BRAIN AND LANGUAGE **41,** 43–51 (1991)

Data, Statistics, and Theory: A Comment on Bates, McDonald, MacWhinney, and Applebaum's "A Maximum Likelihood Procedure for the Analysis of Group and Individual Data in Aphasia Research"

ALFONSO CARAMAZZA

The Johns Hopkins University

Although there certainly is much room for disagreement on specific cases, all scientists will undoubtedly agree that not all putatively theoretically relevant observations are in fact so, and, more importantly, that not all forms of inferences from observation to theory (or vice versa) are equally acceptable. Indeed, a significant part of scientific practice in all sciences is concerned with distinguishing between theoretically useful and nonuseful observations, and between valid and invalid arguments in reasoning from theory to data, and vice versa. From time to time, when someone attempts to distinguish explicitly between valid and invalid methods in a scientific domain, these concerns become the focus of heated, public controversy. This seems to have been the case with my colleagues' and my effort to define the range of valid inferences that may be drawn from the performance of brain-damaged subjects for hypotheses about normal cognition and the brain (Badecker & Caramazza, 1985; Caramazza, 1984, 1986; Caramazza & Badecker, 1989; Caramazza & McCloskey, 1988; McCloskey & Caramazza, 1988).

In several papers, I have argued that in neuropsychological research only single-patient studies provide the sorts of data that would allow valid inferences about normal cognition. The claim that patient-group studies based on the a priori classification of patients (e.g., Broca's aphasia, deep dyslexia, neglect, agrammatism, and so forth) do not allow meaningful

Several friends (B.B., R.B., A.H., B.R.) provided helpful comments on an earlier version of this paper. Address all correspondence and reprint requests to Alfonso Caramazza, Department of Cognitive Science, The Johns Hopkins University, Baltimore, MD 21218.

0093-934X/91 $3.00

conclusions about cognition or the brain has not gone unchallenged, however (Bub & Bub, 1988; Caplan, 1988; Newcombe & Marshall, 1988; Zurif, Gardner, & Brownell, 1989; but see Caramazza & Badecker, 1989; McCloskey & Caramazza, 1988, for responses to these papers). Bates, McDonald, MacWhinney, and Applebaum (1991) also take issue with the arguments and conclusions my colleagues and I have presented on this matter. In their paper, Bates et al. begin by summarizing what they take to be the principal arguments against patient-group studies, followed by a discussion of what they consider to be the major limitations of single-patient methodology. They conclude by proposing a "compromise" methodology, the maximum likelihood estimation (MLE) procedure, which can be applied to individual patients in groups, as well as to whole groups, thus preserving a central role in cognitive neuropsychological research for such categories as Broca's aphasia, Wernicke's aphasia, agrammatism, and so forth.

In this discussion of the Bates et al. paper, I briefly assess their proposed use of the MLE procedure in neuropsychological research. Because the authors do not restrict themselves to presenting an abstract argument for the procedure, but actually discuss in detail an example from their own research, it is possible to evaluate the level of theoretical sophistication that can be reached with this procedure. Before doing so, however, I will very briefly argue that they have misunderstood the arguments against the patient-group methodology. I will go on to show that their criticisms of the single-patient methodology are mostly irrelevant and those that are not irrelevant are not the sort of criticisms that would undermine the valid use of this methodology.

ON THE ARGUMENTS AGAINST PATIENT-GROUP STUDIES

In various publications I have presented detailed arguments against the use of the patient-group study methodology as the basis for drawing inferences about the structure of normal cognitive processes. These detailed arguments will not be repeated here (the interested reader may consult Caramazza, 1986; Caramazza & Badecker, 1989; McCloskey & Caramazza, 1988). Briefly, the arguments principally turn on the observation that patients are experiments of nature in which brain damage has resulted in an unspecified transformation of the normal cognitive system. An implication of this observation is that interpretation of a patient's performance requires identification of the specific manipulations introduced by nature in any particular patient. Thus, the basic unit of experimental analysis must be the individual patient. In this view, "The only nonarbitrary classification of patients that is possible is a posteriori; that is, on the basis of the theoretically relevant performance which allows the identification of a functional lesion in a cognitive system. But the latter is equivalent to saying that patient classification cannot play any

significant role independently of the single-patient research projects that are required to determine that each of the patients in question has the appropriate functional lesion for a posteriori classification" (Caramazza & Badecker, 1989, p. 260). Consequently, only single-patient studies can allow meaningful interpretation of impaired performance in informing theories of normal cognitive function.

Bates et al. conveniently overlook these arguments and instead summarize my position as a claim about "the importance of minority patterns" and the "nonrepresentativeness of the mean." This reformulation of my analyses of the *logical insufficiency* of patient-group studies is designed to reduce the arguments against patient-group methodology to a claim about statistical problems with this methodology. If this reduction were successful, then, one could seek statistical solutions for the statistical problems, as Bates et al. have attempted to do in their paper. However, until Bates et al. offer valid counterarguments to those I have presented, their assertions are without force and should be treated accordingly.

ON THE ARGUMENTS AGAINST SINGLE-PATIENT METHODOLOGY

One objection raised by Bates et al. concerns the problem of idiosyncratic findings. As discussed in Caramazza (1986) this is indeed a problem: it is entirely possible that various reports will represent anomalous findings either because of premorbid peculiarities in the cases in question (e.g., a case with a developmental disability superimposed on an acquired disorder) or because of willful or unwitting errors in the reported data. The solution to this problem, however, is not found by turning to the patient-group methodology. Since this methodology cannot allow meaningful theoretical conclusions it can hardly be offered as the solution to a real problem with the single-patient methodology[1] (see also McCloskey & Caramazza, 1988, for further discussion). In Caramazza (1986) I discuss the problem of idiosyncratic findings and I consider possible solutions.

In discussing the problem of idiosyncratic findings, Bates et al. give the example of category-specific semantic deficits to illustrate the pitfalls of single-patient methodology. They seem to suggest that category-specific impairments should be treated as idiosyncratic results. What is the basis for this claim? Are such results to be treated as idiosyncratic simply because they are relatively rare?[2] Or is the reason for relegating category-specific impairments to the bin of "unusual results" merely because these

[1] Bates et al. seem to think that all one has to do to solve a problem is to combine it with another problem so that they will magically cancel each other out: "If group and case studies are compared side by side, there are sound reasons to believe that their methodological flaws will cancel out to yield a deeper truth." However, Bates et al. never make explicit the "sound reasons" for this remarkable insight.

[2] See Marshall and Newcombe (1984) for a discussion of the fallacy of treating *rare* as equivalent to *atypical*.

results do not conform with the theoretical preconceptions of Bates et al. about the range of possible consequences from brain damage? Bates et al. do not offer a justification for their claim—they simply assert it. What are we to conclude from this practice? Perhaps Bates et al. are not familiar with the recent, extensive literature which has documented consistent and richly articulated patterns of category-specific effects (e.g., Basso, Capitani, & Laiacona, 1988; Baxter & Warrington, 1985; Goodglass & Budin, 1988; Hart, Berndt, & Caramazza, 1985; Hillis & Caramazza, in press; Miceli, Silveri, Nocentini, & Caramazza, 1988; Miceli, Silveri, Villa, & Caramazza, 1984; McCarthy & Warrington, 1985, 1989; Sartori & Job, 1988; Semenza & Zettin, 1989; Silveri & Gainotti, 1988; Warrington, 1975, 1981; Warrington & McCarthy, 1983, 1987, 1989; Warrington & Shallice, 1984; Zingeser & Berndt, 1988, 1990; see Berndt, 1988; Shallice, 1988, for reviews).

Another issue raised by Bates et al. in the context of their discussion of idiosyncratic findings concerns the problem of "adaptation." Presumably this issue is raised to make the point that different patients with the same underlying deficit may "adapt" differently in response to the deficit, leading to theoretically uninteresting variation in performance. The relevance of this issue is not obvious. Are we to take this observation as indicating that single-patient studies are flawed if patients adapt differently to the same underlying deficit? What is the argument for such a contention? None is offered. Suppose that patients with the same underlying deficit did adapt differently (whatever this might mean). What are we to make of this? Does it mean that we will be unable to uncover the real underlying causes of a disorder? Bates et al. seem to imply this when they say that there may not be a "transparent" relation between impaired performance and cognitive mechanisms.[3] But, if this were so, then, the problem would not concern specifically single-patient methodology—it would undermine the whole enterprise of neuropsychology: In that case, impaired performance would, *by hypothesis,* not be usable as a basis for inferring the structure of normal cognitive processes. Perhaps, Bates et al. mean something else. Perhaps they mean that adaptation is "more" problematic for single-patient studies than for patient-group studies. If this is what they meant, what are the arguments for the claim? None are offered.[4]

[3] Bates et al. have not understood what I mean by the "transparency assumption" (Caramazza, 1986). Briefly, this assumption simply asserts that the modifications introduced to the cognitive system are of a sort that would not impede the use of impaired performance to infer the structure of normal cognitive processes, no matter how complex or tortuous the link may be. In my use of the term there is no implication that inferences from pathology to normal systems will be easy or immediate.

[4] In discussing the problem of idiosyncratic findings, Bates et al. take issue with arguments I have presented against claims about "adaptation." They claim that I have argued that "neuropsychology would be impossible, and nothing could be learned, if it were true that

The second set of "objections" raised against the single-patient methodology concerns "measurement problems inherent in the study of individuals." Issues of measurement are a standard concern in all experimental situations. There are many excellent treatments of the statistical precautions to be followed in single-patient research. Bates et al. add nothing new to this familiar literature except some confusion. Thus, they discuss the problem of the shape of the underlying distribution in interpreting interactions as if this concerned specifically single-patient methodology. This is simply an error (see Loftus, 1978; Pachella, 1974). The other confusion they add concerns the problem of "base rates and expected values." Once again the problem does not concern specifically single-patient research; and, it makes no sense to suppose that averaging the performance of patients with potentially different underlying deficits could magically solve the problems inherent in this area.[5] In short, Bates et al. have not presented any serious criticisms specifically concerning the single-patient methodology that my colleagues and I have not already addressed in some detail.

AN EXAMPLE IS WORTH A THOUSAND ARGUMENTS

In their paper, Bates et al. not only present a critical analysis of single-patient methodology, they also discuss a methodology that is supposed to be ideally suited to the study of subjects with acquired cognitive disorders. As already mentioned, this discussion is accompanied by a specific application to the authors' own research: the application of the maximum likelihood (MLE) procedure to the analysis of sentence comprehension in normal and aphasic speakers of English, Italian, and German. This gives us the opportunity to assess concretely the merits of the procedure.

Bates et al. reanalyze results from a study originally published as Bates,

---

patients are capable of developing or modifying symptoms through some process of adaptation." I have never made such claims. Whatever I could be "accused" of, it could not be for claiming things such as: "patients are capable of . . . modifying symptoms through adaptation." It is patent nonsense to say that "patients modify symptoms." Symptoms are the consequences of damage to a cognitive system. What I have argued is that nothing could be learned from pathology were it to result in the creation of new processing procedures since in that case the observed performance would not concern the normal cognitive system (see Caramazza, 1984, 1986).

[5] In discussing the issue of base rates and expected values, Bates et al. illustrate problems in this area by considering a study by Miceli and Caramazza (1988). They point out that Dell's study (1988) has found a larger proportion of inflectional than derivational errors in corpora of normal subjects' speech errors. They use this result to argue that perhaps inflectional morphology is less "robust" and, therefore, more susceptible to breakdown than derivational morphology. It is not clear to me what Bates et al. mean by "robust." The reason that there are more inflectional than derivational errors in the corpora of speech errors produced by normal subjects could mean that different processes are engaged in processing the two kinds of morphology, consistent with our results. The appeal by Bates et al. to an undefined "robustness" is simply unhelpful.

Friederici, and Wulfeck (1987). In this study they compared the performance of "25 Broca's aphasics (8 English, 7 German, and 10 Italian), 15 Wernicke's aphasics (9 German and 6 Italian), and a series of different control groups" (Bates et al., 1991). The subjects were asked to act out word strings consisting of the following sequences of nouns (N) and verbs (V): NVN, NNV, and VNN.[6] The authors hypothesized that word order, morphology, and semantic information would be differentially important in determining performance in the different languages. Thus, for example, since Italian is a morphologically rich language with relatively free word order structure, it was hypothesized that subjects would be more sensitive to morphological cues than to word order cues. Assuming that word order, morphological, and semantic cues are differentially important across English, German, and Italian, it is possible to specify the ordering of importance of these linguistic cues for the three languages.

Bates et al. present an impressive set of analyses leading to a remarkable set of conclusions. In this paper, the authors focused on the performance of Broca's aphasics. The first set of analyses considered group performance. Here the results are unambiguous:

> English patients fit the English equation with an RMSD of .2393 (with individual fits ranging from .3072 to .4477). These figures are of course much higher (i.e., much worse) than the range of individual fits observed with English normals. At the group level, German patients fit the German model at a group value of .3136, with individual figures ranging from .3107 to a .4840. Finally, Italian aphasics fit the Italian equation at .2463, but individual statistics range from .1517 to a high .5243.

Bates et al. go on to present individual subject fits for different language models:

> It is equally clear, however, that the best fit is still provided by the patients' own native language. At the group level, the RMSD of .2393 provided by the English equation is considerably better than the RMSDs for German (.4377) or Italian (.4447). Furthermore, the English model also provided the best fit for all eight individual Broca's aphasics. In German, the group RMSD of .3136 compares favorably with the models for English (.4977) or Italian (.4681). The German model also provided the best fit of the three for 7 of the 8 individual German patients. The Italian group data is best explained by the Italian equation (.2463), but the German equation is only slightly worse (.2532) and the English equation provides a surprisingly close fit (.3287). The spread of variation observed in the Italian group is also illustrated by the fact that the Italian equation provides the best fit for only 4 of the 10 individual Italian patients; 2 are better fit by the German model, and 4 individual Italians are actually fit best by the English equation.

[6] It should be noted that some of these sequences are ungrammatical in the languages tested. The relevance of patients' performance on these tasks for normal comprehension processes is not entirely obvious.

These results led Bates et al. to note that "the native language equation provides a better fit [than the nonnative language equations] for 18 of 25 patients . . . ($p < .01$)." And, thus, they conclude that "the English model gives a better fit for English patients, the German model gives a better fit for German patients, and the Italian model provides a better fit for Italian patients, despite individual deviations."

It is clear from the above that Bates et al. have provided convincing empirical evidence that it is not only extremely unlikely that a premorbidly monolingual English speaker would, as the result of brain damage, become an Italian or German speaking aphasic, but it is equally unlikely that premorbidly monolingual German speakers would become English or Italian speaking aphasics as the result of brain damage. The only discordant note in the analyses presented by Bates et al. is that there seems to be an indication that some premorbidly monolingual Italian speakers may, as the result of brain damage, become English or German aphasics. Although Bates et al. do not present an explanation for this puzzling wrinkle to their story, the overall results are, nonetheless, remarkable.

Bates et al. go on to provide equally striking analyses and results when they consider various hypotheses about the underlying nature of Broca's aphasia. Although they propose four possible hypotheses about the nature of agrammatism, only three of these are taken seriously: the *Closed-Class Loss hypothesis;* the *Closed-Class Access hypothesis;* and, the *Normalcy through Noise hypothesis*. The results are striking. Thus,

> At the individual level, Normalcy through Noise provides the best fit for all 8 individual English patients. This model also provides the best fit for 4 of the 10 Italian patients. Among the other six individuals in the Italian group, three are best fit by simply assuming that the patient is normal; with these patients we might conclude that the 50% noise level is much too strong, and that better fits would have been obtained with either Normalcy through Noise or Closed-Class Access if we had assumed a smaller overall deficit level (e.g., a 10–20% decrease in cue strength). The remaining three Italians are explained best by Closed-Class Access. The seven German Broca's aphasics break down as follows: four are best fit by Closed-Class Loss, two by Closed-Class Access, with a single Broca's aphasic (out of all 25 in the study) receiving his lowest RMSD score under the assumptions of Strong Agrammatism.

Although it is not clear whether Bates et al. would conclude from their results that the so-called Broca's aphasics included in their 1991 study and 1987 study have damage to different processing components or whether they prefer to assume that the patients do in fact have damage to the same underlying mechanisms but have "adapted" differently to the damage, this *subtle* distinction does not seem to be of particular concern to the authors. In their view, differences among patients in a pre-established category cannot be nearly as theoretically important as the original clinical classification itself and, thus, we should not be deterred from using

the patient-group methodology. After all, as they would have it, the differences among patients are likely to "cancel out" when we average their performance. And, in any case, what difference does it make if, for example, a patient has a deficit to morphological processing mechanisms and another to thematic mapping mechanisms (see Berndt, 1991, for a detailed discussion of the various patterns of deficits observed in so-called agrammatic Broca's aphasics). What seems to matter most to Bates et al. is that we preserve the crucially important categories that have been handed down to us—categories such as Broca's and Wernicke's aphasia[7]—even at the price of entertaining the possibility that brain damage could transform an Italian speaker into an English or a German aphasic.

REFERENCES

Badecker, W., & Caramazza, A. 1985. On considerations of method and theory governing the use of clinical categories in neurolinguistics and cognitive neuropsychology: The case against agrammatism. *Cognition,* **20,** 97–125.

Basso, A., Capitani, E., & Laiacona, M. 1988. Progressive language impairment without dementia: A case with isolated category-specific naming defect. *Journal of Neurology, Neurosurgery, and Psychiatry,* **51,** 1201–1207.

Bates, E., Friederici, A., & Wulfeck, B. 1987. Comprehension in aphasia: A cross-linguistic study. *Brain and Language,* **32,** 19–67.

Bates, E., McDonald, J., MacWhinney, B., & Applebaum, M. 1991. A maximum likelihood procedure for the analysis of group and individual data in aphasia research. *Brain and Language,* **40**(2), 231–265.

Baxter, D., & Warrington, E. K. 1985. Category-specific phonological dysgraphia. *Neuropsychologia,* **23,** 653–666.

Berndt, R. S. 1991. Sentence processing in aphasia. In M. Sarno (Ed.), *Acquired aphasias.* Orlando: Academic Press. 2nd ed.

Berndt, R. S. 1988. Category-specific deficits in aphasia. *Aphasiology,* **2,** 337–340.

Bub, D., & Bub, D. 1988. On methodology of single-case studies in neuropsychology. *Cognitive Neuropsychology,* **5**(5), 565–582.

Caplan, D. 1988. On the role of group studies in neuropsychological and pathopsychological research. *Cognitive Neuropsychology,* **5**(5), 535–548.

Caramazza, A. 1984. The logic of neuropsychological research and the problem of patient classification in aphasia. *Brain and Language,* **21,** 9–20.

Caramazza, A. 1986. On drawing inferences about the structure of normal cognitive processes from patterns of impaired performance: The case for single-patient studies. *Brain and Cognition,* **5,** 41–66.

Caramazza, A., & Badecker, W. 1989. Patient classification in neuropsychological research. *Brain and Cognition,* **10,** 256–295.

Caramazza, A., & McCloskey, M. 1988. The case for single-patient studies. *Cognitive Neuropsychology,* **5,** 517–528.

[7] The attentive reader of the Bates et al. paper may have wondered what it was precisely that those authors concluded about Broca's aphasia, agrammatism, or normal language processing. What, in other words, was learned about language or the brain from all the interminable number of numbers presented by the authors? More specifically, what was concluded about the category Broca's aphasia? Does it represent a theoretically coherent notion? What in their many numbers would support such a conclusion?

Goodglass, H., & Budin, C. 1988. Category and modality specific dissociations in word comprehension and concurrent phonological dyslexia. *Neuropsychologia,* **26,** 67–78.

Hart, J., Berndt, R., & Caramazza, A. 1985. Category-specific naming deficit following cerebral infarction. *Nature,* **316,** 338.

Hillis, A., & Caramazza, A. in press. Category specific naming and comprehension impairment: A double dissociation. *Brain.*

Loftus, G. R. 1978. On interpretation of interactions. *Memory and Cognition,* **6,** 312–319.

Marshall, J. C., & Newcombe, F. 1984. Putative problems and pure progress in neuropsychological single-case studies. *Journal of Clinical Neuropsychology,* **6,** 65–70.

McCarthy, R. A., & Warrington, E. K. 1985. Category specificity in an agrammatic patient: The relative impairment of verbal retrieval and comprehension. *Neuropsychologia,* **23,** 709–723.

McCarthy, R. A., & Warrington, E. K. 1989. Evidence for modality-specific meaning systems in the brain. *Nature,* **334,** 428–430.

McCloskey, M., & Caramazza, A. 1988. Theory and methodology in cognitive neuropsychology: A response to our critics. *Cognitive Neuropsychology,* **5,** 583–623.

Miceli, G., Silveri, M. C., Nocetini, U., & Caramazza, A. 1988. Patterns of dissociation in comprehension and production of nouns and verbs. *Aphasiology,* **2,** 351–358.

Miceli, G., Silveri, M. C., Villa, G., & Caramazza, A. 1984. On the basis of agrammatic's difficulty in producing main verbs. *Cortex,* **20,** 217–220.

Newcombe, F., & Marshall, J. C. 1988. Idealization meets psychometrics: The case for the right groups and the right individuals. *Cognitive Neuropsychology,* **5,** 549–564.

Pachella, R. G. 1974. The interpretation of reaction time in information processing research. In B. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition.* Hillsdale, NJ: LEA.

Sartori, G., & Job, R. 1988. The oyster with four legs: A neuropsychological study on the interaction of visual and semantic information. *Cognitive Neuropsychology,* **5,** 105–132.

Semenza, C., & Zettin, M. 1989. Evidence from aphasia for the role of proper names as pure referring expressions. *Nature,* **342,** 678–679.

Shallice, T. 1988. *From neuropsychology to mental structure.* Cambridge: Cambridge Univ. Press.

Silveri, M. C., & Gainotti, G. 1988. Interaction between visual and language in category-specific semantic impairment. *Cognitive Neuropsychology,* **5,** 677–709.

Warrington, E. K. 1975. Neuropsychological studies of verbal semantic systems. *Quarterly Journal of Experimental Psychology,* **27,** 635–657.

Warrington, E. K. 1981. Neuropsychological studies of verbal semantic systems. *Philosophical Transaction of the Royal Society of London B,* **295,** 411–423.

Warrington, E. K., & McCarthy, R. A. 1983. Category-specific access dysphasia. *Brain,* **106,** 859–878.

Warrington, E. K., & McCarthy, R. A. 1987. Categories of knowledge: Further fractionations and an attempted explanation. *Brain,* **110,** 1273–1296.

Warrington, E. K., & McCarthy, R. A. 1988. Evidence of modality-specific meaning systems in the brain. *Nature,* **334,** 428–430.

Warrington, E. K., & Shallice, T. (1984). Category specific semantic impairments. *Brain,* **107,** 829–853.

Zingeser, L. B., & Berndt, R. S. 1988. Grammatical class and context effects in a case of pure anomia: Implications for models of language production. *Cognitive Neuropsychology,* **5,** 473–516.

Zingeser, L. B., & Berndt, R. S. 1990. Retrieval of nouns and verbs in agrammatism and anomia. *Brain and Language,* **39,** 14–32.

Zurif, E., Gardner, H., & Brownell, H. 1989. The case against the case against group studies. *Brain and Cognition,* **10,** 237–255.