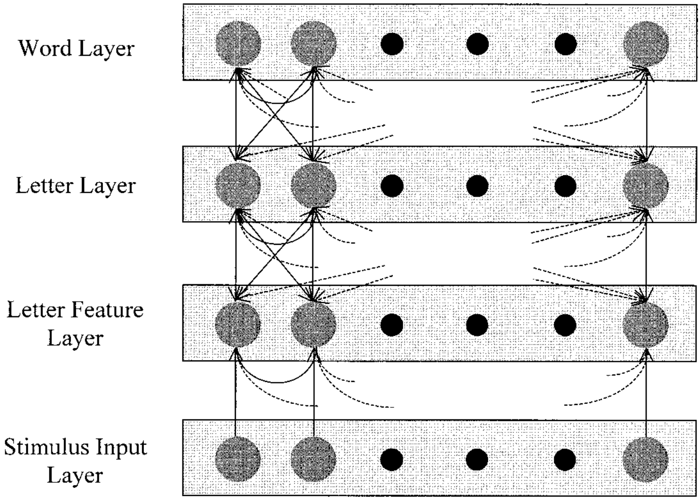

Fig. 2. An example of an interactive activation architecture.

McClelland, 1982). Within this model, units in the lower-most layer correspond to simple geometric features, such as the crossbar of a letter 't,' units in the next layer correspond to the letters themselves, and units in the uppermost layer correspond to words composed of these letters. Although McClelland & Rumelhart (1981) primarily employ response probabilities as measures of model performance, the time course of the generation of these response probabilities is also seen as an important feature of this evaluation. For example, the response time for the model to recognise a letter as part of a four character word is demonstrated to be faster than when the letter is presented alone.

The dominant contemporary layered architectural paradigm, known as GRAIN (McClelland, 1991), is a direct descendant of the interactive activation model. The primary difference between the two approaches is that the mutual interconnections established between layers of GRAIN networks are constrained to be purely excitatory. The motivation for the exclusion of inhibitory connections between layers is discussed by Usher and McClelland (1995), and concerns problems inherent in the possibility of units within one layer 'over-inhibiting' a unit corresponding to a plausible alternative at a subsequent layer. Instead, the necessary comparison of plausible units within a given GRAIN layer is accomplished solely by lateral inhibition.

The GRAIN class of models also differ from both McClelland and Rumelhart's (1981) letter-perception model and Anderson's (1991) BSB-based model in their incorporation of intrinsic processing variability. Anderson (1991) suggests that the determinism of the BSB model may be appropriate to the extent that experimental factors, such as the sequence of the stimuli presented, are responsible for the variability of response times in relation to the same stimulus. While acknowledging this claim, McClelland (1991) introduces processing variability on the grounds that "there are surely many other factors (mood, context, motor preparation, accommodation of the eye, etc.) that introduce trial-to-trial variation" (p. 660). In any case, as noted by Usher and McClelland (1995), and pursued by Ratcliff and van Zandt (1996), there is considerable scope for the extension of BSB-type models to incorporate processing variability.

Ratcliff and van Zandt (1996) also employ their signal detection task to evaluate a GRAIN-based model. Their particular model contains three layers and incorporates adaptive learning capabilities through learning rules based on mean-field theory, the discrete counterpart of the Boltzmann learning algorithm (see Hertz, Krogh, & Palmer, 1991; Peterson & Hartman, 1989). Ratcliff and van Zandt (1996) summarize their findings by noting that the model "captures some features of the data" (p. 24), but also has limitations. These include: the mistaken prediction that "error responses [are] almost always slower than correct responses" (p. 24); an inappropriate prediction of an excessive number of lengthy response times; and an inability to capture sequential effects because of the fundamental assumption that learning is dependent on external feedback.

Usher and McClelland (1995) explore the response time modeling capabilities of a simpler GRAIN-based model which incorporates only an input and a response layer. Inter-layer excitatory connections are established only from the input to the response layer, and there are only two units in the lateral inhibition response layer. These two units are aligned with the potential responses associated with a two-choice task, although Usher and McClelland (1995) note the ease with which an extension can be made to a multi-choice paradigm through the addition of further response units. The GRAIN assumption that the representation of an input varies stochastically is implemented through the addition of Gaussian noise to the activation values of the response units received from the input units. Unlike the GRAIN-based model examined by Ratcliff and van Zandt (1996), Usher and McClelland's (1995) model does not incorporate any form of learning.

Usher and McClelland (1995) consider data from time-controlled signal detection tasks, and show that their model predicts that, at short times, or for easy tasks, d' will increase as the square root of processing time, as predicted by most fixed sample models. At longer times, their model produces d' curves that converge exponentially to asymptotic behavior, as observed by Wickelgren (1977). They go on to examine the time-controlled speed-accuracy tradeoff in some detail, following Loftus, Busey, and Senders (1993) in assuming that the probability of a correct response, corrected for guessing, is given by $P_c = 2P - 1$, where P is the probability of a correct response and P_c is the corrected estimate. However, it may be noted, in passing, that this correction for guessing makes assumptions about the relative proportions of 'catastrophic' and 'process-induced' errors, which may not be justified (Pietsch & Vickers, 1997). These assumptions, in turn, are related to questionable assumptions about the representation of incoming stimulus elements (e.g., that there is no effective capacity limitation).

Usher and McClelland (1995) also evaluate the behavior of their model in information-controlled, 'standard' choice reaction time tasks. A limiting case of their model, in which there is neither leakage nor lateral inhibition, is equivalent to a continuous-time version of an accumulator process. They compared the distribution of reaction times obtained for their network of leaky, competing, laterally-inhibited accumulators with data from an experiment by Vickers, Caudrey and Willson (1971), in which observers were required to judge the relative frequency of sequences of light flashes on the left- or right-hand of two lamps. They found a good qualitative fit to the pattern of latency-probability functions, obtained by plotting the mean reaction time for a response against the probability of making that response for different relative frequencies of the two stimulus alternatives. The shapes of empirical response time distributions for correct and incorrect responses were also well described.

Finally, Lacouture and Marley (1991; see also Lacouture and Marley 1995) use a three-layered variation on the well-established 'auto-encoder' (Ackley, Hinton & Sejnowski, 1985) connectionist network to model absolute identification and choice reaction time tasks. The architectural recurrence of their network resides in the 'decision module,' which consists of a layer of integrator units which accumulate activation values through self-connections, and fire once they reach a set threshold. This decision layer may be thought of as a lateral inhibition layer without inhibitory connections, and with self-excitatory connections greater than unity which function as 'gain' parameters. Lacouture and Marley (1991) focus upon modeling the identification of unidimensional stimuli, and argue on both theoretical and empirical grounds that this is best achieved by using a hidden layer with only one unit. Stimuli are represented by Gaussian distributions, with appropriate activation values being coarsely distributed across the units in the input layer. Learning is accomplished using gradient-descent principles applied to an error measure which appends to the standard mean-squared-error term an additional 'penalty' or 'regularizing' term measuring the variance of the squared error. Lacouture and Marley (1991) argue that this variance term corresponds to a form of 'selective attention' which leads the network to attempt to learn to identify each stimulus in a set with equal accuracy.

Lacouture and Marley (1991) demonstrate their model's ability to capture classic set-size effects (e.g., Merkel 1885), effects on changes in d', and range effects. They also suggest plausible means by which established features of reaction time distributions (Luce 1986) and differences in the latencies of correct and incorrrect responses might be accounted for using the same basic network architecture.

EVALUATION OF NEURAL NETWORK APPROACHES

Modeling Possibilities

Inevitably, the fact that applying the neural network framework to response time modeling is a relatively recent development means that a number of promising avenues are currently unexplored. There are a large number of established recurrent network architectures which have not been directly applied to the modeling of response time phenomena. One obvious candidate is the Adaptive Resonance Theory (ART) (see Grossberg 1982, 1987a, 1987b) family of models, particularly ART-EMAP models (Carpenter & Ross, 1995) which include evidence accumulation mechanisms reminiscent of those employed in the accumulator module we described in part I. The class generically known as 'sequential network architectures' (see Hertz, Krogh & Palmer, 1991) also afford an architectural flexibility which might prove useful in attempting to provide natural psychological interpretations for various components of the model. Finally, bi-directional associative memory networks (Kosko 1988) constitute a compromise between fully-recurrent and layered architectures which seems worth exploring to gain further insight into the relative merits of both architectural approaches. Beyond these possibilities, there remain countless network configurations which might sensibly and profitably be applied to the modeling of response time phenomena.

Similarly, a number of established cognitive models, founded upon recurrent neural networks, remain largely unexplored in relation to response time issues. A typical example is provided by the connectionist implementation of schematic memory structures developed by Rumelhart, Smolensky, McClelland and Hinton (1986). In this model, a set of units corresponding to household furniture and room properties, such as 'small,' 'cupboard,' and 'table,' is placed in a fully connected architecture. By considering the network's stable state when the activation values of some units are fixed at certain values, Rumelhart et al. (1986) argue that room schemata such as 'kitchen,' 'dining room,' and 'bathroom' are naturally emergent properties of the model. Although response times are not used to evaluate the model, the fully recurrent nature of the network means that response time measures, such as the number of processing iterations needed to include a 'toaster' in a 'kitchen,' are readily available. It is possible that a consideration of such measures would provide useful additional perspectives on the neural network modeling of response times.

Current Shortcomings

All of these neural network models (or potential models) of response times hold promise, and it is perhaps too early to make definitive judgments regarding their various merits. However, there are a number of aspects of these models in which shortcomings suggest themselves, and there are also a number of aspects in which clear deficiencies are evident.

First, perhaps only one of the models which focus primarily on response times employs principled distributed representations. While several of the fully-recurrent iterate-to-attractor type of models employ distributed representations of a sort, it is difficult to provide an objective and psychologically well-founded means of generating appropriate stimulus representations for a given experimental task. In Ruppin and Yeshurun's (1991) modeling, for example, the only representational guidance is that the similarities between various distributed stimulus patterns (as measured by a Hamming metric) "reflects" (p. 384) the similarity of the stimuli. Anderson's (1991) generation of representations is guided in much the same way. While the method by which these distributed stimulus representations are generated remains largely unconstrained, it seems pertinent to recall Smolensky's (1987, cited in Smolensky 1988) general warning that "a poor representation will often doom the model to failure, and an excessively generous representation may essentially solve the problem in advance" (p. 69).

Of the layered connections model, those which are GRAIN-based use entirely local stimulus representations, although this practice is regarded by McClelland (1991) as a feature that "arises not as a matter of principle but as a simplification" (p. 656), and Usher and McClelland (1995) make a similar concession. The stimulus representations used by Lacouture and Marley (1991), however, given their close correspondence to those developed by Shepard and Kannappan (1991) from a first-principles theory of mental representation (Shepard 1987), seem more promising. Furthermore, the extension of their representational approach to multidimensional stimuli has been anticipated by Shepard and Tenenbaum (1991), although Lacouture and Marley (1991) suggest that developing an appropriate configuration for the hidden layer within their architecture for the multidimensional case is a non-trivial exercise.

Secondly, the ability of the response time neural network models to learn is often either inappropriate or remains to be implemented. The predominance of pre-learning, evident in the derivation of the models of Anderson (1991), Ruppin and Yeshrun (1991), and Chappell and Humphreys (1994), effectively precludes consideration of the learning which occurs during an actual experiment. Usher and McClelland (1995) discuss ways in which this type of learning might be incorporated in their model, including, in particular, a self-organizing learning rule. However, these features are not incorporated in the modeling they report. Even those models which address response time issues less specifically, such as the letter-perception (McClelland & Rumelhart, 1981) model, have connection weights which are entirely pre-determined. One which does incorporate a significant learning component is the GRAIN-based model evaluated by Ratcliff and van Zandt (1996). Unfortunately, however, one of the most decisive conclusions made by Ratcliff and van Zandt (1996), following their empirical evaluation of this model, is that the feedback-driven assumption on which the learning rules are founded is inappropriate. Once again, the modeling of Lacouture and Marley (1991) appears more promising, although the 'selective attention' justification for their learning rule, while plausible, is somewhat qualitative and heuristic. The more principled learning rules they suggest, based largely on notions of information preservation, remain to be fully implemented and evaluated.

Ironically, these two criticisms, involving issues of representation and adaptation, are directly aimed at areas which were identified earlier as potential strengths of the neural network modeling approach. A third general weakness relates to the overwhelming explanatory burden borne by the nature of the architectural recurrence in many models. The means by which overt decisions are triggered in all of the models we have described are relatively simple. Proximity to a response state in a fully recurrent network, or the attainment of a threshold activation value in a response unit of a layered network, are typically the sole mechanisms for converting underlying information processing into response behavior. Effectively, therefore, the ability of these models to accommodate response time phenomena is the result of the rich interconnectedness of the network architectures which perform the bulk of the information processing.

From this perspective, Anderson's (1991) practice, also canvassed by Ruppin and Yeshurun (1991), of not establishing every connection in a fully recurrent architecture, suggests that the density of unit interconnection consequently required by recurrent architectures may be inappropriate.

Usher and McClelland (1995) provide a biological rationale for the GRAIN network architectures they advance, but we are inclined to concur with Ratcliff and van Zandt's (1996) view that "until the models become truly models of neural functioning, "neural plausibility" should be attended to but not be decisive in decisions about processing" (appendix 1). Our reservation concerning the massive interconnectedness of some recurrent architectures, particularly fully recurrent architectures and those based on lateral inhibition layers, is based more on issues of model interpretability and computational tractability. In particular, it seems likely that sophisticated decision rules could lessen the computational burden on the network architecture, thus reducing the required density of architectural interconnections and facilitating the intuitive identification of meaningful components.

PAGAN

With these weaknesses of previously suggested neural network models of response times in mind, a neural network implementation of the adaptive accumulator module, described by Vickers and Lee (1998), would appear to hold some promise. Fortunately, it is generally easy to recast traditional stochastic latency models, such as the random walk, diffusion and accumulator models, as neural network models, by employing the types of network structure described by Lippman (1987). In particular, self-organizing learning rules would naturally implement the self-adjustment of standards in the model. In a similar fashion, the confidence-based threshold adjustment rule is a form of self-supervised learning, where the 'teacher' values are derived internally from the over- and under-confidence accumulators of each module. Quite possibly, other neural network learning techniques could be applied to extend the model. However, the empirical necessity for such extensions would need to be established beforehand.

Structure and Processing of PAGAN

Figure 3 shows a representation of a module within the proposed PAGAN (Parallel Adaptive Generalized Accumulator Network) model in information-flow terms. This corresponds to the self-regulating accumulator module for three-category judgments described by Vickers and Lee (1998), except that this version is configured to make only judgments of identity, or equality ($v = s$), between a variable stimulus, v, and a standard, s. As a result, the evidence accrual processes leading to the response outcomes

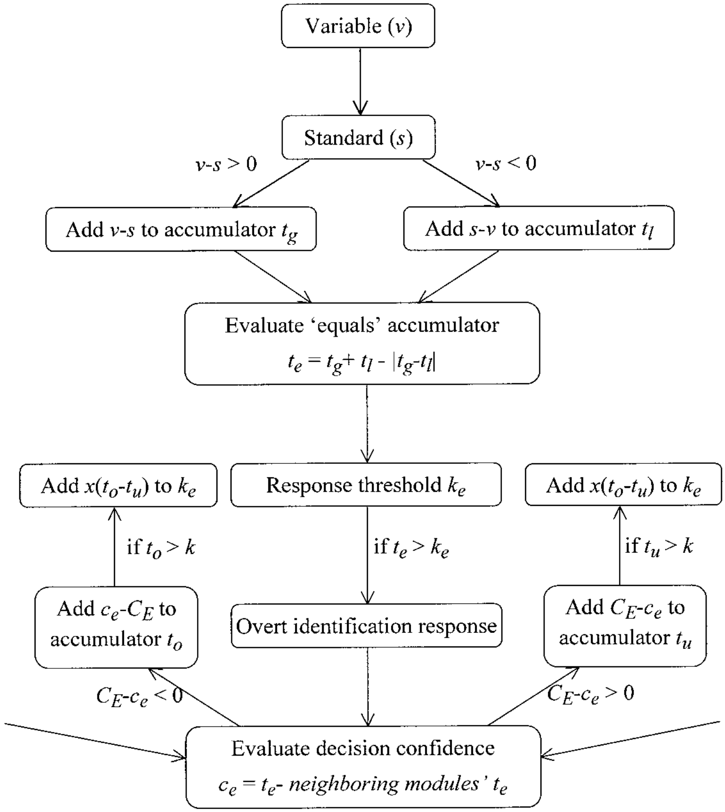

Fig. 3. A PAGAN identification module.

$v > s$ and $v < s$ are assumed to be inoperative and are therefore not shown. In other respects, however, the operation of this identification module is similar to that of the original three-category mechanism. On a given trial, successive values of the representation of a presented variable, v, are compared with those of a remembered standard, s, and the momentary positive and negative $(v - s)$ differences are separately accumulated until the evidence, t_e, for an identification response reaches or exceeds a threshold value, k_e. That is, unlike the situations considered by Vickers and Lee (1998), where both variable and standard are assumed to be externally presented, this application deals with tasks in which a variable is presented, and has to be identified as equal to an internally stored standard (or, more usually, to one of a number of standards). The latter situation is exemplified by choice reaction tasks, in which subjects must identify which one of a (usually linear) array of stimuli has been presented (Welford, 1980), and by absolute identification experiments, in which the linear ordering of the possible standards, such as a set of line lengths, is more explicit (Lacouture & Marley, 1991; 1995).

A particular feature of this identification module is that the response threshold is expressed in terms of the sum of the totals of the unsigned $(v - s)$ differences minus their modulus. This means that evidence is directly accrued in favour of an identification response, rather than this being a default conclusion from a failure to find some difference. On each trial, the confidence, c_e, in the identification response that eventuates is compared with a target level of confidence, C_E, and the amount of over- or under-confidence is accrued in the respective secondary, or control accumulators (over-confidence, t_o, in one accumulator and underconfidence, t_u, in the other). As soon as one of these totals reaches a predetermined threshold, K, this triggers an internal adjustment. If a critical amount of over-confidence has been accrued, then the threshold, k_e, in the primary accumulator (which is responsible for the identification response), is reduced. Conversely, if a critical amount of under-confidence is accrued first, then the primary response threshold is increased. The amount by which the threshold is increased or decreased is proportional to the difference between the amounts of over- and under-confidence accrued at the time that one of these totals reaches the critical amount, and an internal adjustment is triggered. (In other words, the internal adjustments are themselves proportional to the 'confidence' that an adjustment is appropriate).

Meanwhile, the coefficient of proportionality, x, serves to determine the coarseness of control exerted by the secondary accumulators. It does this in conjunction with the thresholds assumed for these accumulators. For the present, these are simply assigned a uniform, moderate value. (However, in principle, any parameter of any accumulator process could be altered by any other accumulator process.) Similarly, the coefficient of proportionality is also assigned a uniform value ($x = 0.75$), which is intermediate between what produces minimal, trial-to-trial adjustments in the primary thresholds (0.25) and a value that gives rise to intermittent but more dramatic changes (e.g., 2). As shown by Vickers (1979, p. 211), the coarser the control exercised, the lower the dicriminative efficiency of the system.

In the identification module shown in Fig. 3, there is only one overt response. In order to apply the model to choice reaction and absolute identification tasks with multiple response alternatives, all that is required is to assign one such module to each of the representations, or remembered standards, corresponding to each response alternative. For example, Fig. 4 shows an array of four adaptive identification modules, appropriate for

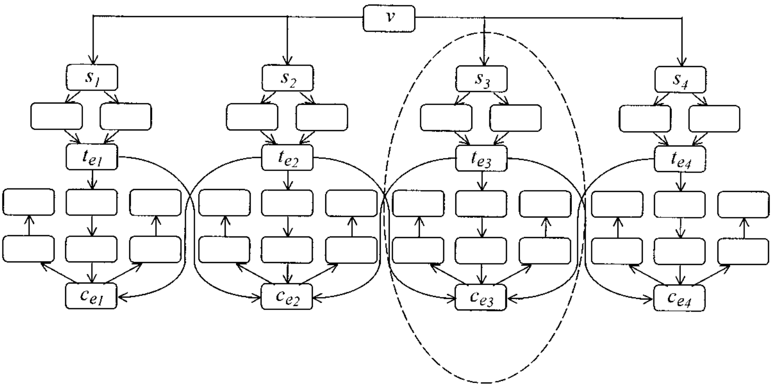

Fig. 4. A PAGAN array of four adaptive identification modules.

a 4-alternative choice reaction or absolute identification task. On any given trial, the presented stimulus is compared simultaneously by each module with the standard assigned to that module. The process resembles a horse-race, in which the eventual overt response is determined by the first module to reach its threshold. When this happens, the evidence in favor of that particular identification response is then compared with the evidence accrued in favor of the nearest single response (in the case of end responses, at the extremes of a linear array), or with the average of that accrued in favor of the two nearest neighboring responses (in the case of responses in intermediate positions), to give a measure of the confidence with which that decision is reached. Other conventions for assessing confidence are also conceivable. However, this approach was chosen because it is arithmetically simple and involves the same kind of local comparison as in the original three-category decision module. It also plausibly represents the way in which we might expect an observer to operate. (For example, if an observer identifies a variable as being equal to the smallest standard, we might expect them to consider the second-smallest standard as a possible alternative response. However, it seems less likely that the observer would consider the largest standard as a possible alternative, to be used in assessing the confidence with which the actual identification response was made.)

Each module in this array preserves the nonlinear characteristics and dynamical properties of the original three-category module. These include: the assumption of a fixed rate at which sensory input is sampled, which provides a temporal metric for the operation of the model; a nonlinear relation between stimulus difference and the probability of a particular, overt response (and between theoretical confidence measures, averaged over trials, and the probability of an internal threshold adjustment); hysteresis, or a differential response to steady increases or decreases over trials in the value of a presented variable; and the capacity to adapt, despite constant parameter values, to any change in the range or distribution of presented values of the variable stimulus or in the prior probabilities of any of the response alternatives. At the same time, the response of each individual module depends upon the existence and behavior of all other modules in the array. Conversely, the performance of the array, as a whole, is not deducible from the sum of the histories of the individual modules considered in isolation. Thus, the PAGAN model inherits the nonlinear, dynamical properties of its component modules, while acquiring a new set of properties, which are not intuitively obvious from its composition. A small selection of its properties are described below.

Demonstrations of PAGAN

Figure 5 illustrates the evolution of the totals in the "equals" accumulators over successive iterations (time). In this case module 4 is the first to reach a (current) threshold of 15. As in the simple, three-category case, it may be assumed that target levels of confidence are set by the observer so

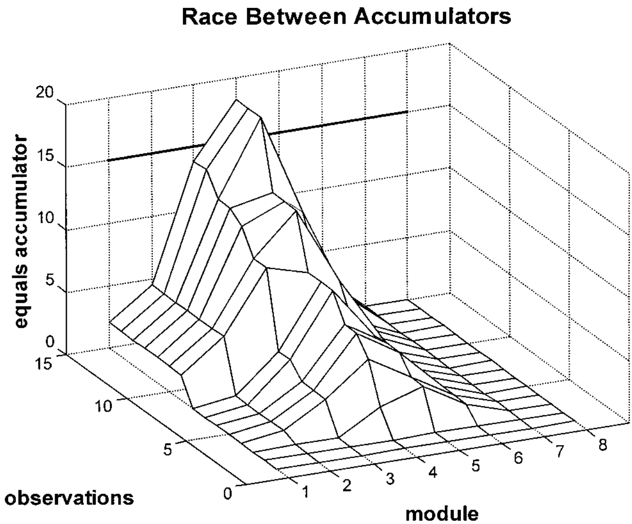

Fig. 5. "Equals" accumulator totals for 8 PAGAN modules during an identification trial.

as to make certain classes of response with, on average, a certain degree of confidence. Once these levels are set, the system will automatically configure the thresholds of the individual modules so as to adapt to any change in the pattern of stimuli presented to it. The list of potential experimental manipulations is too long to consider here. However, it is useful to look at a number of situations where some comparison with empirical data can be made.

Effects of Set Size

Figure 6 shows how the mean number of observations required by the PAGAN model to make a correct identification response increases as a (slightly S-shaped) function of (the logarithm of) the number, *n*, of alternative stimuli (and standards). The form of this increase is closely similar to that of the classical data of Hick (1952) and Merkel (1885), as plotted by Laming (1966). Meanwhile, Hale's (1968, 1969) finding that the times for errors in multi-choice experiments were shorter than those for correct

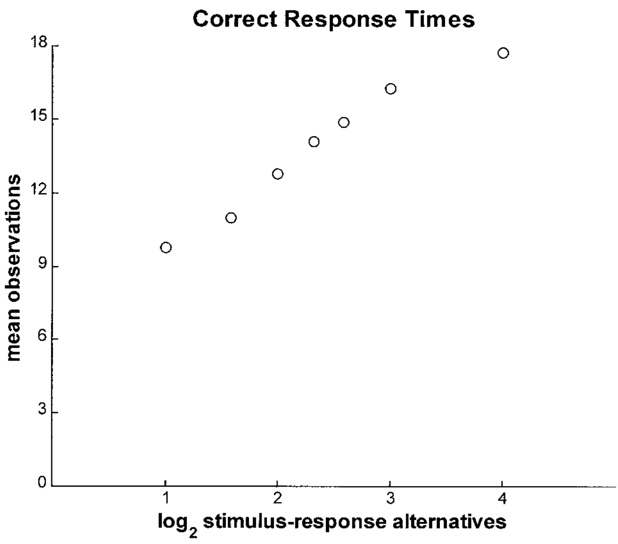

Fig. 6. Mean response times, produced by PAGAN, vs. log (base 2) of number of stimulus response alternatives for 2, 3, 4, 5, 6, 8 and 16 choice identification.

responses, irrespective of whether instructions emphasized speed or accuracy, is also echoed by the present model, as shown in Fig. 7.

The relation between predicted response time and the number of response alternatives, shown in Figs. 6 and 7, is a consequence of the assumption that the standards corresponding to each response alternative must be accommodated within a finite representational space. As the number of response alternatives is increased, the represented standards become crowded together and are less discriminable. Increasing the number of response alternatives should therefore lead to less accurate responding, as predicted by the model and found empirically. However, the accompanying predicted and observed increase in response time is also an important consequence of the self-regulation of thresholds of the individual modules. As shown by Vickers (1979, p. 258), in the absence of such adaptation, a simple horse-race model would predict that increasing the number of response alternatives should lead to faster responding, particularly for the less discriminable alternatives, such as those in the middle of a linear array. The reason is that rival alternatives, which are linked to standards that are quite different from the standard associated with the 'correct' module, provide only poor competition for that module. However, as more standards are added, and are packed more closely together, competition for the module in question is intensified. When the number of response alternatives

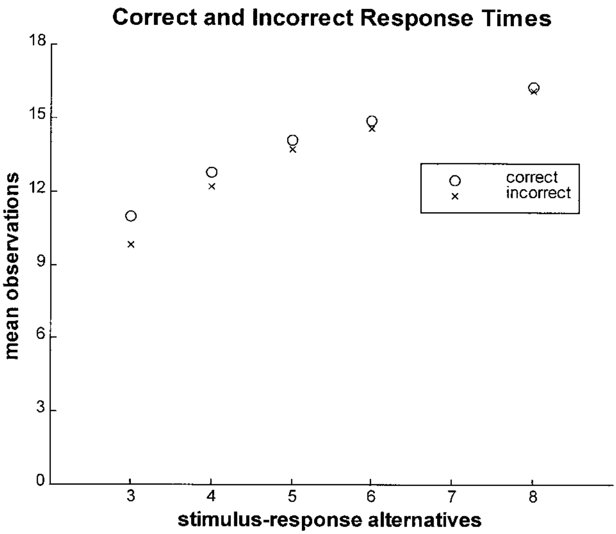

Fig. 7. Mean correct and incorrect response times, produced by PAGAN, for 3, 4, 5, 6, and 8 choice identification.

is increased, therefore, we should expect the correct module to win less often, but in a faster time on average. In contrast, in the adaptive version, when input to a module becomes less discriminable, the threshold for that module is automatically increased. This results in longer response times and serves to counteract, though not to completely cancel, the accompanying increase in the number of errors.

Serial Position Effects

A second set of striking qualitative correspondences between theory and data concerns the predicted and empirical patterns of serial position effects obtained when accuracy and response times are considered separately for each of a linear array of different stimuli, each presented for identification in terms of one of a similar array of standards. Figures 8(a)–(c) illustrates the performance of the PAGAN model in a range of multi-choice tasks. In general, end stimuli are associated with faster and more accurate responses, particularly when the degree of choice is low (e.g., 3 or 4). However, the interplay between stimulus discriminability and serial position, on the one hand, and the adaptive adjustments in threshold levels, on the other, means that, with higher degrees of choice (e.g., 8), the middle stimuli can also show an advantage with respect to both speed and accuracy. Figure 9 shows the asymptotic mean criterion levels reached by the system after 5,000 trials, and part of their evolution towards this asymptote. A pattern of empirical results, qualitatively similar to the early stages of this evolution, has been reported by Welford (1971, 1973), Nettelbeck and

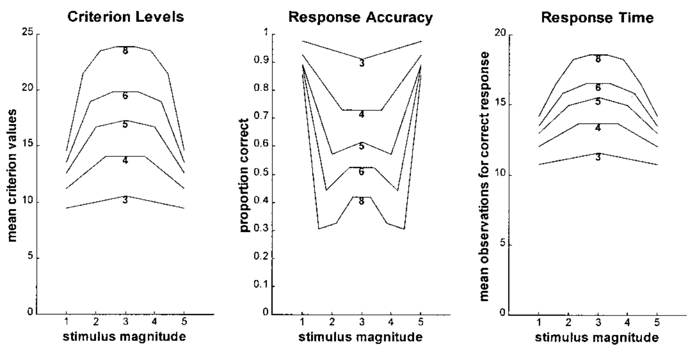

Fig. 8. Performance of PAGAN across 5,000 trials for 3, 4, 5, 6, and 8 choice identification, in terms of (a) criterion levels, (b) response accuracy, and (c) response time.

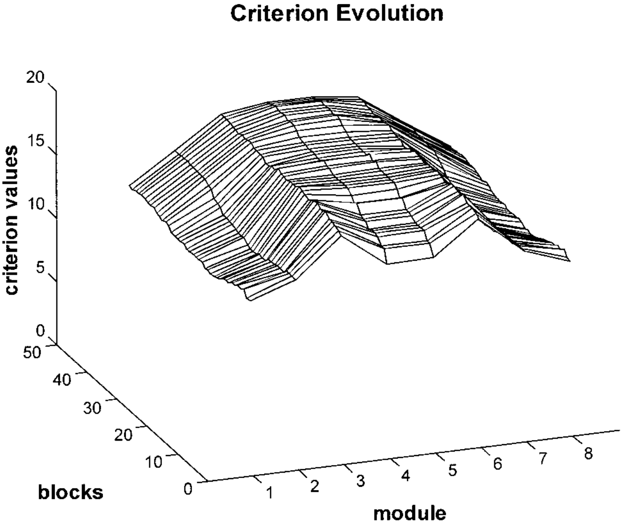

Fig. 9. The evolution of criterion levels in PAGAN over 5,000 trials for an 8 choice task.

Brewer (1976) and Smith (1977). A convenient summary of similar end and bow effects in serial list learning and symbolic comparison studies is given by Leth-Steenson (1998).

A number of alternative explanations of such effects have been advanced. These include an *end anchor* strategy (Trabasso & Riley, 1975; Trabasso, Riley, & Wilson, 1975), an *end-inward scanning* process (Woocher, Glass, & Holyoak, 1978), and a *response bias* explanation (Birnbaum & Jou, 1990). However, such explanations are inevitably ad hoc. In contrast, in the PAGAN model, end and bow effects follow as a direct consequence of the nature of a linear array, terminated at each end; no further special mechanism is necessary. Moreover, the model makes predictions concerning detailed changes in the relations between accuracy, response time, and confidence, on the one hand, and stimulus position in the linear array, as the number of response alternatives is increased.

Speed-Accuracy Tradeoffs

A rich variety of other behavior and predictions can also be explored without any major development of the model. For example, Fig. 10 presents illustrative speed-accuracy tradeoff functions for individual responses as

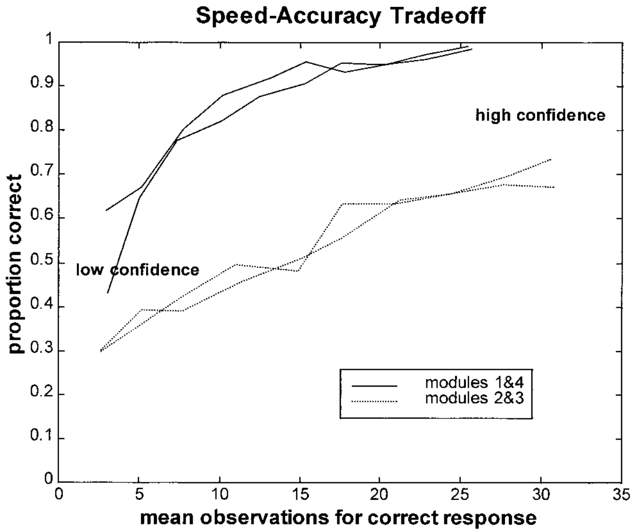

Fig. 10. Mean response accuracy of PAGAN vs. mean time required for a correct response for 4 choice identification with varying levels of target confidence.

target levels of confidence for a 4-choice identification task are subjected to global increases. Such negatively accelerated functions are typical of tradeoff relations between the percentage of correct responses and mean response time as the subject's degree of caution is varied (e.g., Pachella, Fisher, & Karsh, 1968; Pew, 1969; Schouten & Bekker, 1967; Swenssson, 1972; Taylor, Lindsay, & Forbes, 1967). In addition, the less bowed functions, obtained with the more difficult middle identifications is consistent with Swensson's finding that "increased stimulus difficulty . . . clearly seemed to make tradeoff functions less steep," while "unlike estimates of tradeoff slopes, the intercept estimates showed little systematic change with stimulus similarity (1972, p. 27).

Manipulations of Relative Caution

Finally, the readiness to make one or another response can also be manipulated, with results that capture the qualitative features of the results obtained in corresponding empirical studies. For example, Vickers (1979, pp. 273–283) has presented simulation results for a situation in which the target level of confidence for the fifth response in an array of eight is lowered, while the target levels for all other responses are maintained at a constant value. This simulation of the PAGAN model is intended to mimic an experiment in which the subject is instructed to watch particularly for one specific stimulus (in this case, the fifth). As shown by Vickers, the effect of this differential reduction in target confidence for one response is to make simulated responses to the corresponding stimulus faster, more likely to be correct, and (in defiance of the usual relations holding between these three dependent variables) *less* confident. At the same time, neighbouring responses become somewhat less accurate. Although responses to the two end stimuli, remain virtually unchanged, those to the seventh stimulus become slightly faster, while those to the second stimulus in the array become slower. Meanwhile, predicted confidence measures for all except the end stimuli are very slightly reduced.

Although these simulation results provide predictions that are both detailed and complex, they accord well with the few results on this topic. For example, in an 8-alternative choice reaction task, in which subjects were instructed to be especially ready to respond to the fifth or the sixth stimulus in a linear array of eight, Welford (1973) found that response times for the primed stimulus were substantially reduced, as were, to a lesser extent, responses to adjacent stimuli in the same (upper) half of the array. Conversely, times for responses in the other (lower) half were increased, while times for the two end stimuli remained virtually unchanged. Analogous results were obtained by Welford (1973) in a similar experiment, in which subjects were instructed to be especially ready to respond to stimuli in either the left or the right half of the array. As shown by Vickers (1979, pp. 282-3), the PAGAN model is successful in capturing the qualitative features of these results also. More generally, the behavior of the model is consistent with general suggestion by Falmagne (1965) and Audley (1973) that differential preparation should be associated with short response times and absence of special preparation with relatively longer times.

Effects of Changes in Experimental Conditions

Despite the dynamic interaction between response modules, evident in these examples, the above experimental manipulations are all designed to investigate primarily static properties of any process of choice reaction or absolute identification. They correspond roughly to the manipulations of response bias, caution, and discriminability considered by us at the beginning of our first paper on the self-regulating individual decision module Vickers and Lee (1998). Nevertheless, as shown by the evolution of the threshold levels in Fig. 9, the PAGAN model does not behave in a static way, even within the framework of a static experimental design. Although the parameters of the model remain constant throughout a simulated experimental session, its behavior changes continually, with the result that, at any given trial, the response of the model is a function of its entire past history of stimuli and responses.

As we have tried to indicate, the PAGAN model will adapt automatically, continually and differentially to any change in the number of different stimuli presented, in the mean, range, and distribution of their values, and to any change in the frequency with which they are presented (i.e., to *any* imaginable experimental manipulation within the fields of choice reaction and absolute identification). Unlike alternative models, including that of Lacouture and Marley (1991; 1995), such adaptations and improvements with practice are made *without the need for any external feedback or training whatsoever*. Since most human adaptation, learning, and improvement with practice is unsupervised, and takes place without the need for trial by trial feedback, this self-organising feature constitutes a major advantage of the PAGAN approach. Also in contrast to many models, whose predictions are limited to the small subset of experiments, that they were specifically designed to explain, the predictive scope of the PAGAN model far exceeds the currently available empirical evidence. For instance, the model makes detailed predictions concerning the accuracy and distributions of response time and confidence measures for each possible stimulus-response pairing in a multiple-alternative task. However, few (if any) experiments analyse this number of response measures in such fine detail.

A further important set of predictions concerns the ways in which the PAGAN model might adapt to changes in conditions within a particular experiment. For example, Vickers and Lee (1998) described adaptations by the individual three-category module to progressive changes in the magnitude (and sign) of differences between a variable and a standard stimulus, and also examined the response of the module to step changes in the probability of alternative responses. As evident from its performance in the above static situations, the PAGAN model will respond adaptively to any such change. For example, if the model is first exposed to a complete set of variable values (evoking each of the alternative responses), and is then presented with a subset of these values, it will automatically adjust its threshold values so as to maintain the target levels of confidence. As illustrated above, these target levels may be individually predetermined (because subjects can decide to concentrate on a particular stimulus), or the same target level may be assumed for all responses. In the latter case, it will be the average confidence over all responses that the model seeks to maintain. Unfortunately, while data from the psychophysical method of limits and from vigilance studies provide some illustration and test for the individual module, such experimental designs have not been used in tasks with multiple response alternatives.

Noise and Optimality

The PAGAN model also possesses a number of incidental but interesting properties that follow directly from its basic assumptions. For example, although the 'greater' and 'lesser' accumulators will receive input from a variable and a standard stimulus that are noise-free, provided their values are different, none of the three accumulators will receive any input from a noise-free variable and standard that have the same unvarying value. This means that the presence of variability in values of the variable and/or those of the standard is essential to the operation of the model. Further, as shown by Vickers (1979, pp. 284–290), the adaptive adjustments, made by the model in response to increases in noise, are such that it is possible to identify an optimum noise level. Above or below this level, the model performs less efficiently, in the sense of producing fewer correct responses in unit time. In addition, this optimum level is related to the number and discriminability of the stimulus alternatives (i.e., to task difficulty), as well as being determined by the target level of confidence. Although the notion of an identification process that depends upon the stimuli in question being at least momentarily different may seem paradoxical, the existence of noise at all levels in the human nervous system can safely be assumed. Moreover, the simulated occurrence of optimal levels of efficiency with varying amounts of noise, and their interaction with task difficulty and target confidence, is reflected in empirical phenomena, such as the U-shaped relations between activation and various measures of efficiency in human performance, often referred to as Yerkes-Dodson effects (Broadhurst, 1959; Eysenck, 1955; Welford, 1968). As pointed out by Broadbent (1965), these effects also show interactions between motivation and task difficulty.

Multidimensional Stimuli

Finally, it should be stressed that the model presented here has been developed to account for the pattern of results obtained in studies of choice reaction or absolute identification. For this reason, it is assumed that the standards have already been learned (and have stable representations) and that the modules are tuned to registering variations in the stimulus with respect to the relevant dimension (e.g., position in a linear array or line length). However, by adding further linear arrays of modules in parallel, the network could easily be extended to enable it to identify stimuli varying on several dimensions, and to show a preference for those dimensions on which the stimuli can be most effectively discriminated. In the same way, by utilising the latent capacity of each module to make 'greater' and 'lesser' decisions as well as 'equal' ones, such a network could quickly adapt so as to classify multidimensional stimuli into different categories in an optimally efficient way. For example, in a paired comparisons task, with multidimensional stimuli, such a network would automatically adjust so as to be more ready to make relative judgments on the basis of the most discriminable dimension(s). The capacity of each module to make greater, lesser and equal decisions, even when these may be redundant, also means that the network can use such decisions to revise its standards, to generate new standards in an unsupervised way, or to scale stimuli in terms of a number of given standards. Moreover, because of the modular nature of the approach, it is not difficult to maintain a morphological correspondence between the structure of the task and that of the model, so that its application to new tasks is relatively straightforward.

Further Extensions of the PAGAN Model

PAGAN seems to address two of the three weaknesses of neural network response time models described earlier. The dynamics of PAGAN are comparatively simple, operating within a modular architecture that is only locally connected, and learning is accommodated naturally within the model in the form of both self-organization of internal stimulus standards, and the adaptive setting of confidence-based decision criteria. Not suprisingly, therefore, the most immediate potential extension of PAGAN afforded by a neural network interpretation, comes through the remaining criticism relating to the distribution of information. In this regard, recent results by Pietsch and Vickers (1997), involving the capacity properties of information accumulation within PAGAN have strong implications for the way in which stimulus information is assumed to be represented in networks of stochastic modules, such as PAGAN.

Pietsch and Vickers (1997) challenge the 'leaky' accumulation model adopted by Usher and McClelland (1995), based on empirical data gathered from a task, like that of Vickers *et al.* (1971), in which observers discriminated the relative frequency of sequences of visual or auditory stimuli. Pietsch and Vickers evaluated numerous models, using multiple comparisons between response patterns of 47 individual (and pooled) observers and predictive measures based on separate simulations for each of 300 trial sequences. They found that results were clearly inconsistent with mechanisms of discrete sampling (with random registration failure), retroactive attenuation, or accumulation with leakage over time. Instead, the data implied that the memory of clusters of similar stimuli was represented as a set of discrete micro-representations. All of these might be accessed as part of the decision-making process, provided they were retained. However, any number of them might be lost (completely) during the course of a trial.

Pietsch and Vickers (1997) conclude that a 'loose capacity' model of the evidence accumulation process provides the best account of the data. The loose capacity account assumes that information is stored as a set of discrete micro-representations, which represent clusters of consecutive, identical stimulus elements (e.g., a run of flashes on the same side). These are subject to an all-or-nothing form of loss, rather than to any progressive reduction in value. The stochastic loss of micro-representations can potentially be induced by any additional registration of input, but with a probability which depends on the number of units held in memory at the time. Thus the loose capacity model is intermediate between accounts based on interference by subsequent activity and those based on a strict, fixed limit to the number of items that can be held at one time. This process implies a probabilistic steady state capacity in terms of information storage, but also allows considerable flexibility in this capacity.

Although it is not immediately obvious how to achieve a neural network implementation of this loose capacity mechanism, which is directed at modeling the representation of a stimulus varying within the space of a single trial, it is possible that trial-by-trial representations might be well captured by some form of distributed memory within a neural network. For example, Leth-Steenson (1998) has described a connectionist process for learning the linear ordering of symbolic stimuli, based on the pairwise relations between adjacent stimulus items, and has argued that this process, in conjunction with parallel evidence accrual processes, can give a good account of a range of results (including end and bow effects) from a symbolic comparison task. Meanwhile, the information storage properties of various neural networks have been widely explored, particularly relating to 'cross-talk' in associative memories (e.g., Knapp & Anderson, 1984; Kohonen, 1984; Willshaw, Buneman & Longuet-Higgins, 1969) and the 'catastrophic forgetting' phenomenon (e.g., French, 1991; Kruschke, 1993), in which connection weight changes which store new information concurrently serve to remove substantial amounts of previously stored information. Indeed, it is precisely these types of considerations which catalysed the development of the sparsely distributed memories mentioned earlier, and we suspect that these types of neural network memory structures could act as loose capacity mechanisms. However, the justification of this claim requires the specification of an exact means by which this correspondence is achieved, and this is obviously a priority for future research.

CONCLUSION

The strengths of neural network models have generally been seen as the capacity for learning and generalization, the ability to achieve solutions to otherwise intractable optimization problems, and the potential for distributed representation of stimulus information. Conversely, a perceived weakness of cognitive models developed within the neural network framework is that their flexibility and complexity make empirical verification difficult. In contrast, the comparatively well-researched features of simple judgments, and the relatively constrained development of the traditional stochastic latency mechanisms which have evolved to account for them, means that, while there is not consensus on the theoretical mechanisms involved, the most plausible candidates share a number of family resemblances. At the same time, only one group, which may labelled as accumulators or parallel stochastic integrators, has been extensively developed to account for the pattern of confidence ratings, as well as the accuracy and response time measures, which characterise performance in simple judgments tasks. Moreover, it is only within this group that a demonstrated potential exists for capturing dynamic aspects of performance adjustment, adaptation, self-regulation, and unsupervised learning.

In comparing the strengths and weaknesses of neural network and stochastic latency models, with respect to their accounts of simple judgment, our review suggests that the most fruitful approach may not be to regard these as opposing categories of model. Both styles of modeling share a number of assumed processes (such as integration over time, memorial representation, and activation thresholds), and it is not difficult to construct a neural network representation of a stochastic latency model. It seems arguable that the best strategy is to attempt to combine what seem to be the most advantageous features of each modeling framework. As an example of this strategy we have outlined a PAGAN model, based on a parallel array of self-regulating generalized accumulator modules.

Such an approach would seem to have a number of advantages. Firstly, in the form of confidence, it embodies a workable mechanism for self-regulation (which, incidentally, has considerable empirical justification, as well as being plausible from a Bayesian perspective). Secondly, it is open to extension and development to account for other sources of control (e.g., by the independent adaptation of referents or standards, or by shifts in starting values). Thirdly, the modules making up the system do not simply transfer excitation, but transmit the results of processing. This increased computation means that each module is essentially a self-contained, intelligent unit (cf. Minsky, 1986), and there is much less need for the network to be more richly interconnected than is computationally efficient or neurophysiologically plausible. Fourthly, the modules can be connected in a variety of ways, so there is no obvious constraint on the kind of network architecture in which they can be embodied.

Where both stochastic latency mechanisms and (paradoxically) neural networks both appear to be deficient is in the simplifying assumptions usually made concerning the representation of information. Evidence recently to hand clearly shows that, at least in the case of temporally extended stimuli, consisting of a series of elements, it may be necessary to assume that these elements are represented in a distributed form (e.g., as a set of discrete micro-representations), rather than in the form of a condensed summary or tally. In addition, it seems likely that it will be necessary to take into account the human observer's limited, if flexible, capacity for such representation. In developing a neural network implementation of models such as PAGAN this aspect emerges as a priority for future research.

REFERENCES

Ackley, D. H., Hinton, G. E., & Sejnowski, T. J. (1985). A learning algorithm for Boltzmann machines. *Cognitive Science, 9,* 147-169.

Anderson, J. A. (1991). Why, having so many neurons, do we have so few thoughts? In W. E. Hockley and S. Lewandowsky (Eds.), *Relating theory to data: Essays on human memory* (pp. 477-507). Hillsdale, NJ: Erlbaum.

Anderson, J. A. (1995). *An introduction to neural networks.* Cambridge, MA: MIT Press/ Bradford.

Anderson, J. R. (1990). *The adaptive character of thought.* Hillsdale, NJ: Erlbaum.

Arbib, M. A. (Ed.). (1995). *Handbook of brain theory and neural networks.* Cambridge, MA: MIT Press.

Audley, R. J. (1973). Some observations on theories of choice reaction time: Tutorial review. In: S. Kornblum (Ed.). *Attention & performance IV*. New York: Academic Press, 509-545.

Birnbaum, M. H., & Jou, J. (1990). A theory of comparative response times and "difference" judgments. *Cognitive Psychology, 22,* 184-210.

Broadbent, D. E. (1965). A reformulation of the Yerkes-Dodson law. *British Journal of Mathematical and Statistical Psychology, 18,* 145-157.

Broadhurst, P. L. (1959). The interaction of task difficulty and motivation: The Yerkes-Dodson law revived. *Acta Psychologica, 16,* 321-338.

Brooks, R. A. (1991a). *Intelligence without reason.* AI memo 1293, Massachusetts Institute of Technology. (ftp://publications.ai.mit.edu/ai-publications/1000-1499/AIM-1293.ps.Z).

Brooks, R. A. (1991b). Intelligence without representation. *Artificial Intelligence, 47,* 139-159.

Carpenter, G. A., & Ross, W. D. (1995). ART EMAP: A neural network architecture for object recognition by evidence accumulation, *IEEE Transactions on Neural Networks, 6(4),* 805-818.

Chappell, M., & Humphreys, M. S. (1994). An auto-associative neural network for sparse representations: Analysis and applications to models of recognition and cued recall. *Psychological Review, 101(1),* 103-128.

Cussins, A. (1990). The connectionist construction of concepts. In M. A. Boden (Ed.), *The philosophy of artificial intelligence* (pp. 368-441). Oxford: Oxford University Press.

Eysenck, H. J. (1955). A dynamic theory of anxiety and hysteria. *Journal of Mental Science, 101,* 28-51.

Falmagne, J. C. (1965). Stochastic models for choice-reaction time with applications to experimental results. *Journal of Mathematical Psychology, 2,* 77-124.

French, R. M. (1991). Using semi-distributed representations to overcome catastrophic forgetting in connectionist networks. *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society* (pp. 172-178). Hillsdale, NJ: Erlbaum.

Gigerenzer, G., & Goldstein, D.G. (1996). Reasoning the fast and frugal way: Models of bounded rationality. *Psychological Review, 103(2),* 650-669.

Glenberg, A. M. (1997). What memory is for. *Behavioral and Brain Sciences, 20,* 1-55.

Grossberg, S. (1976). Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. *Biological Cybernetics, 23,* 121-134.

Grossberg, S. (1980). How does the brain build a cognitive code? *Psychological Review, 87,* 1-51.

Grossberg, S. (1982). *Studies of mind and brain: Neural principles of learning, perception, development, cognition, and motor control.* Boston, MA: Reidel.

Grossberg, S. (Ed.). (1987a). *The adaptive brain, I: Cognition, learning, reinforcement and rhythm.* Amsterdam: Elsevier/North-Holland.

Grossberg, S. (Ed.). (1987b). *The adaptive brain, II: Vision, speech, language, and motor control.* Amsterdam: Elsevier/North-Holland.

Hale, D. J. (1968). The relation of correct and error responses in a serial choice reaction task. *Psychonomic Science, 13,* 299-300.

Hale, D. J. (1969). Speed-error trade-off in a three-choice serial reaction task. *Journal of Experimental Psychology, 81,* 428-435.

Hanson, S. J., & Burr, D. J. (1990). What connectionist models learn: Learning and representation in connectionist networks. *Behavioral and Brain Sciences, 13,* 471-518.

Haykin, S. (1994). *Neural networks: A comprehensive foundation.* Englewood Cliffs, NJ: Macmillan.

Hertz, J., Krogh, A., & Palmer, R. G. (1991). *Introduction to the theory of neural computing.* Redwood City, CA: Addison-Wesley.

Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4,* 11-26.

Hofstadter, D. R. (1985). *Metamagical themas: Questing for the essence of mind and pattern.* London: Viking.

Kanerva, P. (1988). *Sparse distributed memory.* Cambridge, MA: MIT Press/Bradford.

Keeler, J. D. (1988). Comparison between Kanerva's SDM and Hopfield-type neural networks. *Cognitive Science, 12,* 299-329.

Knapp, A. G., & Anderson, J. A. (1984). Theory of categorization based on distributed memory storage. *Journal of Experimental Psychology: Learning, Memory and Cognition, 10,* 616-637.

Kohonen, T. (1984). *Self-organization and associative memory.* New York: Springer-Verlag.

Kosko, B. (1988). Bidirectional associative memories. *IEEE Transactions on Systems, Man, and Cybernetics, 18,* 49-60.

Kruschke, J. K. (1993). Human category learning: Implications for backpropagation models. *Connection Science, 5,* 3-36.

Lacouture, Y., & Marley, A. J. (1991). A connectionist model of choice and reaction time in absolute identification. *Connection Science, 3,* 401-433.

Lacouture, Y., & Marley, A. J. (1995). A mapping model of bow effects in absolute identification. *Journal of Mathematical Psychology, 39,* 383-395.

Laming, D. R. J. (1966). A new interpretation of the relation between choice reaction time and the number of equiprobable alternatives. *British Journal of Mathematical and Statistical Psychology, 19,* 139-149.

Leth-Steenson, C. (1998). A connectionist, evidence accrual model of response times in symbolic comparison. *Unpublished thesis submitted for the degree of Ph.D.,* McGill University, Montreal.

Lippmann, R. P. (1987). An introduction to computing with neural networks. *IEEE acoustic speech and signal processing magazine, April,* 4-22.

Loftus, G. F., Busey, T. A., & Senders, J. (1993). Providing a sensory basis for models of visual information acquisition. *Perception & Psychophysics, 54,* 535-554.

Luce, R. D. (1986). *Response times.* New York: Oxford University Press.

McClelland, J. L. (1991). Toward a theory of information processing in graded, random, and interactive networks. In D. E. Meyer and S. Kornblum (Eds.), *Attention and performance 14: Synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (pp. 655-688). Cambridge, MA: MIT Press.

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: part 1. An account of basic findings. *Psychological Review, 88,* 375-407.

Merkel, J. (1885), Die zeitlichen Verhältniss der Willensthätigkeit. *Philosophische Studien, 2,* 73-127.

Minsky, M. (1986). *The society of mind.* New York: Simon and Schuster.

Nettelbeck, T., & Brewer, N. (1976). Effects of stimulus-response variables on the choice-reaction time of mildly retarded adults. *American Journal of Mental Deficiency, 81,* 85-92.

Norman, D. A. (1993). Cognition in the head and in the world: An introduction to the special issue on situated action. *Cognitive Science, 17,* 1-6.

Pachella, R. G., Fisher, D. F., & Karsh, R. (1968). Absolute judgments in speeded tasks: Quantification of the trade-off between speed and accuracy. *Psychonomic Science, 12,* 225-226.

Peterson, C., & Hartman, E. (1989). Explorations of the mean-field theory learning algorithm. *Neural Networks, 2,* 475-494.

Peterson, C., & Söderberg, B. (1995). Neural optimization. In M. A. Arbib (Ed.), *Handbook of brain theory and neural networks* (pp. 617-621). Cambridge, MA: MIT Press.

Pew, R. W. (1969). The speed-accuracy operating characteristic. *Acta Psychologica, 30,* 16-26.

Pietsch, A., & Vickers, D. (1997). Memory capacity and intelligence: Novel techniques for evaluating rival models of a fundamental information processing mechanism. *Journal of General Psychology,* 124, 229-339.

Ratcliff, R. & van Zandt. (1996). *Comparing connectionist and diffusion models of reaction time.* Unpublished manuscript, Northwestern University, Evanston, IL. [http://www.psych.nwu.edu/~roger].

Rumelhart, D. E. (1989). The architecture of mind: A connectionist approach. In: M.I. Posner (Ed.), *Foundations of cognitive science.* Cambridge, MA: MIT Press, 133-159.

Rumelhart, D. E., & McClelland, J. L. (1982). An interactive activation model of context effects in letter perception: part 2. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89(1),* 60-94.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., & Hinton, G. E. (1986). Schemata and sequential thought processes in PDP models. In J. L. McClelland and D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition, vol. 2* (pp. 7-57). Cambridge, MA: MIT Press.

Ruppin, E., & Yeshurun, Y. (1991). Recall and recognition in an attractor neural network model of memory retrieval. *Connection Science, 3(4), 381-400.*

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica, 27,* 143-153.

Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317-1323.

Shepard, R. N., & Kannappan, S. (1991). Connectionist implementation of a theory of generalization. In R. P. Lippmann, J. E. Moody, D. S. Touretzky (Eds.), Advances in neural information processing systems 3 (pp. 665-671). San Mateo, CA: Morgan Kauffman.

Shepard, R. N., & Tenenbaum, J. (1991). Connectionist modeling of multidimensional generalization. Paper presented at the Thirty Second Annual Meeting of the Psychonomic Society, San Francisco, November.

Simon, H. A. (1982). *Models of bounded rationality.* Cambridge, MA: M.I.T. Press.

Smith, G. A. (1977). Studies of compatibility and a new model of choice reaction time. In S. Dornic (Ed.), *Attention and Performance VI,* (pp. 27-48). Hillsdale, N.J: Lawrence Erlbaum.

Smolensky, P. (1988). On the proper treatment of connectionism. *Behavioral and Brain Sciences, 11,* 1-23.

Swensson, R. G. (1972). The elusive trade-off: Speed versus accuracy in visual discrimination tasks. *Perception and Psychophysics, 12,* 16-32.

Taylor, M. M., Lindsay, P. H., & Forbes, S. M. (1967). Quantification of shared capacity processing in auditory and visual discrimination. *Acta Psychologica, 27,* 223-229.

Trabasso, T. R., & Riley, C. A. (1975). On the construction and use of representations involving linear order. In: R. L. Solso (Ed.), *Information processing and cognition: The Loyola symposium.* Hillsdale, NJ: Lawrence Erlbaum, 201-229.

Trabasso, T. R., Riley, C. A., & Wilson, E. G. (1975). The representation of linear order and spatial strategies in reasoning: A developmental study. In: R. Falmagne (Ed.), *Reasoning: Representation and process.* Hillsdale, NJ: Lawrence Erlbaum, 201-229.

Usher, M., & McClelland, J. L. (1995). *On the time course of perceptual choice: A model based on the principles of neural computation.* Technical report PDP.CNS.95.5: Center for the neural basis of cognition. [ftp://hydra.psy.cmu.edu/pub/pdp.cns/pdp.cns.95.5.ps.Z].

Vickers, D. (1979). *Decision processes in visual perception.* New York: Academic Press.

Vickers, D., Caudrey, D., & Willson, R. (1971). Discriminating between the frequency of occurrence of two alternative events. *Acta Psychologica, 35,* 151-172.

Vickers, D., & Lee, M. D. (1998). Dynamic models of simple judgments: I. Properties of a self-regulating accumulator module. *Nonlinear Dynamics, Psychology, and Life Sciences, 2,* 169-194.

Welford, A. T. (1968). *Fundamentals of skill.* London: Methuen.

Welford, A. T. (1971). What is the basis of choice-reaction time? *Ergonomics, 14,* 679-693.

Welford, A. T. (1973). Attention, strategy, and reaction time: A tentative metric. In S. Kornblum (Ed.), *Attention and Performance IV, and (pp. 37-53). New York: Academic Press.*

Welford, A. T. (1980). Reaction times. New York: Academic Press.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica, 41,* 67-86.

Willshaw, D. J., Buneman, O. P., & Longuet-Higgins, H. C. (1969). Non-holographic associative memory. *Nature, 222,* 960-962.

Woocher, F. D., Glass, A. L., & Holyoak, K. J. (1978). Positional discriminability in linear orderings. *Memory and Cognition, 6,* 165-173.

Yuille, A .L. (1995). Constrained optimization and the elastic net. In M.A. Arbib (Ed.), *Handbook of brain theory and neural networks* (pp. 250-255). Cambridge, MA: MIT Press.