British Journal of Psychiatry (1993), **163**, 308–314

Recognising Psychiatric Symptoms
Relevance to the Diagnostic Process

G. E. BERRIOS and E. Y. H. CHEN

Current overemphasis on nosological diagnosis has led to a neglect of the process of symptom recognition. There is evidence, however, that the perception of the symptom alone does not guarantee symptom ascertainment since a decision-making component is also involved. To achieve the latter, additional information must be provided by the contextual cues implicit in the ongoing diagnostic hypothesis. Current diagnostic systems, however, still assume a two-stage model according to which symptom and disease recognition are independent cognitive events. This paper suggests that this model is inadequate and that descriptive psychopathology is not *transparent*. It then describes a neural network simulation to make various aspects of the problem explicit. This takes into account the multidimensional and probabilistic aspects of symptom recognition and is, from this point of view, superior to traditional algorithmic models. It also has the capacity to represent the different cognitive styles involved in symptom recognition.

Current overemphasis on nosological diagnosis and on the reliability of diagnostic check-lists – for example, DSM–III–R (American Psychiatric Association, 1987) and ICD–10 (World Health Organization, 1992) – has led to a neglect of the problems involved in symptom recognition. Most diagnostic systems consider symptom ascertainment to be unproblematic, and descriptive psychopathology as 'transparent'. This assumption is unwarranted; indeed, there is evidence that the recognition of the symptoms and signs of mental illness is criterion dependent, that is, it also involves a decision-making component.

Current diagnostic instruments also assume that any problems associated with symptom recognition can be resolved by providing operational definitions (of delusions, hallucinations, depersonalisation, etc.). However, these definitions are primarily addressed at 'constructing' the symptom, hence are of little value; furthermore, there are no clear rules for their application. For example, DSM–III–R defines 'depersonalisation' as:

> ". . . an alteration in the perception or experience of the self so that the feeling of one's own reality is temporarily lost. This is manifested in a sense of self-estrangement or unreality, which may include the feeling that one's extremities have changed in size, or a sense of seeming to perceive oneself from a distance (usually from above)."

This definition includes a list of experiential events and dimensions, the presence of any of which signifies the existence of the symptom, and which can occur in various clinical contexts. The observer, however, is given no rules to cope with such ambiguities.

In reality, the perception of such symptom-related events or dimensions does not complete the process of symptom ascertainment. A decision-making component and a context are always required. In other words, decision making requires additional information which the observer obtains from contextual cues which are usually provided by the diagnostic hypothesis.

Psychiatric training has traditionally focused on disease recognition. Symptom recognition, on the other hand, is taught by 'ostensible' definition, that is by the iterative demonstration of cases with the 'same' symptom. Trainees are expected to generate and memorise 'prototypes' which then can be used as templates for subsequent symptom recognition. This learning occurs at the same time as disease recognition. It is likely that, therefore, both processes become inextricably linked in the mind of the psychiatrist early on in training. However, current diagnostic systems, such as DSM–III–R, assume that symptom recognition and disease recognition are successive and independent cognitive events. Indeed, their reliability (and perforce validity) requires that such events do not contaminate each other. According to this model, the first event is the recognition of the 'units of analysis' or 'building bricks'; the second, their synthesis into a 'diagnosis'. The objective seems to be the creation of a tight decisional cascade according to which, 'given that a, b, and c are obtained, then D is the case', where a, b, and c are 'criteria' (that is, symptoms), and D, 'diagnosis'. The reliability of this model is based on the view that whenever a, b, and c are present, D will have to be diagnosed.

What we want to suggest in this paper is that these diagnostic systems are caught in the horns of a

https://doi.org/10.1192/bjp.163.3.308 Published online by Cambridge University Press

dilemma. On the one hand, if a two-stage model is adopted, then the fact will have to be faced that a clandestine contamination of symptom recognition is taking place and new foundations for the putative reliability of the system sought for; on the other, if the two stages are made truly independent, then new ways will have to be found to make symptoms less ambiguous, that is, a far more complex science of symptom recognition will have to be created.

An illustration of this problem can be found in recent attempts to computerise the diagnostic process. Whatever the level of sophistication of the programs involved, the fact remains that few provide subroutines to deal with independent symptom recognition. Thus, the validity (and reliability) of the computerised instrument depends entirely upon the quality of the information provided as input data (i.e. symptoms). This paper offers an account of this problem and describes a neural network simulation that models the situations described above.

Conceptual issues

'Descriptive psychopathology', 'psychiatric semiology', or the 'language' of psychiatry can be said to have developed during the 19th century. The process of development has been described in detail elsewhere (Berrios, 1984, 1988). Suffice it to say that during the early 20th century this 'language' was gratuitously called 'phenomenological', thus starting a confusion that has lasted to the present day. Like all languages, psychopathology includes a vocabulary, a grammar, a syntax, and a set of rules for their application. While the vocabulary has become enshrined in traditional teaching and has acquired a 'sacred' status, the application rules (i.e. the guidelines according to which a psychiatrist decides whether a particular fragment of behaviour is to be 'named' as symptom a, b, or c) have been sadly neglected. These rules, which were doubtlessly articulated during the early stages of the formation of the language, seem to survive only in the oral tradition built into the apprenticeship system.

The redundant nature of psychiatric diagnosis allows for the possibility of occasional lapses in symptom recognition; for example, if out of six criteria for schizophrenia four are correct and 2 misidentified, the final dignosis may still be reached. This may cause the belief that symptom recognition is far more efficient than in fact it is. The issue, therefore, is whether a notion of 'efficiency' can be developed with regard to symptom recognition. We would suggest that efficiency might be defined here as the capacity of a language to achieve a high 'hit rate' in its naming function; for example, every time a fragment of behaviour is called symptom a, b, or c, the name will fit. An efficient language would be one that extracts the maximum amount of information available in a given behaviour.

This leads to the question of whether all the symptoms listed as the components of a particular disease have, in fact, the same informational import. Clinical practice suggests that they may not. However, current diagnostic systems do not offer means of assigning such weights or otherwise creating symptom hierarchies. The fact that in some cases symptom misidentification does not seem seriously to affect diagnosis suggests some variation in the quantum of information carried by each symptom. Interestingly, the fact that descriptive psychopathology has so far been unable to make such differences explicit, suggests that its efficiency is limited.

The points listed above suggest that there is little reason to consider descriptive psychopathology as transparent, and that such language may, in fact, be conceptually parasitical upon the diagnostic process itself. If so, more research will be needed into the real nature of this language, on its conceptual boundaries, and on whether the two-events model of diagnosis is sufficiently heuristic.

A computational model

Within the last decade, artificial neural networks (ANNs) (or parallel distributed processing) have been recognised as useful models of cognitive processes. These consist of sets of interconnected processing units, each capable of independent computation. Computed outcomes are communicated to other units via weighted connections. The latter can be modified according to definable learning algorithms. The emergent properties of ANNs have been applied to cognitive processes such as associative memory, categorisation, and generalisation.

'Constraint satisfaction networks' are a particular class of ANN (McClelland & Rumelhart, 1988). These can be defined as systems where the weights of the internal connections 'constrain' the state of the various associated units. This limits the number of possible states the system can evolve into from a starting point. A typical example is the network studied by Hopfield (1982). Hopfield networks have an analogy with a known physical system (Ising ferromagnetism), and this has allowed a formal description of their behaviour. Rumelhart *et al* (1986) have also used 'constraint satisfaction networks' to model cognitive processes such as the representation of schemata. We propose here that this type of network may also be useful in modelling symptom recognition.

https://doi.org/10.1192/bjp.163.3.308 Published online by Cambridge University Press

Constraint satisfaction networks

In a constraint satisfaction network, each processing unit can be taken to represent a hypothesis (for example a particular 'feature' or 'dimension' of a symptom). Connections between units represent constraints among hypotheses. Thus, a positive connection between A and B would imply that whenever A is present B is 'expected' (probabilistically) to be present; a negative connection that whenever A is present B is 'expected' to be absent. When A and B are not connected, their behaviours would be independent of each other (i.e. unrelated). A positive input from outside the network to a particular unit would mean that there is evidence that the relevant feature is present; a negative input would mean the opposite. The strength of the evidence is reflected in the magnitude of the input.

When such a network is run, it will eventually settle into an 'optimal' state, defined as one in which as many as possible of the constraints are satisfied, and in which priority is given to the strongest constraints. The network is then said to have 'relaxed' into a 'local' solution. In contrast, a 'global' solution is attained when a stochastic (non-deterministic) rule is used to determine activation of the units (by a process called simulated annealing – for details see Hinton & Sejnowski, 1986).

Evaluation of psychiatric symptoms

The cognitive processes involved in symptom recognition can be modelled on the basis of a system of constraint satisfaction networks. As a first approximation, a two-stage process is considered (Fig. 1).

Recognition of symptoms

At stage 1 there are N networks (Y_A, Y_B, . . . Y_N) each dedicated to recognising a symptom (say, a hallucination). Individual networks receive a number (k) of inputs (features 1 to k), each representing a particular symptom characteristic or magnitude (e.g. vividness, frequency, timing). For each symptom, a set of external observations is represented in the form

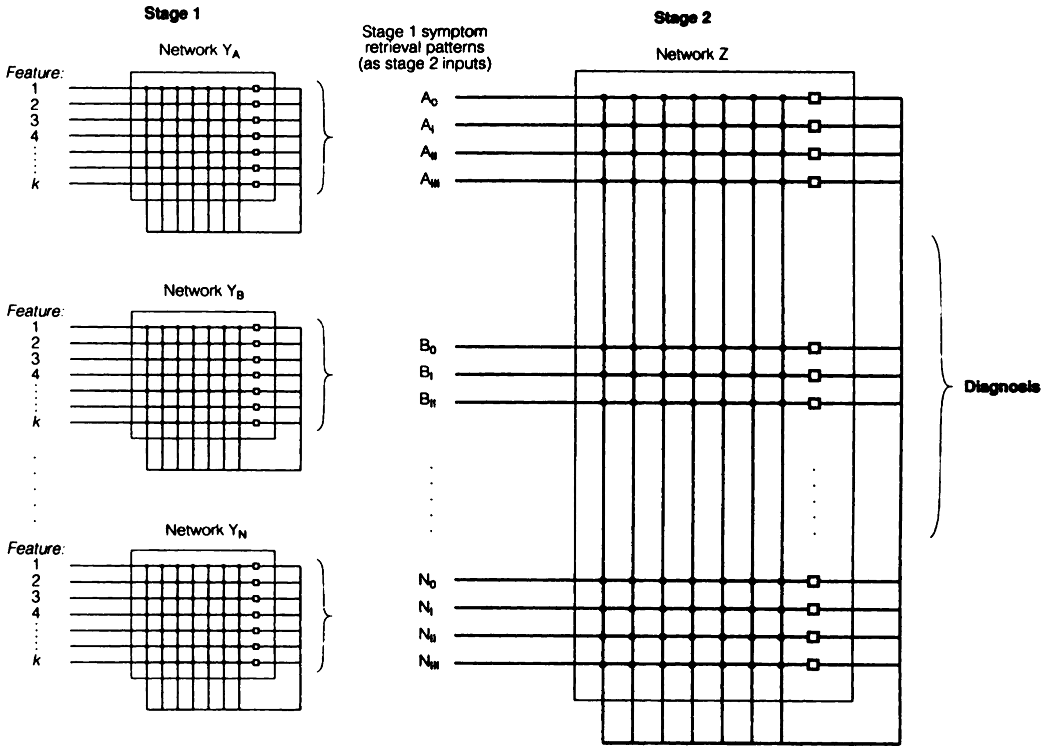

Fig. 1 A hierarchical system of networks processing psychiatric symptoms.

of an input vector with k elements which is then mapped onto the k units of the network. One vector will thus act as the starting point for, say, network Y_A. From previous learning, network Y_A has stored a number p of stable patterns as k-dimensional vectors (retrieval patterns): for example, for symptom A, there are four stored patterns (A_0, A_I, A_{II}, A_{III}). These are encoded as connection weights to guide inputs entering the network to settle in one of the retrieval patterns. The set of input vectors which will relax to a particular pattern, say A_{III}, is called the 'basin of attraction of A_{III}'.

In the second stage, end interpretations of stage 1 networks serve as input for network Z. This receives input from each stage 1 network (Y_A, Y_B . . . Y_N). Inputs to network Z are recognised symptoms. Internal weights in network Z store information on association between symptoms. The stored stable patterns correspond to symptom clusters or 'diagnoses'. Input vectors for network Z are N-dimensional vectors resulting from stage 1 processing, and represent the particular combination of symptoms. Network Z settles into an interpretation by retrieval of one of the stored patterns of 'diagnosis'.

When such state has been reached, the internal constraints are said to be optimally satisfied, and an interpretation of the input data to have taken place. This process is repeated for each symptom, that is, one at a time.

Recognition of disease

Stage 2 shows a network dedicated to disease recognition. Symptom recognition, as carried out during stage 1, provides the input for this network. The internal weights of this network store information about putative symptom associations. Stable patterns will correspond to symptom clusters or 'syndromes'. The input vectors for this network (retrieval patterns from stage 1) represent particular combinations and permutations of symptoms (as determined by the examination of the patient). The network finds an interpretation by settling into one of the stored patterns or 'syndromes'.

According to this model, the 'diagnostic process' starts with an 'unbiased' observation of phenomenological data, assumed to be theory free (stage 1). This is followed by a diagnostic decision based on the symptoms collected (stage 2).

Processing under a 'primed' condition

We suggest that the above model – which is assumed by current diagnostic systems – is inadequate to represent the real clinical situation for symptom recognition as it is not entirely unbiased. As mentioned in the introduction, such recognitions are likely to be affected by decisions taken on the basis of contextual features (for example demographic details such as age, sex, etc.), and on-going diagnostic hypotheses. (These influences are referred to here as 'priming'.) To model 'priming', the two-stage multi-network system described above needs to be modified by introducing 'bidirectional' connections between stage 1 and stage 2 (Fig. 2a,b).

In such a model, stages 1 and 2 are no longer considered as occurring in series but in parallel. For example, after a symptom A has been recognised by retrieving A_I from among the stored patterns (A_0, A_I, A_{II}, . . ., A_N), the pattern A_I is entered into the stage 2 network. In the absence of other inputs, the activation of A_I will force a corresponding state of activity in the stage 2 network. Now, suppose that among the stored patterns of symptom B (B_0, B_I, B_{II}, . . ., B_N), A_I is particularly associated with B_{II}. In such a case, the association would be encoded in the stage 2 network as a large positive weight between A_I and B_{II}. In the presence of A_I, B_{II} becomes partially activated. If the relative activation of B_{II} was fed-back to stage 1, the effect is that during evaluation of symptom B there would be a tendency for the network to settle towards pattern B_{II}. Analytically, such a tendency is equivalent to the presence of a weak external input field corresponding to pattern B_{II} (Amit, 1989).

Another way of describing this 'priming' effect is to say that the 'energy function landscape' becomes distorted so that the set of input vectors that is attracted to B_{II} is enlarged (see below). Some input vectors which previously would have settled in, say, B_I may now settle in B_{II}. In other words, *exactly the same external features* for symptom B would be interpreted *differently* because of the priming effect of A_I.

The processing of symptom B could be generalised to the recognition of subsequent symptoms. Symptom recognition is, therefore, a function of external data, stored patterns (from long-term store), as well as of 'priming' from preceding assessments via stage 2 network.

Simulation experiment

To illustrate this theoretical discussion, a simple simulation experiment is described. It shows the processing of a symptom at stage 1, and the effect of 'priming' on the dynamics of the network in retrieving two hypothetical patterns.

Method

The processing of a symptom at stage 1 will be modelled by means of a 16-unit constraint satisfaction network. This

https://doi.org/10.1192/bjp.163.3.308 Published online by Cambridge University Press

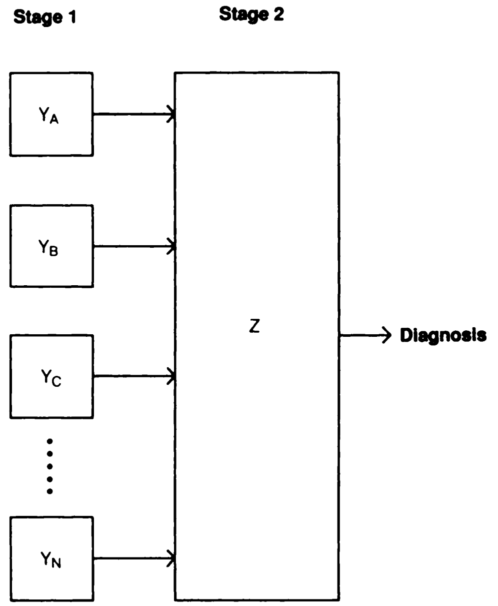

Fig. 2a Each symptom in stage 1 is recognised independently. The resulting information is used as data for the stage 2 network Z.

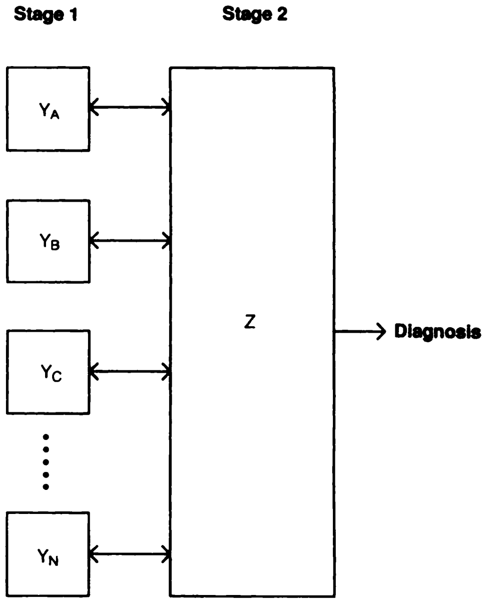

Fig. 2b With bidirectional connections between stages 1 and 2, the state of network Z can influence processing in stage 1. For example, after processing of the symptom A in network Y_A, the result is registered in network Z. The resulting pattern of partial activation in network Z then 'primes' unit Y_B and distorts its energy landscape, affecting the dynamics of the recognition of symptom B.

is a fully connected network with bounded continuous activation levels, employing an asynchronous update rule (for details of the original program, of which the one used here is a modification, see McClelland & Rumelhart, 1988). Two patterns are stored in the network by prior specification of the weight matrix. The same storage could be achieved by *training* the network with repeated presentation of the two patterns using an unsupervised learning algorithm (such as the Hebbian rule).

Simulation is performed by running the network with randomised update sequences. After each run, the final state (the retrieval state) and the measure of 'energy' corresponding to this state are noted. The 'priming' effect is modelled by applying a weak external field (see above) to the network. Thus, each unit which is active in the primed pattern receives a weak 'priming' input of 0.1 on a scale of 0 to 1. The primed network is run with randomised update sequences as in the unprimed network.

Both networks were run with a constant stochastic activation function (temperature = 2), and a baseline 'bias' of 0.5 (that is, in the absence of input, each updated unit has a probability of 0.5 of being activated). Coefficients for relative strength of internal connections and external input were 0.4 and 0.2 respectively. Since there was a 'bias' factor, the initial network states were strongly related to the early sequence of updating. Thus, using a randomised update sequence, random initial states are represented.

Results

Results of the simulation (Table 1) can be visualised as changes in the energy landscape associated with the primed and unprimed network states. In Fig. 3a,b, the vertical dimension represents the energy function. The horizontal plane is a simplified two-dimensional representation of network states' space. Retrieval states of the two patterns are represented by two minima on the energy surface. The size of the basin of attraction, that is the area around each retrieval state within which the network will be attracted to the retrieval state, is related to the probability of attaining that

Table 1
Changes in the energy landscape associated with the primed and unprimed networks

	Unprimed		Primed	
	Energy[1]	Probability[2]	Energy	Probability
Pattern 1	-6.4	0.38	-7.2	0.61
Pattern 2	-6.4	0.33	-6.4	0.12

1. Relative 'energy' is expressed as the negative magnitude of a 'goodness-of-fit' measure (McClelland & Rumelhart, 1988).
2. Probability refers to the probability of retrieving the pattern with random sequence updating. For each condition, 100 trials were performed.

https://doi.org/10.1192/bjp.163.3.308 Published online by Cambridge University Press

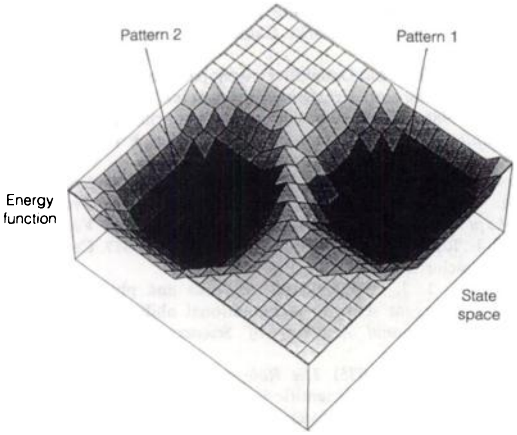

Fig. 3a Energy landscape for 'unprimed' network.

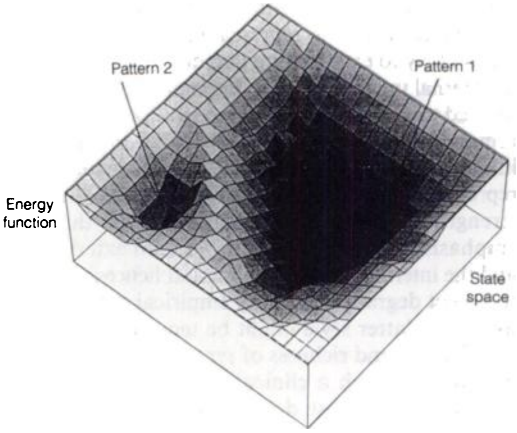

Fig. 3b Energy landscape for network 'primed' for pattern 1.

retrieval state from a random starting network state. The depth of the minima is a measure of energy (i.e. the extent to which external and internal constraints are satisfied in the network). In Fig.3a,b, spurious retrieval states (Amit, 1989) are ignored as they are not directly relevant to the present model.

In the unprimed condition, the basins of attraction (i.e. the set of initial states which end up being interpreted as a particular pattern) (Amit, 1989) of each of the stored patterns occupy about 36% of the state space, and the energy associated with each pattern is equal. In the primed condition, the primed pattern has a basin of attraction extending over 60% of the state space, at the expense of that of the unprimed pattern, which shrinks to 12%. The energy associated with the primed pattern is also decreased.

Discussion

The simple simulation shown above takes into account the multidimensional and stochastic nature of the cognitive processes that might be involved in the recognition of psychiatric symptoms. It postulates the existence of a cognitive system that strives to maximise satisfaction of internal and external constraints to arrive at a 'best-fit' interpretation of the incoming data concerning the symptom.

Top-down versus bottom-up processing

The effect of the 'top-down' type of influence has thus been illustrated as a 'priming' effect on the network. The model suggests a way of formally articulating the hypothesis mentioned earlier that individual symptom recognition is never an *independent* exercise. Instead, at every stage of the process there seems to be an interaction between symptom and disease recognition.

In contrast to the 'prototype' approach to diagnosis (as used in ICD-9; World Health Organization, 1978), where the actual clinical picture of the patient is compared as a *whole* with an ideal conception of the illness, the 'operational definition' approach (as in DSM-III-R) assumes that independent symptom recognition must precede disease recognition. Such assumption does not take into account the possibility of an interaction between these two cognitive events. In theory, the reliability of the 'operational definition' approach depends on the crucial assumption that disease recognition is independent from symptom recognition; in practice, however, this does not seem to be the case as the model described in this study suggests. It is more likely that an ongoing interaction takes place between symptom and disease recognition; in fact, diagnostic possibilities are already narrowed down early in the clinical process (see Kendell, 1975).

Typical versus atypical symptoms

Although it is suggested that typical symptoms, that is those located close to the prototype patterns in the state space, may be less affected by 'priming', the converse also holds true, namely, that those symptoms falling on the boundary between two prototypes are more susceptible to 'priming'. This, as shown in Fig. 3a,b, is a direct consequence of a distortion of the energy landscape caused by the expansion of one pattern at the expense of the other (which is, in turn, caused by priming).

Cognitive styles and internal versus external connection strengths

In the above simulation, the relative strengths of internal and external connections in the networks at

https://doi.org/10.1192/bjp.163.3.308 Published online by Cambridge University Press

stage 1 were set as constants. However, they could potentially be used to represent the different cognitive styles likely to exist among doctors. A high weighting to external input (relative to internal connections) may be used to describe a style of recognition that wants to remain faithful to empirical observation, and which has a good degree of tolerance for dissonant internal representations. In contrast, a high internal connection strength suggests a recognition style that puts emphasis on goodness-of-fit between external data and the internal representation, and hence is prepared to accept degradation of the empirical data. Doctors using this latter style might be tempted to suppress the diversity and richness of symptom manifestations in order to reach a clinical conclusion.

In practice, most doctors are likely to operate between these two styles. One lesson to be drawn from this analysis is that trainee psychiatrists should develop an early awareness of their cognitive style. This, and the other problems outlined in this paper, constitute areas of urgent investigation, particularly since any future maximisation of diagnostic validity depends on their solution.

References

American Psychiatric Association (1987) *Diagnostic and Statistical Manual of Mental Disorders* (3rd edn, revised) (DSM-III-R). Washington, DC: APA.

Amit, D. J. (1989) *Modelling Brain Function: The World of Attractor Neural Networks*. Cambridge: Cambridge University Press.

Berrios, G. E. (1984) Descriptive psychopathology: conceptual and historical aspects. *Psychological Medicine*, **14**, 303–313.

—— (1988) Historical background to abnormal psychology. In *Adult Abnormal Psychology* (eds E. Miller & P. J. Cooper), pp. 26–51. Edinburgh: Churchill Livingstone.

Hinton, G. E. & Sejnowki, T. J. (1986) Learning and relearning in Boltzmann machines. In *Parallel Distributed Processing: Exploration in the Microstructure of Cognition Vol. 1* (eds D. E. Rumelhart & J. L. McClelland), pp. 282–317. Cambridge, Massachusetts: The MIT Press.

Hopfield, J. J. (1982) Neural networks and physical systems with emergent selective computational abilities. *Proceedings of the National Academy of Sciences of the USA*, **79**, 2254.

Kendell, R. E. (1975) *The Role of Diagnosis in Psychiatry*. Oxford: Blackwell Scientific Publications.

McClelland, J. L. & Rumelhart, D. E. (1988) *Explorations in Parallel Distributed Processing: A Handbook of Models, Programs and Exercises*. Cambridge, Massachusetts: The MIT Press.

Rumelhart, D. E., Smolensky, P., McClelland, J. L., *et al* (1986) Schemata and sequential thought processes in parallel distributed processing models. In *Parallel Distributed Processing: Exploration in the Microstructure of Cognition* (2nd edn) (eds D. E. Rumelhard & J. L. McClelland), pp. 7–57. Cambridge, Massachusetts: The MIT Press.

World Health Organization (1978) *Mental Disorders: Glossary and Guide to their Classification in Accordance with the Ninth Revision of the International Classification of Diseases* (ICD-9). Geneva: WHO.

—— (1992) *The ICD-10 Classification of Mental and Behavioural Disorders*. Geneva: WHO.

*G. E. Berrios, MA (Oxon), MD, FRCPsych, FBPsS, *Consultant and University Lecturer in Psychiatry, Department of Psychiatry, University of Cambridge, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ*; E. Y. H. Chen, MA (Oxon), MBChB, MRCPsych, *Lecturer in Psychiatry, University of Hong Kong, formerly Senior Registrar in Psychiatry, Department of Psychiatry, University of Cambridge*

*Correspondence

https://doi.org/10.1192/bjp.163.3.308 Published online by Cambridge University Press