PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Is Specific Language Impairment a Valid Diagnostic Category? Genetic and Psycholinguistic Evidence

D. V. M. Bishop

Phil. Trans. R. Soc. Lond. B 1994 **346**, 105-111
doi: 10.1098/rstb.1994.0134

Is specific language impairment a valid diagnostic category? Genetic and psycholinguistic evidence

D. V. M. BISHOP

MRC Applied Psychology Unit, 15 Chaucer Road, Cambridge CB2 2EF, U.K.

SUMMARY

Specific language impairment (SLI) is diagnosed when a child fails to develop language normally for no apparent reason: hearing and intelligence are adequate and the social environment is unexceptional. Definitions of SLI typically specify that the child must have a substantial discrepancy between language ability and non-verbal IQ. However, data from a twin study question the validity of this requirement, and indicate that SLI is not genetically distinct from less specific disorders where language impairment occurs in the context of low average or borderline non-verbal ability. A second question concerns the heterogeneous language symptoms seen in SLI: do these correspond to distinct conditions, or to different phenotypic manifestations of a common underlying disorder, or are they merely random variations resulting from unreliable assessments? The last of these possibilities is ruled out by the finding that twins who are concordant for language disorder show good agreement in terms of the pattern of language impairment. However, systematic variation in the age and ability of children in different SLI subgroups suggest that these may correspond to variable manifestations of a core inherited language disorder, rather than distinct diagnostic entities.

1. INTRODUCTION

It has long been recognized that some children fail to develop normal language for no apparent reason; hearing and intelligence are adequate, the home environment is unexceptional, yet ability to produce and/or comprehend language is well below age level. The terminology used to refer to such children has varied over the past few decades: 'developmental aphasia' and 'developmental dysphasia' have waned in popularity as it has become recognized that they imply more understanding of the neurological basis than is the case. More neutral diagnostic terms with fewer medical overtones, such as 'developmental language disorder' or 'specific language impairment' (SLI) are now usually preferred.

Formulations of diagnostic criteria (see, for example, Stark & Tallal 1981; American Psychiatric Association 1987) stress three key features: (i) language must be significantly below age level; (ii) poor language functioning must be out of keeping with other aspects of development (this is normally interpreted as a substantial discrepancy between a language test score and non-verbal IQ); (iii) possible physical causes (e.g. hearing loss, acquired brain damage or abnormal structure of the articulators) and recognized syndromes associated with language delay (e.g. autism) are excluded.

This definition embodies conventional wisdom about SLI, but few attempts have been made to validate the criteria, i.e. to establish that they do define a coherent disorder with a distinct aetiology. Two questions that merit further scrutiny are: (i) how distinct is specific language impairment from more general developmental delay affecting all cognitive functions? and (ii) how may subgroups of SLI are there?

(*a*) *Is the distinction between specific language impairment and more global cognitive impairment valid?*

Conventional definitions of SLI stress the importance of differentiating specific speech and language disorders from more global developmental delays. An analogy may be drawn with cases where language is impaired for an obvious reason: children with profound congenital hearing loss tend to have poor oral language but normal non-verbal intelligence. In such cases, the non-verbal IQ is regarded as providing an estimate of 'true' intellectual potential. In cases of SLI, we do not know the cause of the language impairment, but the existence of a large verbal–non-verbal discrepancy is taken as circumstantial evidence of a specific aetiology that interferes selectively with language.

This approach to the diagnosis of SLI has a clear parallel with the definition of 'developmental dyslexia' or specific reading retardation, where the diagnostic criteria specify that the child's poor reading must be out of keeping with age and ability. However, as Stanovich (1994) has pointed out, there is remarkably

Phil. Trans. R. Soc. Lond. B (1994) **346**, 105–111
Printed in Great Britain

little evidence for the validity of this approach with respect to dyslexia. Children with a large IQ–reading discrepancy are similar to poor readers of low ability in terms of genetic risk factors, psychological profile and associated problems with phonological processing (Pennington *et al.* 1991; Fletcher *et al.* 1992). It would seem that to regard the IQ–reading discrepancy as a cornerstone of the diagnostic criteria for dyslexia is misguided.

Exactly the same issues apply in the case of SLI. Stark & Tallal (1981) and Aram *et al.* (1992) found that a high proportion of children who are thought on clinical grounds to have SLI fail to show a substantial discrepancy between language and non-verbal ability on psychometric tests. Such findings suggest that we need to examine carefully the validity of using a non-verbal–verbal discrepancy to identify a clinical group. Cole *et al.* (1992) have argued that one reason why we should be cautious of using a discrepancy criterion is that the *reliability* of a non-verbal–verbal discrepancy may be unsatisfactory; they found that many children who met discrepancy criteria for SLI on one occasion failed to do so on another, and vice versa.

(*b*) *Are there distinct subgroups of SLI?*

A further issue is whether SLI is a unitary condition or a heterogeneous collection of disorders. There is no doubt that the clinical manifestations of SLI can be extremely variable. Some children have problems predominantly with production of speech sounds; others make many grammatical errors but appear to understand normally. Cases have been described of children with severely restricted comprehension, and others of children who speak fluently in complex sentences but who give tangential answers to questions (Bishop & Rosenbloom 1987).

Although most specialists recognize that there is considerable variation from child to child, attempts to devise a classification of SLI have not been conspicuously successful. Early studies applying multivariate methods such as cluster analysis to derive subgroups from language test data used samples far too small to give reliable results (Aram & Nation 1975; Wolfus *et al.* 1980). Beitchman *et al.* (1989) applied cluster analysis to data from five speech–language measures from a large sample of five-year-olds, including 301 children who failed a speech–language screening. They distinguished three subgroups: those with pure articulation disorders ($n = 174$), those with impairment only on an auditory comprehension test ($n = 56$), and those with low performance on all language measures ($n = 87$). The problem with this result is that it lacks face validity; in clinical experience it is most unusual to find a child with poor comprehension and normal expression, whereas the converse is relatively common. Furthermore, as Wilson & Risucci (1986) note, the results of a cluster analysis depend crucially on the measures entered into the analysis. Standardized tests are not sensitive to all aspects of clinical presentation: for example, it is difficult to identify pragmatic difficulties or dyspraxic speech except by clinical appraisal. Thus critical features may go unrecognized if one relies solely on test scores. Wilson & Risucci suggest that an approach that starts with clinically defined categories and then attempts to validate these by quantitative analyses is more appropriate. They report modest success in validating a clinically derived five-subgroup classification against a classification derived from a cluster analysis. Agreement between the clinicians and the cluster analysis was 46% (Cohen's $\kappa = 0.32$), which is above chance, but far from impressive, especially when one considers that the clinicians and the cluster analysis based their classifications on the same data. Furthermore, out of 93 children, only 78 could be categorized by the clinicians, and the remainder were excluded. Out of these 78, only 69 were assigned classifications by the cluster analysis program.

Two factors pose particular problems when attempting to devise a classification of SLI. First, when looking for patterns in language test data, great importance is attached to the profile of language test scores. However, discrepancies between verbal and non-verbal test scores, or between two language scores, are common in normal samples. We may illustrate this point with data from a representative sample of 179 English children who were assessed by Bishop & Butterworth (1979) in the course of studying the suitability of the revised version of the Wechsler Intelligence Scale for Children (WISC-R) (Wechsler 1974) for use in the UK. The average scaled score difference between the highest and lowest subtest was 7.6 (s.d. = 2.23), i.e. 2.5 standard deviations. The average scaled score difference between the lowest subtest and the mean of the remaining 9 subtests was 4.5 (s.d. = 1.65), i.e. 1.5 standard deviations. A flat subtest profile is very unusual; peaks and troughs in performance are commonplace and cannot be assumed to have clinical significance. The cause of such variation in normal samples is not well understood, but at least some of the unevenness in scores reflects error of measurement. Single subtests are less reliable than composites derived by summing several subtests, and differences between two subtests will be less reliable than the individual tests on which they are based. As well as error of measurement, there may be systematic differences in cognitive strengths and weaknesses of individuals that reflect specific experiences and preferences. For instance, the child who is fascinated by animals may do particularly well on a naming test that includes several animals. Factors such as these can produce peaks and troughs in test profiles that may be reliable without being very meaningful. The problem, then, is to identify systematic patterns of language impairment from a data set that is bound to contain a great deal of meaningless variation.

Second, where one does find evidence of distinct patterns, it does not necessarily follow that these correspond to different conditions. Lewis (1992) found that relatives of children with pre-school phonological disorders were affected by a wide range of speech, language and literacy problems; she concluded that these different disorders may be variable manifestations of a broad verbal trait deficit. There may also be

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

systematic variability in the pattern of test scores with age. For instance, Bishop & Edmundson (1987) found that a generalized expressive disorder involving speech and language often resolved into a pure phonological disorder as the child grew older. The message from these studies is clear: we should beware of assuming that different linguistic profiles correspond to aetiologically distinct disorders. The possibility remains that a single disorder might show very variable manifestation.

2. TWIN DATA AS A MEANS OF VALIDATING DEFINITIONS

The twin study method is typically used in the context of behaviour genetics to address questions about aetiology. Where both twins have a disorder, they are said to be concordant, whereas when only one twin is affected, the pair is discordant. By comparing concordance rates for monozygotic (MZ) twins, who are genetically identical, and dizygotic (DZ) twins, who share only half their genes on average, one can deduce whether genes are important in causing the disorder (see Plomin *et al.* 1989). Furthermore, the twin studies can also help us arrive at a clearer definition of a disorder. One way of tackling questions of diagnosis and classification is to compare different characterizations of the phenotype when attempting to fit a genetic model. Rutter *et al.* (1993) suggest an iterative approach where one starts with a hypothesized definition of disorder and uses genetic findings to refine the concept of the phenotype, then validating this in subsequent studies. In addition, concordant twins provide a natural form of replication study and enable one to study the question of whether different patterns of disorder simply reflect random error of measurement. If so, twins who are concordant for presence of disorder may be discordant for language profile.

In this paper, data from a twin study are used to consider the questions outlined above. A full account is reported by Bishop *et al.* (1994) and only brief details of sampling and methodology will be given here. Bishop *et al.* (1994) recruited 90 same-sex twin pairs where one or both twins had a specific speech or language impairment, using stringent diagnostic criteria based on definitions in the Diagnostic and Statistical Manual of the American Psychiatric Association (DSM-III-R) (APA 1987). The majority of twins (90%) were aged between 7 and 10 years; the remainder were teenagers or young adults. The diagnosis of SLI was made on the basis of significantly impaired performance (scaled score of 80 or below) on any one of four language tests, with a substantial discrepancy (20 points) from non-verbal IQ, as measured by Raven's Matrices (Raven *et al.* 1986). The tests were selected to be well standardized and sensitive to different domains of language function. They included two tests of receptive language (Test for Reception of Grammar (TROG) (Bishop 1983) and WISC-R Comprehension (Wechsler 1974)), a test of sentence repetition (Semel *et al.* 1980), and a word-finding test formed by combining items from a children's and adult's picture-naming scale (Renfrew 1980; McKenna & Warrington 1983). The proportion of pairs concordant for having any DSM-III-R diagnosis (i.e. speech and/or language disorder) was 54% in MZ twins (34 out of 63 pairs) and 30% in DZ twins (8 out of 27 pairs).

(*a*) *Specific and non-specific language impairment*

Although concordance was significantly higher for MZ, than for DZ twins, it was well below 100 per cent. However, when Bishop *et al.* (1994) considered the discordant 'unaffected' MZ twins, it was clear that few of them could be regarded as having normal language development. As can be seen from figure 1, 'unaffected' MZ twins frequently obtained low scores on language tests and failed to be included as affected cases simply because the discrepancy between verbal and non-verbal ability was not large enough to meet stringent criteria for SLI. The 'unaffected' MZ twins did not differ significantly from their affected co-twins on two of the four language measures (TROG and word finding), but they did differ in terms of the Matrices non-verbal IQ, obtaining significantly *lower* scores than the affected twin. Figure 1 also shows the comparable

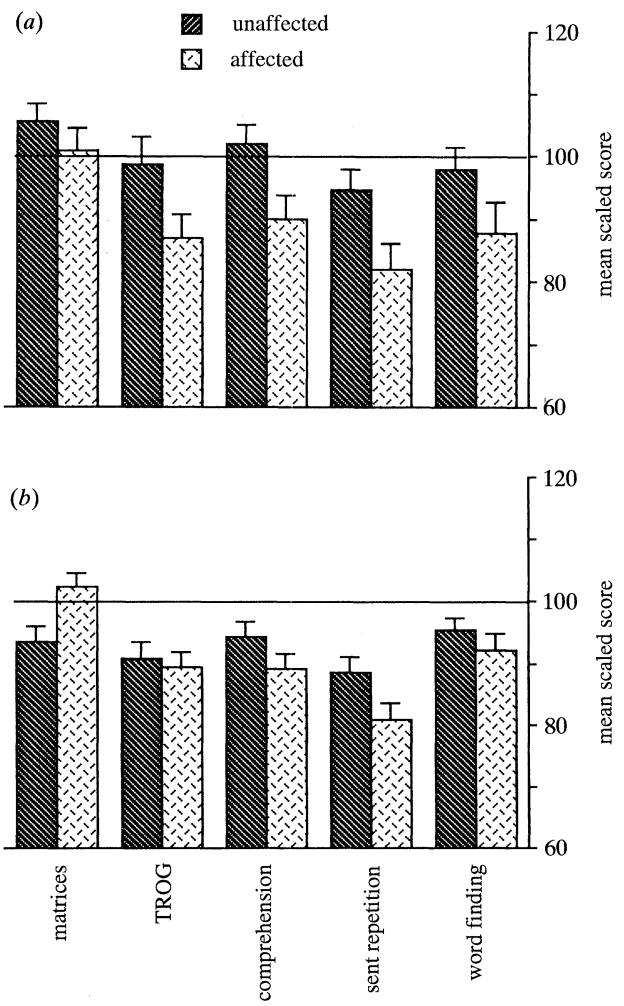

Figure 1. Mean non-verbal and language test scores for (*a*) DZ and (*b*) MZ twins discordant for SLI.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

data for discordant DZ twins, where the pattern was far more in line with what one would expect if the unaffected co-twins were linguistically and cognitively normal; language scores for unaffected individuals were much closer to the population mean and significantly different from affected co-twins for all four language measures, whereas non-verbal IQ was slightly above average and did not differ from that of affected individuals.

If we adopt a laxer diagnostic criterion for language disorder that includes all children who obtain any language test score below 81 and who are not mentally handicapped (i.e. IQ of 70 or above), then the concordance for MZ twins rises to 70% (with DZ concordance of 41%). Furthermore, the majority of the remaining 'unaffected' MZ co-twins had received speech therapy for a period of two years or more after the age of four years. Including these cases as 'affected' brings the MZ concordance up to 89%, compared with DZ concordance of 48%. These results lead one to draw a conclusion very similar to that stated by Stanovich (1994) regarding the dyslexia literature: namely, that there is no fundamental difference between children with language impairments who have a large discrepancy between IQ and verbal functioning, and those who do not.

The concern is that, if we are to relax discrepancy criteria, we may end up with an over-inclusive definition that will select all children with borderline intelligence that has no known cause. Only empirical studies can determine how serious a problem this is. We know far less about the causes of borderline intelligence than we do about mental handicap: it could be that the majority of cases are aetiologically similar to children with SLI, or it could be that there is a host of different causal factors that can lead to the final common path of low verbal ability. If the latter is the case, then by abandoning the IQ–language discrepancy criterion we will identify a much higher proportion of cases of genetically based developmental language disorder but only at the cost of including a great many false positive cases who have other aetiologies. It may prove to be the case that we need qualitative indicators of communicative function, rather than language test scores, in order to discriminate aetiologically distinct groups of children with low verbal abilities.

Clearly, there are no simple answers to the question of definition. In our current state of knowledge, research studies may be best advised to continue to use discrepancy criteria, simply to avoid the possibility of selecting a heterogeneous mix of children with diverse aetiologies. However, it must be recognized that, in doing so, they are drawing an artificial distinction; many children with the same aetiology have equally severe language difficulties without a large IQ–language discrepancy. A diagnostic term which does not include the word 'specific', such as 'developmental language disorder' seems preferable to describe the condition. In terms of practical policy, there seems little justification in continuing to place heavy reliance on IQ–language discrepancies in determining who should receive extra help at school. Such decisions should be based more on whether or not the child has communicative difficulties that impair daily life and educational progress, rather than on whether they have a large difference between two test scores (Fletcher 1992).

(*b*) *Concordance for subgroup of language impairment*

When one turns to look at specific aspects of language difficulty, one is greeted by a bewildering array of symptoms. Some children are poor on all verbal tests; others have much more specific impairments with particular aspects of expression or comprehension. Concordant twins provide a useful means of testing whether patterns of language impairment reflect stable individual characteristics, or whether they are largely the result of error of measurement. The crucial question is how far there is consistency between concordant twins in terms of the pattern of impairment shown on language tests.

This analysis considered data from 54 twin pairs only, who were concordant for language disorder; these included those meeting strict criteria for SLI or developmental articulation disorder, and those who have a less specific developmental language disorder, with at least one language scaled score of 80 or below, and non-verbal IQ above 70. MZ and DZ twins are treated together for this analysis; the question here is not whether genetic similarity is related to phenotypic similarity, but rather whether the pattern of disorder reflects a stable behavioural profile or whether it merely arises through random variation.

A previous analysis of this kind had been conducted by Bishop *et al.* (1994) to look at stability of diagnostic category across twin pairs using the diagnostic groupings of DSM-III-R (expressive language disorder, receptive language disorder and articulation disorder). However, it was clear that these groupings did not correspond to any natural categories; many children met criteria for two or three of the subgroups and there were many instances where both twins were affected but one met criteria for receptive language disorder and the co-twin did not. However, if one ignored the child's receptive status, there was excellent agreement between twins in terms of whether or not an expressive language disorder was accompanied by articulation disorder.

The tests used in the twin study had been selected with the aim of differentiating between different types of receptive and expressive problem, so it seemed worth while extending the study of twin similarity by looking at a more detailed classification system that had evolved through clinical experience. Wilson & Risucci's (1986) diagnostic framework was taken as the starting point. These authors related clinically derived categories to test scores covering nine major domains of neuropsychological functioning. Their test protocol follows a branching procedure, whereby an initial test in a given domain screens for general problems in that area, allowing further follow-up tests if the child appears to have difficulties. Thus information is potentially available

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 1. *Operational criteria for assigning individuals with SLI to diagnostic subgroups*

construct:	–	AC3	AM1	RS	VM3
test:	TROG	comp	repetition	word find	vis mem
subgroup					
W: global	–	–	–	–	–
W′	–	–	–	–	+
C: receptive 1	+	–	·	·	+
HB: receptive 2–3	+	–	·	·	–
P: receptive 4	–	–	–	+	–
F: aud mem/retrieval	+	+	–	–	+
F′:	+	+	–	–	–
A: expressive 1	·	–	–	+	+
S: expressive 2	+	+	+	–	·
G: short-term memory	+	+	–	+	·
G′	–	+	–	+	·
N: no deficits on test[1]	+	+	+	+	+

Symbols: +, scaled score of 85 or above; –, scaled score below 85; ·, indeterminate (score can take any value).
[1] In this sample, the only children in this group would be those with pure articulation disorder.

on a very wide range of measures. Depending on the pattern of performance, Wilson & Risucci assigned children to one of 11 language disorder subgroups; however, because there were very small numbers in some subgroups, these were collapsed to five subgroups in their validation study.

Data from our twin study were available for only four language measures, but three of these corresponded fairly closely to measures used by Wilson & Risucci. Thus, in relation to their model, WISC-R Comprehension can be regarded as an index of Auditory Cognition (equivalent to their measure AC3), Word Finding can be regarded as an index of Retrieval (RET), and Repeating Sentences as an index of Auditory Memory (equivalent to their AM1). Their battery contained no measure analogous to TROG, which is a multiple-choice test specifically designed to evaluate comprehension of grammatically complex sentences when vocabulary is kept simple. This type of test, which is known to reveal grammatical difficulties in children with predominantly expressive problems (Bishop 1979), is very different from the semantic comprehension measures used by Wilson & Risucci. They also used a range of tests designed to test visual aspects of cognitive processing. We had no tests directly analogous to their battery, but we had included the Immediate Visual Memory subtest from the British Ability Scales (Elliot *et al.* 1979). The test is not a pure measure of visual memory; the child is shown a card with 20 pictured items which are named by the tester, and is asked for verbal recall when the card is removed. However, there is some overlap with the Visual Memory construct of Wilson & Risucci, and this test was used here as a proxy for their VM3 test.

On the basis of the descriptions and criteria for categories given by Wilson & Risucci, operational definitions for each subgroup were specified, as shown in table 1. Cases were assigned a code for each measure as impaired or unimpaired simply on the basis of whether their score fell above or below 85. The pattern of codes was directly translated into categories according to the operational definitions in table 1. Because the Wilson & Risucci classification depends on a much more complex database than was available here, it was not possible to derive criteria for

Table 2. *Agreement in Wilson–Risucci subgroup between 54 pairs of twins concordant for developmental language disorder*

twin B category	twin A category												
	W	W′	C	HB	P	F	F′	A	S	G	G′	N	none
W	4	3	·	·	·	1	1	·	·	·	·	·	1
W′		·	2	·	1	·	·	·	·	·	·	·	·
C			·	·	1	·	·	2	·	1	·	1	1
HB				1	·	·	·	·	·	·	·	·	·
P					·	·	1	1	·	1	2	·	1
F						4	2	2	·	1	·	·	·
F′							·	·	·	·	·	·	2
A								·	·	·	1	·	·
S									1	·	·	·	·
G										5	2	·	·
G′											3	2	·
N*												1	1
none													1

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES

Table 3. *Mean non-verbal IQ and age for twins according to language disorder subgroup*

	W + W′ global ($n = 20$)	F + F′ repetition + word finding ($n = 20$)	G + G′ repetition only ($n = 28$)	significance[1]
age	7.88 (1.01)	11.49 (5.38)	8.11 (1.48)	$F(2, 65) = 8.9$ $p < 0.01$ (W = G) < F
Matrices IQ	90.2 (11.03)	98.3 (12.54)	101.1 (12.25)	$F(2, 65) = 5.0$ $p < 0.01$ W < G

[1] One-way analysis of variance; Scheffé test for specific contrasts.

every subgroup, and some categorical assignments had to be made on the basis of information that was considerably more limited than that specified by Wilson & Risucci. Subgroup H was not discriminable from subgroup B on the basis of available data, and no tests were available for diagnosing auditory processing disorders. Wilson & Risucci usually had several indices of each area of functioning and so could specify that a child showed an impairment if two out of three indices were low. Because only one index per area was available for this study, it was necessary to specify impairment on the basis of that particular test. Where one category was a logical subset of another (for example, subgroups C and A), the more under-specified category was assigned only if the more specified category did not apply. Where a pattern was common enough in the data set to merit a category of its own, but did not exactly correspond to one of the Wilson & Risucci subgroups, a related subgroup was created, i.e. W′, G′ and F′. The number of possible patterns on five tests is 32; ten of these do not correspond to any subgroup in the Wilson & Risucci scheme, and twins showing these patterns were assigned no code.

Table 2 shows the conjoint classification of language disorder subgroups for each pair of twins. The extent of agreement between twins, although far from perfect, is substantially above chance levels, with 20 out of 54 pairs sharing the same diagnosis (direct maximum likelihood estimation for block triangular tables (Bishop *et al.* 1975): $\chi^2 = 97.91$, d.f. = 66, $p < 0.01$). If we group W with W′, G with G′ and F with F′ agreement is even stronger, rising to 52% ($\chi^2 = 50.19$, d.f. = 36, $p < 0.01$).

What can we conclude from this analysis? At least some of the subgroups identified by Wilson & Risucci were relatively frequent in this sample and yielded agreement between twins that was well above chance levels, demonstrating that these patterns of language score are reasonably stable phenomena and not merely reflective of error of measurement. Some subgroups, notably the receptive categories and category A (expressive 1) did not seem very coherent, although it must be recognized that assignment of cases to subgroups was based on considerably less detailed information than had been used by Wilson & Risucci.

However, good agreement between twins for subgroup of disorder does not necessarily mean that the subgroups correspond to different conditions with different causes. It could be that the same underlying disorder leads to different linguistic profiles depending on other factors such as age and cognitive abilities. Table 3 shows data compatible with that interpretation from twin pairs who are concordant for the three most common subgroups. Subgroups W (global) and G (repetition deficit) do not differ in age, but do differ in non-verbal IQ. Those in subgroup F (poor repetition and word finding, with normal comprehension) are significantly older than twins in the other two subgroups. These systematic differences between subgroups in terms of age and IQ suggest that it would be premature to dismiss the null hypothesis of a single core inherited language deficit whose manifestation depends on level of other cognitive abilities and on stage of development.

We must remember that a classification is only as good as the data from which it is derived. Clinically based classifications (such as those of Bishop & Rosenbloom 1987; Rapin & Allen 1983) rely on language features that would not be detected with standardized language tests. For example, key features of 'semantic pragmatic disorder' are verbosity, tangential answers to questions, and problems in understanding discourse.

The twin data demonstrate that some patterns of language impairment are reliable. The challenge for the future is to discover how far these patterns correspond to different disorders. Two research strategies are required. We need longitudinal studies to evaluate how language phenotypes change as the child develops, and we need to develop new assessment procedures that provide valid and reliable measures of those behaviours that appear, on clinical grounds, to characterize distinct subgroups of language disorder.

REFERENCES

American Psychiatric Association 1987 *Diagnostic and statistical manual*, 3rd edn, revised (DSM-III-R). Washington, D.C.: American Psychiatric Association.

Aram, D.M., Morris, R. & Hall, N.E. 1992 The validity of discrepancy criteria for identifying children with

developmental language disorders. *J. Learn. Disabil.* **25**, 549–554.

Aram, D.M. & Nation, J.E. 1975 Patterns of language behaviour in children with development language disorders. *J. Speech Hearing Disord.* **18**, 229–241.

Beitchman, J.H., Hood, J., Rochon, J., Peterson, M., Mantini, T. & Majumdar, S. 1989 Empirical classification of speech/language impairment in children I. Identification of speech/language categories. *J. Am. Acad. Child Adolesc. Psychiatr.* **28**, 112–117.

Bishop, D.V.M. 1979 Comprehension in developmental language disorders. *Devel. Med. Child Neurol.* **21**, 225–238.

Bishop, D.V.M. 1983 *The test for reception of grammar.* (Published by the author and available from Age and Cognitive Performance Research Centre, University of Manchester.)

Bishop, D.V.M. & Butterworth, G.E. 1979 A longitudinal study using the WPPSI and WISC-R with an English sample. *Br. J. educ. Psychol.* **49**, 156–168.

Bishop, D.V.M. & Edmundson, A. 1987 Language-impaired four-year-olds: distinguishing transient from persistent impairment. *J. Speech Hearing Disord.* **52**, 156–173.

Bishop, D.V.M., North, T. & Donlan, C. 1994 Genetic basis of specific language impairment: evidence from a twin study. *Devel. Med. Child Neurol.* (In the press.)

Bishop, D.V.M. & Rosenbloom, L. 1987 Classification of childhood language disorders. In *Language development and disorders* (ed. W. Yule & M. Rutter) (*Clinics in Developmental Medicine* (double issue) nos. *101–2*), pp. 16–41. London: MacKeith Press.

Bishop, Y.M., Fienberg, S.E. & Holland, P.W. 1975 *Discrete multivariate analysis.* Cambridge, Massachusetts: MIT Press.

Cole, K.N., Dale, P.S. & Mills, P.E. 1992 Stability of the intelligence quotient – language quotient relation: is discrepancy modeling based on a myth? *Am. J. mental Retard.* **97**, 131–143.

Elliot, C.D., Murray, D.J. & Pearson, L.S. 1979 *British ability scales.* Windsor: NFER-Nelson.

Fletcher, J.M. 1992 The validity of distinguishing children with language and learning disabilities according to discrepancies with IQ. *J. Learning Disabil.* **25**, 546–548.

Fletcher, J.M., Francis, D.J., Rourke, B.P., Shaywitz, S.E. & Shaywitz, B.A. 1992 The validity of discrepancy-based definitions of reading disabilities. *J. Learning Disabil.* **25**, 555–561.

Lewis, B.A. 1992 Pedigree analysis of children with phonology disorders. *J. Learning Disabil.* **25**, 586–597.

McKenna, P. & Warrington, E.K. 1983 *Graded naming test* (GNT). Windsor: NFER-Nelson.

Pennington, B.F., Gilger, J.W., Olson, R.K. & DeFries, J.C. 1992 The external validity of age- versus IQ-discrepancy definitions of reading disability: lessons from a twin study. *J. Learning Disabil.* **25**, 562–573.

Plomin, R., DeFries, J.C. & McClearn, G.E. 1989 *Behavioral genetics: a primer*, 2nd edn. New York: W. H. Freeman.

Rapin, I. & Allen, D. 1983 Developmental language disorders: nosologic considerations. In *Neuropsychology of language, reading and spelling* (ed. U. Kirk), pp. 155–184. New York: Academic Press.

Raven, J.C., Court, J.H. & Raven, J. 1986 *Raven's progressive matrices and Raven's coloured matrices.* London: H. K. Lewis.

Renfrew, C. 1980 *Word-finding vocabulary scale.* Oxford: C. E. Renfrew.

Rutter, M., Simonoff, E. & Silberg, J. 1993 How informative are twin studies of child psychopathology? In *Twins as tools of behavioral genetics* (ed. T. J. Bouchard & P. Propping), pp. 179–194. Chichester: Wiley.

Semel, E.M., Wiig, E.H. & Secord, W. 1980 *Clinical evaluation of language fundamentals – revised.* San Antonio, Texas: Psychological Corporation.

Stanovich, K.E. 1994 Annotation: does dyslexia exist? *J. Child Psychol. Psychiatr.* **35**, 579–595.

Stark, R.E. & Tallal, P. 1981 Selection of children with specific language deficits. *J. Speech Hearing Disord.* **46**, 114–122.

Wechsler, D. 1974 *Wechsler intelligence scale for children – revised.* New York: Psychological Corporation.

Wilson, B.C. & Risucci, D.A. 1986 A model for clinical-quantitative classification generation 1: application to language disordered preschool children. *Brain Lang.* **27**, 281–310.

Wolfus, B., Moscovitch, M. & Kinsbourne, M. 1980 Subgroups of developmental language impairment. *Brain Lang.* **10**, 152–171.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES