Quantifying Dissociations in Neuropsychological Research

Elizabeth Bates[1], Mark Appelbaum[2], Jesus Salcedo[3], Ayse Pinar Saygin[4],
and Luigi Pizzamiglio[5,6]

[1]Center for Research in Language, University of California, San Diego, La Jolla, CA, USA,
[2]Department of Psychology, University of California, San Diego, La Jolla, CA, USA, [3]Fordham University,
Bronx, NY, USA, [4]Department of Cognitive Science, University of California, San Diego,
La Jolla, CA, USA, [5]IRCCS Fondazione Santa Lucia, Centro Ricerche Neuropsicologia, Italy, and
[6]Department of Psychology, Università degli Studi di Roma "La Sapienza," Rome, Italy

ABSTRACT

Double dissociations play an important role in neuropsychology, but they are often identified through subjective estimates of "high" versus "low" performance, without considering the probability that such an outcome might have occurred by chance. To determine whether two measures "come apart" in an interesting way in brain-damaged patients, it is important to know the degree to which variance in one measure can be predicted by variance in the other. This study introduces a statistical procedure to determine the probability of a double dissociation when the correlation between measures is taken into account. Different quantitative definitions of dissociations were compared in two large samples of neurological patients, and applied to four pairs of measures (two for language, two for hemispatial neglect) with different degrees of intercorrelation (ranging from +.21 to +.84). If the correlation between measures is not taken into account, large numbers of dissociated cases may be missed, especially for measures that are highly correlated. There are also qualitative differences between methods in the identity of those individuals who meet each definition.

Early in the 19th century, investigators interested in the relations between brain and mental faculties became involved in the search for associations between specific brain lesions and the loss of behavioral functions (Gall & Spurzheim, 1809). The birth of this anatomo-clinical correlation method was accompanied by some powerful insights into the "necessary and sufficient conditions" to test hypotheses concerning the association between functional loss and a lesion to some specific location in the brain. The young physician Bouillaud (1825) described with extreme clarity the principle of double dissociation: "Two activities are functionally separate if they can be disrupted in isolation from each other" (discussed in Luzzatti & Whitaker, 2001).

The logic of double dissociations has had an immense influence on neuropsychological research since the second half of the 19th century, but the first clear operational definition did not appear until halfway through the 20th century, in an influential paper by Teuber (1955) proclaiming that a double dissociation of symptoms is necessary to establish specificity of functions. Increasingly sophisticated versions of this concept have developed over the years (Bates, Appelbaum, & Allard, 1991; Bates, McDonald, MacWhinney, & Appelbaum, 1991; Bub & Bub, 1988; Caramazza, 1986; Dunn & Kirsner, 1988, 2003; Goodglass,

1993; Kinsbourne, 1971; Marin, Saffran, & Schwartz, 1976; Shallice, 1988; Weiskrantz, 1968), but the classical formulation can be described as follows:

> Two groups or two single cases (P1 and P2) show differential impairments in Tasks A and B, such that P1 is impaired in Task A but spared in Task B while P2 is impaired in Task B but spared in Task A.

Under these conditions, it must be concluded that the pattern of impairment "may be drawn by two independent functions, F1 and F2 involved in Tasks A and B" (Vallar, 1999).

This methodological tool has been used extensively in both group and single-case studies to "elucidate the multicomponential architecture of mental functions and their neural basis" (Vallar, 1999). However, such proposed dissociations are often based upon subjective estimates of "high performance" on one task and "low performance" on another, without considering the probability that such an outcome might have occurred by chance if patients were drawn randomly from the general population, or from some reference population (e.g., large unselected samples of brain-injured patients at risk for a given class of disorders). When inferential statistics are applied, researchers sometimes make unwarranted assumptions about independence and the equivalence of variances and means between the two measures in question, leading to high risk of false positives and/or false negatives (Bates, Appelbaum, et al., 1991; Bates, McDonald, et al., 1991).

In order to determine whether two measures "come apart" in an interesting way in two or more brain-damaged patients, it is important to understand how well those measures "hang together" (i.e., correlate) across the populations of interest. The goal of this study is to introduce a statistical procedure to determine the probability of a double dissociation when the correlation between measures is taken into account.

We will begin with a discussion of artifacts that occasionally arise when the logic of double dissociations is applied. This will include some problems that have been explored by other authors, involving measurement and population variance, and uncontrolled properties of the

"performance/resource curve" (Shallice, 1988). However, our main focus will be on the problem of measurement intercorrelation, which (to the best of our knowledge) has not yet received serious consideration in neuropsychological research, at least not with reference to the particular issue of quantifying dissociations. Following this review, we will provide a series of concrete examples from large-sample studies of linguistic and nonlinguistic functions, demonstrating why the quantification of both single and double dissociations requires knowledge of the correlation between each pair of measures. Our goal is not to criticize prior research for failing to take this issue into account, but rather to provide neuropsychological researchers with analytical tools that can be used in a variety of ways, depending on the goals of the investigator, to quantify and evaluate profiles of association and dissociation in patients with neurological impairments.

Some Problems in the Quantification of Behavioral Dissociations

Under what conditions can we conclude that a true double dissociation has been found, of the sort that justifies conclusions about the functional and neural separation of two functions? Figure 1a illustrates a potential dissociation between two measures, A and B. In the situation illustrated by this bar graph, Patient P1 is relatively spared on Measure A and impaired on Measure B; P2 shows the opposite profile, with relative sparing on Measure B and impairment on Measure A. This pattern seems to meet the requirements outlined by Bouillaud, Teuber and other proponents of the logic of double dissociation. In fact, if we regraph the same data as a line graph (Fig. 1b), the desired cross-over interaction discussed by many authors is clearly revealed (e.g., Dunn & Kirsner, 1988). However, as noted by Shallice (1988), Bates, Appelbaum, et al. (1991), Bates, McDonald, et al. (1991), Elman et al. (1996), Vallar (1999) and others, this apparent cross-over interaction can be misleading.

The first set of problems concerns the broad variability that is commonly seen in neuropsychological research. This variability is actually the confluence of two quite different sources of variation – both of which can make interpretation
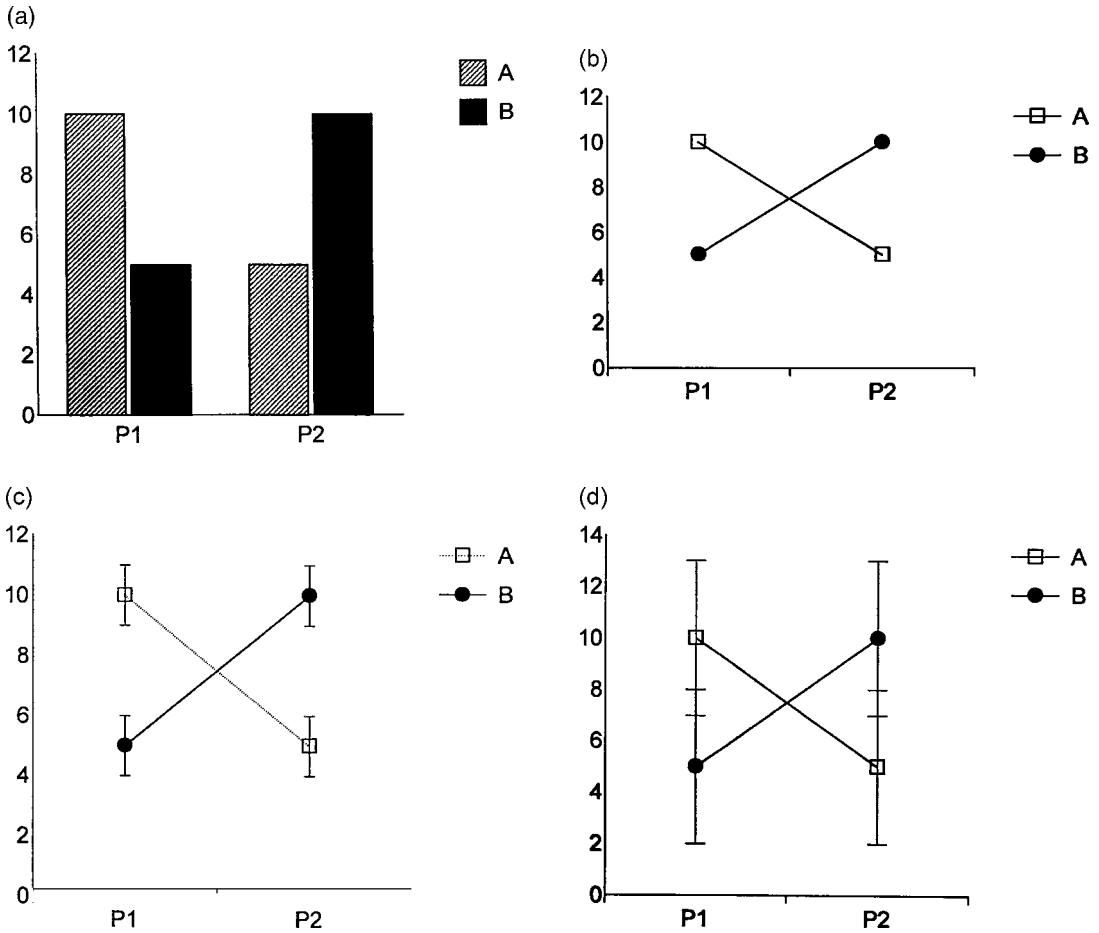
(a)

(b)

(c)

(d)

Fig. 1. (a) Hypothetical double dissociation between two tasks, in two individual patients (or two groups), expressed as a bar graph. (b) Hypothetical double dissociation between two tasks, expressed as a line graph to underscore cross-over interaction. (c) Line graph of a hypothetical double dissociation between two tasks, with error bars to indicate high reliability. (d) Line graph of a hypothetical double dissociation between two tasks, with error bars to indicate low reliability (high variance).

of cross-over interactions difficult. The first source of variation is *measurement variation*, or in psychometric terminology, *measurement unreliability*. The second source of variation is *within-population variation*, that is, differences between individuals that would show up in the data even if our instruments were perfectly reliable, that is, without measurement error variance. We will illustrate both problems with the same set of figures.

With regard to measurement variation, consider the two individuals, P1 and P2 of Figure 1a and b. If we repeat the test on the same individuals

"*n*" times we may find very different outcomes. Figure 1c illustrates the case of a very low standard error, while Figure 1d a very large one. Holding the means constant, it is clear that a meaningful, reliable double dissociation for P1 and P2 can be consistently obtained in the case of a low standard error, but not for the large one. Thus knowledge of the reliability of our two measures and their standard error of measurement in the population under study are essential factors in the interpretation of double dissociations.

The second type of variation has to do with individual differences within the population from

which our patients are drawn, and is relevant if we want to generalize from individual patients to some larger group. For example, P1 may now refer to a population of individuals with a very specific lesion site and type, while P2 may represent a different, equally specific lesion site and type. We may use the same Figure 1c and d to represent this situation, but now the error bars represent the variability in each population. Inferring a double dissociation for P1 and P2 is in this case heavily dependent upon the different degree of variability within the populations.

Thus we can see that the variance around the mean can be quite large for two kinds of reasons: because the measures themselves are unreliable, and/or because there really are large and meaningful differences among individuals within populations in performance on this task. Large variances of either type can lead to situations in which the interpretation of cross-over interactions is problematic. Finally, it should be noted that low reliability also leads to a low correlation between measures, which (as we will demonstrate below) increases the probability of false positives in our search for meaningful dissociations.

A different kind of problem in neuropsychological research involves the relationships between performance and capacity that underlie neuropsychological measures, first discussed in detail by Shallice (1988; see also Bates, Appelbaum, et al., 1991; Elman et al., 1996; Vallar, 1999). The key insight here lies in what Norman and Shallice (1983) call "performance/ capacity curves." Figure 2 illustrates hypothetical relationships between performance and some underlying capacity or resource (e.g., attention, working memory) in three different tasks (A, B, and C). Task A is governed by Function 3, a linear performance/resource relation in which a drop in capacity leads to a proportional drop in performance. Although this kind of relationship is often assumed in neuropsychological research, it is actually fairly rare in research on attention and performance. Task B is governed by Function 1, a common nonlinear function for well-practiced tasks in which performance remains near ceiling until a substantial amount of capacity is lost, and then drops sharply. For Task C, performance and capacity follow a more complex S-shaped relation (Function 2) that is also quite common in the literature on attention and performance. In this third case, decrements in capacity up to 30% have virtually no measurable effect on performance. Given these three functions, consider the patterns of sparing and impairment that might obtain for two hypothetical patients who differ only in the
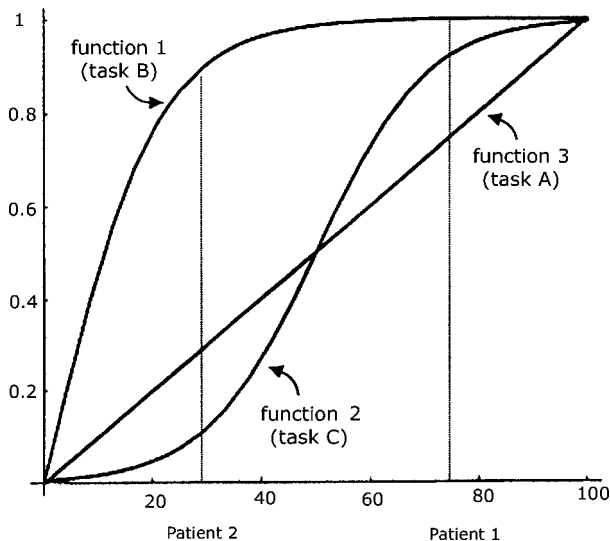

Fig. 2. Illustrations of capacity/performance functions.

amount of capacity that they have lost. P1 has lost 25% of this resource (operating at 75% capacity), P2 has lost 70% of this resource (operating at 30% capacity). Although these two patients actually differ only in quantitative terms, the relations in Figure 2 suggest several qualitative outcomes: (1) that both patients are spared on Task B, and to the same degree, and (2) that there is a double dissociation between Tasks A and C, such that Patient 1 is better on C than A, while Patient 2 is better on A than C. In the absence of information about performance/capacity functions, we might be tempted to draw misleading conclusions about the architecture that underlies these profiles.

These are difficult problems that can only be solved if the investigator has detailed information about the distribution of performance on the measures in question in the reference population from which single cases or groups of patients are drawn. The choice of reference populations is also important, because our understanding of the measurement properties of a clinical instrument must be based on a population that displays significant variation. Suppose, for example, that we were to obtain the means and variances on Measures A and B from a population of healthy normal controls who rarely make an error of any kind. The means for our two measures would hover near 100%, and the variance would be very small. Because of these ceiling effects, we would know next to nothing about the performance/capacity functions that govern our two measures when resources are reduced below normal levels, and as a result we could easily fall prey to the kind of situation illustrated in Figure 2. A better strategy might be to identify a large reference population of unselected patients, or patients who are at risk for a broad class of behavioral deficits and dissociations, including the ones tapped by A and B. This is the situation that we will assume for the remainder of this paper.

The two problems that we have just described (i.e., measurement/population variance, and performance/capacity curves) have been known for some time, although they are not always taken into account in neuropsychological research. In this paper, we will focus on a third (related) problem: the correlation between measures and its consequences for quantification of behavioral

dissociations. To the best of our knowledge, this problem has not been explored in any detail within the field of cognitive neuropsychology, although we will show that its effects on the probability of observing a behavioral dissociation are profound (cf. Bishop, 1997; Tomblin & Pandich, 1999, for a discussion of measurement correlation in diagnoses of childhood language impairments).

The term "correlation" refers to the degree to which performance on one measure is systematically related to performance on another. If two Measures A and B are perfectly correlated, then we can always exactly predict scores on one from scores on the other. When utilizing the Pearson product moment correlation, which is typical in neuropsychology, we further assume that the functional relation of interest is the linear function, and that the two attributes of interest are jointly normally distributed (i.e., bivariate normal). We will follow these standard assumptions for the rest of the paper (although other assumptions are possible, and sufficiently plausible in neuropsychological research that they would merit exploration).

In the hypothetical situations illustrated in Figure 1a–d, we assumed (implicitly) that Measures A and B were independent. However, the assumption of attribute independence is not always valid. Among other things, individual differences in severity of brain injury tend to drive outcomes in a common direction when patients undergo a battery of behavioral tests (Caplan & Hildebrandt, 1988; Schuell & Jenkins, 1959), and premorbid individual differences in education and/or intelligence may continue to have an effect when the same patients have sustained brain injury.

To illustrate the implications of this point, consider another hypothetical situation in Figure 3a, which represents the position of two individuals, Fred and Charlie, on a pair of dimensions X and Y, whose correlation is unknown. Suppose that we are State Troopers who have been told that a criminal is loose in our community, someone who is "very unusual." We have only one chance to make an arrest, and we know that one of these two individuals is the culprit, but the only information that we have to distinguish Fred from
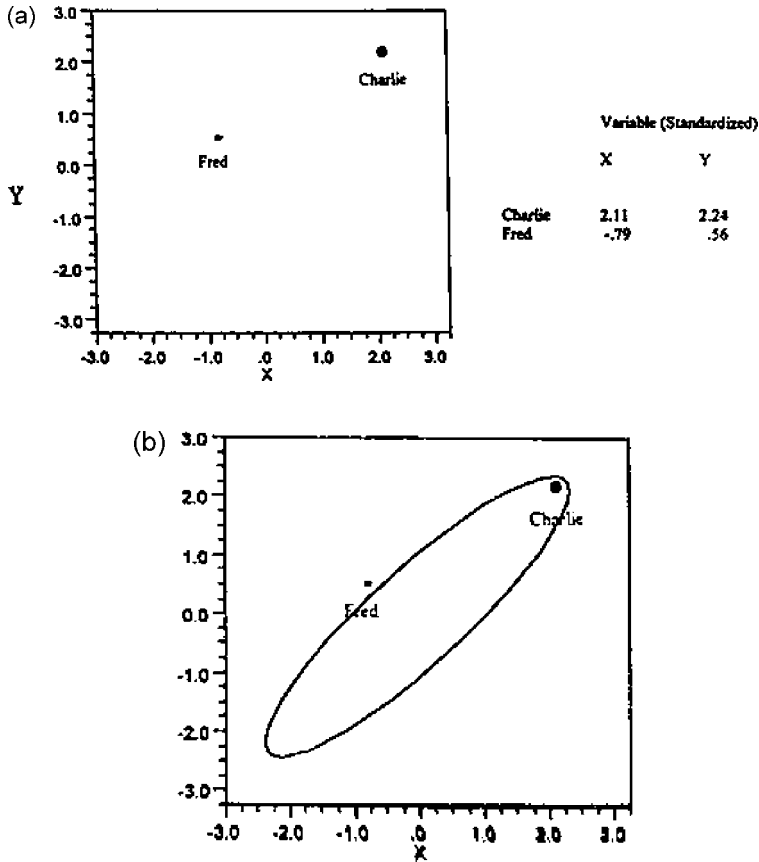
Fig. 3. Why correlation matters: An illustration.

Charlie is the information plotted in Figure 3a. Which of these two gentlemen is "unusual"? If we assume that the two dimensions X and Y are statistically independent (i.e., uncorrelated), then the definition of "unusual" has to lie in the distance of each score from the mean. Fred lies reasonably close to the mean on each dimension, while Charlie seems to be characterized by a pair of extreme scores. We would thus conclude that Charlie is the most likely suspect, and let Fred go.

Suppose, however, that the two dimensions X and Y are highly correlated ($r = +.90$), illustrated by the correlation ellipsoid in Figure 3b. This correlation ellipsoid represents the 95% confidence region, that is, contains 95% of the population (the "typical cases") and excludes 5% of the population (i.e., the "unusual cases"). As we shall see below, correlation ellipsoids can vary in size and shape (though not in symmetry),

depending on the confidence intervals that we assume (i.e., whether we define "unusual" in terms of the outer 5%, 10%, 30%, etc.) and on the magnitude of the correlation that holds between our measures. Under the high correlation assumption, it is Fred who lies outside of the 95% correlation ellipsoid (in the "unusual" 5%), while Charlie falls inside the ellipsoid (within the 95% confidence ellipsoid that we have selected to define a "typical" relationship between measures). Hence Charlie represents a more typical case than Fred, despite his more extreme scores.

The point of this particular illustration is the following: Dissociations are meaningful and interesting because they represent unusual relationships between two measures, that is, relationships that cry out for some kind of an explanation. In neuropsychological research, it is traditionally assumed that these unusual relationships are

caused by a disruption in one or more neural mechanisms that normally operate in concert with the rest of the mind/brain, but have been separated (or "subtracted away") by injury. *Hence one of the factors that must be taken into account in the quantification of dissociations is the correlation that would normally hold in the absence of this particular injury.* The probability that we will observe an "unusual" dissociation between two measures is thus, in part, a function of the correlation between those measures in the reference population from which a candidate has been drawn. Extreme scores are a useful index of pathology, but they can be misleading in the absence of information about measurement correlation.

For the remainder of the paper, we will illustrate this point by presenting information from two large samples of patients from a single neurorehabilitative hospital in Rome. The first sample was drawn from a larger group of 204 patients who were administered a battery of language tests because they were believed to be at risk for aphasia, due in the vast majority of cases to acquired left-hemisphere injury. Note that the patients who received this particular battery (Ciurli, Marangolo, & Basso, 1996) represent a broad range of severity. The second sample was drawn from a separate group of 146 patients who received a battery of tests designed to detect various forms of neglect; these patients were all believed to be at risk for neglect, due in most cases to an acquired right-hemisphere lesion. Because patients were tested on the basis of specific neurological risk factors (especially side of lesion), these are not "unselected samples," and hence they do not represent the general population nor the population of all patients with neurological impairments. However, they do represent (in each case) an otherwise unselected population of patients who are at risk for a class of behavioral disorders (aphasia or neglect) in a given clinic, across a 3-year period. Most important for our purposes here, each group represents a population in which we might expect to find deviations in more than one direction, and hence double dissociations (among language measures in the left-hemisphere sample; among neglect measures in the right-hemisphere sample).

Within each sample (one at risk for language deficits, the other at risk for neglect), we chose two pairs of measures: one pair with a moderate to high correlation in that reference group, and another pair for which the correlation is relatively low (though significantly above zero). These four pairs of measures will be used to illustrate the number of dissociations that we would predict under different quantitative definitions if measures were uncorrelated, and compare those predictions with the number that are actually observed in each data set, with and without taking the correlation between measures into account.

METHOD

Participants

Sample 1 was drawn from 204 patients who were tested for suspected language deficits. Patients had a mean age of 69.3 years (range 15–87 years), and were tested at a mean interval of 13 months post-onset (range 3–180 months). There were 111 males and 93 females, representing a broad range of social class and educational levels (a mean of 8 years of school attendance, with a range from 0 to 18 years). In most cases (93%), patients had unilateral damage to the left hemisphere. Ninety-eight percentage were due to cerebrovascular accidents, 2% were due to tumors.

Although the same battery of language tests was administered to all patients, occasionally a particular subscale was skipped, or the patient received a score of zero when he or she proved to be completely untestable on that particular measure. For the exercises illustrated below, we decided to exclude all cases with zero scores that reflect untestability on either measure, because such cases cluster into a "basement" of floor effects that distort the bivariate normal distribution. We note, however, that we did attempt a replication of all the analyses presented below with and without these untestable cases. Patterns were quite similar in both sets of analyses (testimony to their robust nature in the face of deviations from bivariate normal). For the sake of economy, we will restrict our report to the smaller subsamples that result when these cases are excluded (yielding a total of 157 cases for the noun/verb naming comparison, 130 cases for the sentence repetition/comprehension comparison, both defined below).

Sample 2 includes 146 patients who were tested for suspected hemispatial neglect. Patients had a mean age of 67.3 years (range 23–91 years), and were tested at a mean post-onset interval of 4.5 months (range 1–32 months). There were 84 males and 62 females, with a

mean of 7.7 years of school attendance (range 3–18 years). Of these patients 96% had unilateral damage to the right-hemisphere; 99% of the cases were due to cerebrovascular accidents, 1% to other causes. The problems of missing data and/or untestable zero scores did not arise for this sample (except for the loss of a single case in the Wundt-Jastrow/personal-neglect comparison described below, yielding a total sample of 145).

Measures

Patients with suspected aphasia were given a battery of tests developed by Anna Basso and colleagues (Ciurli et al., 1996), similar in length and structure to standard aphasia batteries in English including the Boston Diagnostic Aphasia Examination (Goodglass & Kaplan, 1983) and the Western Aphasia Battery (Kertesz, 1982). For the analyses presented below, all scores were converted into z scores, so that the mean for each variable is set at zero and individual scores are expressed in standard deviations from the mean.

From the data for the complete aphasia battery, we computed pairwise correlations among the following measures: a 5-point fluency rating scale based on speech samples and performance throughout testing (0 = mute or unintelligible; 1 = primarily single-word speech; 2 = severely telegraphic speech, some connected discourse but large amounts of omission; 3 = mildly telegraphic speech with relatively few errors and omissions; 4 = normal fluency and melodic line), two subscores for naming (naming objects, 20 items; naming actions, 10 items), three repetition subtests (repeat words, 20 items; repeat nonwords, 20 items; repeat phrases and sentences, 10 items), one test of word comprehension (20 items), and two tests of sentence comprehension (the Token Test by De Renzi & Vignolo, 1962, with 36 items) and a separate sentence comprehension scale that was part of the Ciurli et al. battery, with 10 items. Two pairs of measures were chosen for the dissociation analyses presented below: one pair with an exceptionally high correlation (naming objects vs. naming actions, $r = +.84$, $N = 157$, $p < .01$), and another pair with one of the lowest correlations in the battery (repeat sentences vs. comprehend sentences, $r = +.35$, $N = 130$, $p < .01$).

Patients who were at risk for hemispatial neglect were given a battery of measures including a line cancellation test (Albert, 1973, 21 items), a letter cancellation test (Diller & Weinberg, 1977, 104 items), a test based on the Wundt-Jastrow perceptual illusion (Massironi, Antonucci, Pizzamiglio, Vitale, & Zoccolotti, 1988, 40 items), and a test of personal neglect (Zoccolotti & Judica, 1991, six items). Responses to the left and right side were scored separately for each test except the last one. Two measures were chosen for the analyses

presented below: a pair with a moderately high correlation (Wundt-Jastrow left-side and line cancellation left side, $r = -.60$, $N = 146$, $p < .01$) and a pair with a relatively low correlation (Wundt-Jastrow left side and personal neglect, $r = .21$, $N = 145$, $p < .05$).

RESULTS

Theoretical Distributions

We begin this exercise by establishing a series of quantitative definitions of "dissociation," and calculating the number of dissociations that we would expect by chance between two normally distributed measures with a zero correlation. Each of these quantitative definitions was chosen because it reflects standards that are often applied (implicitly or explicitly) in the literature on single and double dissociations.

1. *Quadrants*. In this case, dissociations are defined in terms of quadrants (see Fig. 4a) in a two-dimensional space, with the cross-point representing the means for both measures. Assuming two uncorrelated, normally distributed measures and no ties, we would expect 25% of the sample to fall in the dissociated High *X*/Low *Y* quadrant (lower right), and another 25% would fall in the dissociated Low *X*/High *Y* quadrant (upper left). The remaining 50% of the sample would be defined as "undissociated," including 25% in High *X*/High *Y* quadrant (upper right) and another 25% in Low *X*/Low *Y* quadrant (lower left). This criterion quantifies an informal definition of "double dissociation" that is often applied in single-case studies.

2. *Strong dissociation*. This is a much more stringent criterion, under which a patient is defined as "strongly dissociated" only if he or she falls one standard deviation above the mean on one measure and one standard deviation below the mean on the other. This kind of relationship is also illustrated in Figure 4b, which contains a "box" defined by one standard deviation above and below the means for each measure. Under this most rigorous definition, we would expect only 3% of the population to fall within the space of High *X*/Low *Y* (extreme lower right box) and another 3% in
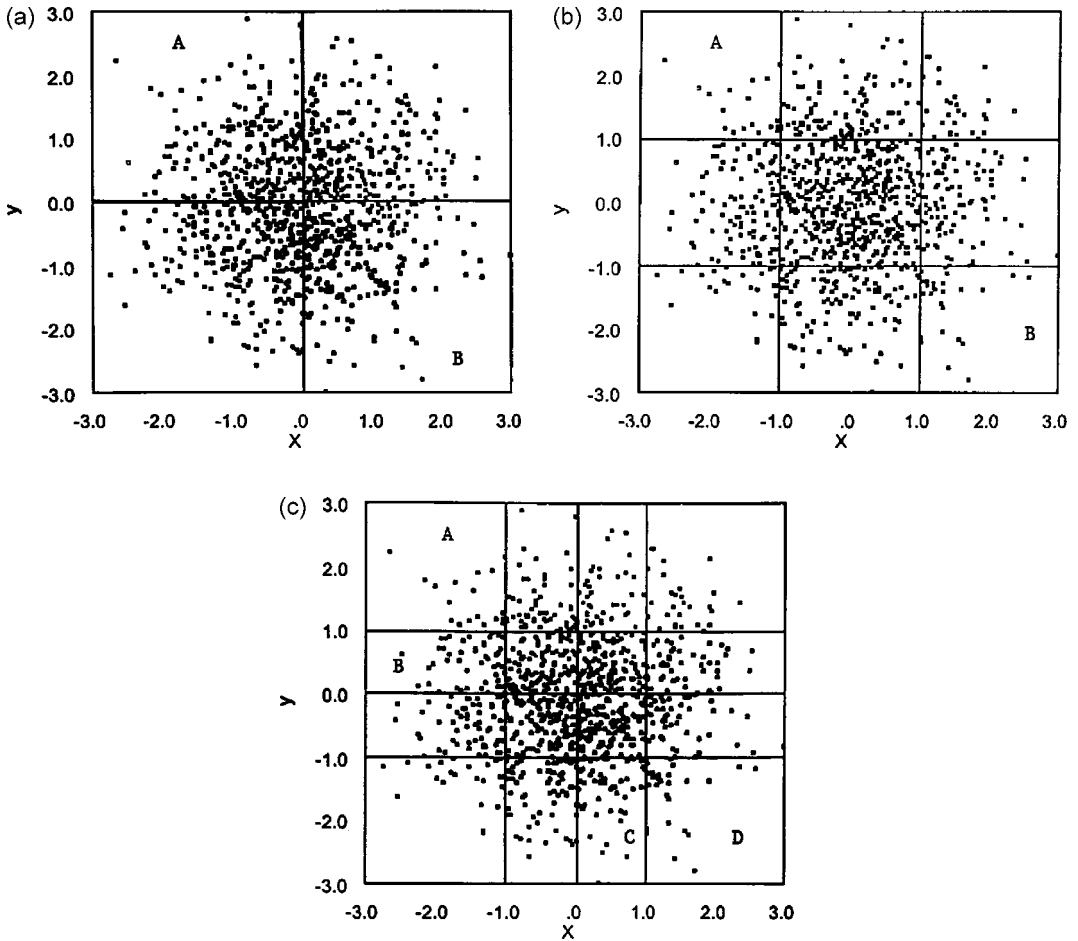
Fig. 4. (a) Dissociations defined in terms of quadrants. (b) Dissociations defined as "Strong Dissociations." Quadrants labelled as "A" and "B" contain dissociated cases by this definition. (c) Dissociations defined as "Weak Dissociations." Cases in areas labelled as "A", "B", "C" and "D" are dissociated by this definition.

the space of Low X/High Y (extreme upper left box), if the measures are uncorrelated. The remaining 94% of the population would not qualify as "truly dissociated." Hence this criterion approximates a conventional two-tailed test in which 5% of the population falls in the "statistically significant" range (2.5% in each tail of the distribution).

3. *Weak dissociation*. This is a somewhat more lenient definition that still takes the standard deviation into account. However, in this case we would only require the patient to fall one standard deviation below the mean on a single

dimension, as long as the patient also fell within one of the High/Low or Low/High quartiles defined in (1) above. This kind of relationship is illustrated in Figure 4c. Under this less stringent definition, we would expect 12.8% of the population to fall in the High X/Low Y space and another 12.8% to fall in the Low X/High Y space, assuming, again, uncorrelated measures. This criterion is close to the definitions that are often used in the literature on language and learning disabilities. For example, Specific Language Impairment is typically defined in terms of a standardized

language score that is one or more standard deviations below the mean in an individual whose performance on a nonverbal intelligence test falls within the normal range (operationalized as scores above 70, 75 or 80, depending on the study – Bishop, 1997; Leonard, 1998).

All of these theoretical distributions are based on the assumption that our measures are uncorrelated and follow the bivariate normal distribution (with the same means and variances). As we shall see below, the difference between predicted and observed outcomes increases proportionally as the size of the correlation goes up, in a direction that makes it increasingly difficult to detect a dissociation of any kind unless the correlation is taken into account.

Observed Distributions

Using the definitions provided above, let us now examine how many dissociations are detected in these data sets if we fail to take the correlation between measures into account. We will illustrate the first pair of measures (noun/verb dissociations) in some detail, and then summarize results for the other three pairs of measures in a more succinct form, referring the reader to the relevant tables for numerical details.

Noun/Verb Dissociations

We begin with the relationship between object (noun) naming and action (verb) naming, which are correlated +.84. This pair of measures is a reasonable place to start, because the potential dissociation between object and action naming has sparked considerable interest in the aphasia literature (for reviews, see Bates, Chen, Tzeng, Li, & Opie, 1991; Bates, Wulfeck, & MacWhinney, 1991; Berndt & Zingeser, 1991; Cappa & Ullman, 1998; Caramazza & Hillis, 1991; Chen & Bates, 1998; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; Goodglass, 1993; Zingeser & Berndt, 1990). Double dissociations between object and action naming have been reported in different languages including English, Italian, and Chinese. When such dissociations are described, the Noun > Verb pattern is typically reported for nonfluent Broca's aphasics, while

the opposite profile (Verbs > Nouns) is more common among fluent aphasics (anomics and/or Wernicke's). Such associations between form class and aphasia type have led to speculation that the mediation of noun versus verb processing may involve different regions of the cortex. Hence the search for noun-verb dissociations is well motivated.

The first three rows of Table 1 summarize the number and percent of dissociations that are actually observed under each of the definitions provided above, compared with the number that we would predict if the measures were truly independent. (Rows 4–9 of Tables 1–4 provide the same kind of information for analyses that will be detailed below.) Clearly, there is a substantial difference between predicted and observed values under most definitions of "dissociation."

Using a simple definition in terms of High/Low and Low/High quadrants, we would expect about 79 cases (50% of 157) to fall into one or the other of the "interesting" quadrants if Noun and Verb naming were independent (39.3 or 25% in Noun > Verb; 39.3 or 25% in Verb > Noun). Instead, we find only 20 dissociations overall, including 12 cases (7.6%) in the Noun > Verb quadrant, and eight cases (5.1%) in the Verb > Noun quadrant (see Table 1). Hence the number of observed dissociations is about one-fourth the number that we would expect to find if the two measures were uncorrelated.

Applying our most stringent definition (requiring one standard deviation above and one below on each measure), we would expect to find between 9 and 10 cases (6% of 157) within one of the two "interesting" portions of the noun-verb space. Instead, we find absolutely no cases that meet the stringent criterion: 100% of our 157 patients fall into the "uninteresting" range.

Finally, loosening this definition to a weaker criterion requiring one standard deviation in a single direction within a High/Low quadrant, we would expect approximately 40 dissociative cases in the sample (12.8% in each direction). Instead, we find only three cases of Noun > Verb (1.9%) and two cases of Verb > Noun (1.2%) – see Table 1. In other words, we find one-eighth the number of dissociations that we would expect on

Table 1. Percent Theoretical Versus Percent Observed Incidence of Dissociations for Noun (Object) Versus Verb (Action) Naming (Number of Cases in Parentheses).

Classification scheme	Noun > Verb		Verb > Noun		No dissociation	
	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)
1. Quadrants (High/Low)						
157 cases	25 (39.3)	7.6 (12)	25 (39.3)	5.1 (8)	50 (78.5)	87.3 (137)
2. Strong dissociation (+1 *SD*/−1 *SD*)						
157 cases	3 (4.7)	0 (0)	3 (4.7)	0 (0)	94 (147.6)	100 (157)
3. Weak dissociation (High/Low)						
(−1 *SD* in one direction)						
157 cases	12.8 (20.1)	1.9 (3)	12.8 (20)	1.2 (2)	74.4 (116.8)	96.8 (152)
4. 90% correlation ellipsoid						
157 cases	5 (7.9)	1.9 (3)	5 (7.9)	4.5 (7)	90 (141.3)	93.6 (147)
5. 90% correlation ellipsoid (High/Low only)						
157 cases	2.5 (3.9)	1.9 (3)	2.5 (3.9)	2.5 (4)	95 (149.2)	95.5 (150)
6. 70% correlation ellipsoid						
157 cases	15 (23.6)	8.3 (13)	15 (23.5)	10.8 (17)	70 (109.9)	80.9 (127)
7. 70% correlation ellipsoid (High/Low only)						
157 cases	7.5 (11.8)	3.2 (5)	7.5 (11.8)	3.2 (5)	85 (133.5)	93.6 (147)
8. 50% correlation ellipsoid						
157 cases	25 (39.3)	22.3 (35)	25 (39.3)	31.8 (50)	50 (78.5)	45.9 (72)
9. 50% correlation ellipsoid (High/Low only)						
157 cases	12.5 (19.6)	5.1 (8)	12.5 (19.6)	3.8 (6)	75 (117.8)	91.1 (143)

Table 2. Percent Theoretical Versus Percent Observed Incidence of Dissociations for Sentence Comprehension Versus Sentence Repetition (Number of Cases in Parentheses).

Classification scheme	Comprehension > Repetition		Repetition > Comprehension		No dissociation	
	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)
1. Quadrants (High/Low)						
130 cases	25 (32.5)	23.8 (31)	25 (32.5)	11.5 (15)	50 (65)	64.6 (84)
2. Strong dissociation (+1 *SD*/−1 *SD*)						
130 cases	3 (3.9)	1.5 (2)	3 (3.9)	0.8 (1)	94 (122.2)	97.7 (127)
3. Weak dissociation (High/Low)						
(−1 *SD* in one direction)						
130 cases	12.8 (16.6)	16.2 (21)	12.8 (16.6)	7.7 (10)	74.4 (96.7)	76.2 (99)
4. 90% correlation ellipsoid						
130 cases	5 (6.5)	0 (0)	5 (6.5)	5.4 (7)	90 (117)	94.6 (123)
5. 90% correlation ellipsoid (High/Low only)						
130 cases	2.5 (3.3)	0 (0)	2.5 (3.3)	1.5 (2)	95 (123.5)	98.5 (128)
6. 70% correlation ellipsoid						
130 cases	15 (19.5)	6.9 (9)	15 (19.5)	16.2 (21)	70 (91)	76.9 (100)
7. 70% correlation ellipsoid (High/Low only)						
130 cases	7.5 (9.8)	6.9 (9)	7.5 (9.8)	6.9 (9)	85 (110.5)	86.2 (112)
8. 50% correlation ellipsoid						
130 cases	25 (32.5)	26.9 (35)	25 (32.5)	32.3 (42)	50 (65)	40.8 (53)
9. 50% correlation ellipsoid (High/Low only)						
130 cases	12.5 (16.3)	16.2 (21)	12.5 (16.3)	8.5 (11)	75 (97.5)	75.4 (98)

Table 3. Percent Theoretical Versus Percent Observed Incidence of Dissociations for Wundt-Jarow Illusion Versus Line Bisection (Number of Cases in Parentheses).

Classification scheme	Wundt-Jarow > Line bisection		Line bisection > Wundt-Jarow		No dissociation	
	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)
1. Quadrants (High/Low)						
146 cases	25 (36.5)	12.3 (18)	25 (36.5)	14.4 (21)	50 (73)	73.3 (107)
2. Strong dissociation (+1 *SD*/−1 *SD*)						
146 cases	3	1.4	3	0	94	98.6
3. Weak dissociation (High/Low)						
(−1 *SD* in one direction)						
146 cases	12.8 (18.7)	8.2 (12)	12.8 (18.7)	3.4 (5)	74.4 (108.6)	88.4 (129)
4. 90% correlation ellipsoid						
146 cases	5 (7.3)	6.2 (9)	5 (7.3)	0.7 (1)	90 (131.4)	93.2 (136)
5. 90% correlation ellipsoid (High/Low only)						
146 cases	2.5 (3.7)	6.2 (9)	2.5 (3.7)	0.7 (1)	95 (138.7)	93.2 (136)
6. 70% correlation ellipsoid						
146 cases	15 (21.9)	21.9 (32)	15 (21.9)	5.5 (8)	70 (102.2)	72.6 (106)
7. 70% correlation ellipsoid (High/Low only)						
146 cases	7.5 (11)	8.9 (13)	7.5 (11)	5.5 (8)	85 (124)	85.6 (125)
8. 50% correlation ellipsoid						
146 cases	25 (36.5)	38.4 (56)	25 (36.5)	30.1 (44)	50 (73)	31.5 (46)
9. 50% correlation ellipsoid (High/Low only)						
146 cases	12.5 (18.3)	8.9 (13)	12.5 (18.3)	11.6 (17)	75 (109.5)	79.5 (116)

Table 4. Percent Theoretical Versus Percent Observed Incidence of Dissociations for Wundt-Jarow Illusion Versus Personal Neglect (Number of Cases in Parentheses).

Classification scheme	Wundt-Jarow > Personal Neglect		Personal Neglect > Wundt-Jarow		No dissociation	
	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)	Predicted (%)	Observed (%)
1. Quadrants (High/Low)						
145 cases	25 (36.3)	24.1 (35)	25 (36.3)	18.6 (27)	50 (72.5)	57.2 (83)
2. Strong dissociation (+1 *SD*/−1 *SD*)						
145 cases	3 (4.4)	0 (0)	3 (4.4)	4.1 (6)	94 (136.3)	95.9 (139)
3. Weak dissociation (High/Low)						
(−1 *SD* in one direction)						
145 cases	12.8 (18.6)	11.7 (17)	12.8 (18.6)	14.5 (21)	74.4 (107.9)	73.8 (107)
4. 90% correlation ellipsoid						
145 cases	5 (7.3)	0 (0)	5 (7.3)	7.6 (11)	90 (130.5)	92.4 (134)
5. 90% correlation ellipsoid (High/Low only)						
145 cases	2.5 (3.6)	0 (0)	2.5 (3.6)	2.1 (3)	95 (137.8)	97.9 (142)
6. 70% correlation ellipsoid						
145 cases	15 (21.8)	15.2 (22)	15 (21.8)	12.4 (18)	70 (101.5)	72.4 (105)
7. 70% correlation ellipsoid (High/Low only)						
145 cases	7.5 (10.9)	9.7 (14)	7.5 (10.9)	6.9 (10)	85 (116)	83.4 (121)
8. 50% correlation ellipsoid						
145 cases	25 (36.3)	19.3 (28)	25 (36.3)	40.7 (59)	50 (72.5)	40 (58)
9. 50% correlation ellipsoid (High/Low only)						
145 cases	12.5 (18.1)	12.4 (18)	12.5 (18.1)	13.8 (20)	75 (108.8)	73.8 (107)

this criterion if object and action naming were independent.

Why do we find so few dissociations in this example under any definition? Because of the very high positive correlation between object and action naming, the data are concentrated along the diagonal connecting High-High and Low-Low. This makes it very difficult to find dissociations between the two measures under any definition, because there are so few cases of any kind in the High-Low and Low-High quadrants.

Dissociations Between Sentence Comprehension and Sentence Repetition

What would happen for two variables with a weaker correlation? Within the language battery, we selected Sentence Comprehension (C) and Sentence Repetition (R) for this purpose. Across the 130 patients, the correlation was a positive but relatively low $+.35$. This is one of the lowest correlations among all of our aphasia subscales, replicating a partial dissociation that has been known for more than 100 years. Table 2 presents the number of dissociations that we observe under each definition, compared with the number that we would expect by chance if the variables were independent. We refer the reader to Table 2 for details, and summarize briefly the disparities seen between observed and predicted outcomes with a relatively weak correlation.

Using the quadrants, we would expect 25% of our cases to fall in $C > R$ and another 25% in $R > C$ if the two measures were bivariate normal and uncorrelated. Observed results are close to these expectations for $C > R$ (23.8%), but fall well short for $R > C$ (11.5%, less than half of the number expected). The asymmetry between $C > R$ and $R > C$ under both these definitions means that the two measures do not strictly adhere to the bivariate normal distribution, even though all scores were transformed into z scores to equate means and variances (we will return to this point in the discussion).

Applying the most stringent definition of dissociation, we would expect to find 3% of the population in each of the two extreme regions (6% total, representing 8–9 cases). Instead, we find only three dissociated cases (two in $C > R$; one in $R > C$), less than half the number that we

would expect by chance if these measures were independent.

Finally, under the weaker definition that requires a standard deviation in only one direction, we would expect to find approximately 33–34 dissociations out of the sample of 130 patients (12.8% in each of the two dissociated quadrants). This time there are fewer discrepancies between predictions and observations: We do find 31 dissociations under this definition (close to the predicted 33–34), although they are asymmetrically distributed, with 16.2% in the $C > R$ group (more than the 12.8% predicted) and 7.7% in the $R > C$ group (a bit more than half the 12.8% predicted). In other words, the weaker definition applied to a data set with a $+.35$ correlation approximates the values that we would obtain if the measures were independent. Again, however, there is an asymmetry in the data (more cases of $C > R$ than $R > C$) indicating a deviation from the assumed bivariate normal distribution.

To summarize, the difference between predicted and observed values is not as dramatic with a correlation of $+.35$ (for sentence comprehension vs. repetition) as it was with a correlation of $+.84$ (the action/object-naming example above). Hence the size of the correlation does matter, and, as it decreases, the data begin to approximate the assumption of measurement independence that is often made implicitly in neuropsychological research. Nevertheless, depending on what definition of dissociation we choose, we tend to find fewer cases than we would expect if the measures were truly independent, and the observed data reflect asymmetries (i.e., deviations from bivariate normal) that would not be predicted on mathematical grounds.

Dissociations in Neglect: Wundt-Jastrow Illusion Versus Line Cancellation

We turn now to the sample of patients at risk for hemispatial neglect, and begin with two variables that have a moderate correlation ($r = +.60$, one of the highest correlations in the neglect battery), between the Wundt-Jastrow illusion test (WJ) and the line cancellation (LC) task. Although the two tasks are both used to detect hemispatial neglect, their cognitive requirements are quite separable. The line cancellation task requires the patients to

sequentially explore the different segments displayed on each stimulus page, while the Wundt-Jastrow test can be solved by examining the stimulus as a whole. Different tasks to evaluate hemispatial neglect have been shown previously to agglomerate into partially separable clusters (Pizzamiglio et al., 1992). For this reason, the two tasks are expected to show a moderate degree of dissociation, but this dissociation is of limited theoretical importance in the study of neglect. For precisely that reason, it is interesting to ask how many dissociations we will find under our various definitions of ''unusual,'' comparing this case with the two previous examples in which dissociations were well motivated, and expected. Table 3 compares the predicted and observed scores under each definition of ''dissociation'' for these two tasks.

When dissociations are defined in terms of quadrants, we see a greater disparity between observed and predicted outcomes. We find only 12.3% in the WJ > LC quadrant and 14.4% in the LC > WJ quadrant, roughly half of the expected number.

Using the most stringent definition ($+1\,SD$ in one direction, $-1\,SD$ in the other), we would expect 6% of all cases to fall in one or the other of the two extreme regions (a total of 8 dissociations, 4 per region). Instead we only find only 1.4% (2 cases) that meet this extreme definition, both of them in the WJ > LC region.

Finally, using the weaker definition (a standard deviation below the mean in only one direction), we would expect 12.8% of cases to fall in each dissociative region if the two measures of neglect were completely independent. The observed numbers are smaller: 8.2% in the WJ > LC region and 3.4% in the LC > WJ region.

To summarize results for this pair of measures, there is clearly a difference between the number of outcomes we observe and the number that we would predict if the two measures were completely independent, showing again that the correlation between measures has a strong influence on the number of dissociations that are observed. A comparison of this exercise with our findings above for noun versus verb naming yields some additional insights. The Wundt-Jastrow and line cancellation tasks are correlated at +.60, but were

selected to measure the same deficit. The noun- and verb-naming tasks are correlated at +.84, but were selected to measure forms of naming that are supposed to dissociate in a substantial number of patients. And yet we have found more dissociations in the former case (where they are of little interest) than the latter case (where they are expected), underscoring the importance of taking the correlation between measures into account.

Dissociations in Neglect: Wundt-Jastrow Versus Personal Neglect

In contrast with the above exercise involving two moderately correlated measures of extrapersonal neglect, there are sound theoretical reasons to expect occasional dissociations between extrapersonal and personal neglect (Galati et al., 2000; Guariglia & Antonucci, 1992; Pizzamiglio et al., 1992). It has been argued that personal neglect involves body-centered representations of space while extrapersonal neglect involves other forms of spatial coding (e.g., object centered or environment centered), and that these representations of space are mediated by different regions of cortex and display a different degree of lateralization (Galati et al., 2000). As expected, the Wundt-Jastrow Illusion (WJ) and the personal neglect measure (PN) are correlated at only +.21, one of the lowest correlations in the neglect battery (constituting a replication of Pizzamiglio et al., 1992). Table 4 compares observed and predicted results for these two measures under each definition.

When dissociations are evaluated in terms of quadrants, observed results in the WJ > PN region were close to the predicted 25% (at 24.1%), but results in the opposite PN > WJ region fell short (at 18.6%). Again, this kind of asymmetry cannot reflect task difficulty per se when z score transforms are used, and must therefore reflect deviations from the bivariate normal distribution.

Applying the most stringent definition of dissociation, we would expect 3% each within the respective WJ > PN and PN > WJ regions (6% of all cases). Instead, we found no cases at all in WJ > PN, and only 4.1% in PN > WJ. Hence, even with a correlation as low as +.21, we still find fewer cases than we would expect when measurement independence is assumed. We also find an asymmetry between WJ > PN and

PN > WJ that may be meaningful (reflecting facts about the neural substrates for extrapersonal vs. personal neglect), although it may also reflect deviations from bivariate normal due to sampling error.

Finally, using the looser definition, the numbers that we observe are a fairly close match to the expected 12.8%: 11.7% of all cases in WJ > PN and 14.4% in PN > WJ.

From these exercises, it should be evident that the correlation between measures matters a great deal when we set out to find dissociations between clinical measures, including widely used and well-motivated measures of aphasia and neglect. In the final section, we will describe a statistical technique that takes the correlation between measures into account, with dissociations defined in terms of the region outside a specified correlation ellipsoid.

Taking the Correlation into Account

For the same four pairs of measures described above, we take the correlation between measures into account by constructing three different correlation ellipsoids, containing 90%, 70% and 50% of the population, respectively. As in the Fred and Charlie example described in the introduction, we use these criteria to decide who is "unusual" (defined as cases that fall outside of the ellipsoid) and who is "typical" (defined as cases that fall inside the ellipsoid). The size of the ellipsoid itself is an arbitrary decision, and depends on the investigator's goals. We chose these three percentages to illustrate a range of options that are often applied (implicitly or explicitly) in the neuropsychological literature, roughly analogous to the three definitions of dissociation applied in the previous section (where measurement independence was assumed).

In all the illustrations that follow, we assume that our measures follow the bivariate normal distribution (although, as we have already seen, deviations from bivariate normal are detectable in these data). Applying the 90% ellipsoid, a "dissociation" is defined as a case that falls somewhere in the "outer space" occupied by 10% of the population. Assuming a bivariate normal distribution, 5% should lie in the $X > Y$ region (on one side of the diagonal) while the

other 5% should lie in the $Y > X$ region (on the other side of the diagonal). Hence this is roughly similar to the most stringent criterion adopted in the previous section (which predicted 3% in each dissociative zone). Similarly, applying the 70% ellipsoid, a dissociation is defined as a case that falls in the "outer space" occupied by 30% of the population, 15% on each side of the diagonal. Hence this criterion approximates the weaker definition of dissociation in which patients must fall at least one standard deviation below the mean on a single measure (predicting 12.8% in each dissociative zone). Finally, applying the 50% ellipsoid, we liberally define "unusual" to include half of the population, 25% on each side of the diagonal, roughly analogous to the quadrant approach illustrated in the previous section. A more detailed description of these computations for a multivariate normal of any dimensionality can be found in Appelbaum (2000).

These three correlation ellipsoids are illustrated in Figure 5a–c for two hypothetical variables with correlation of $+.50$, under each of our three definitions of a dissociation. Notice that these definitions of "dissociation" include cases that fall in the familiar High/Low and Low/High quadrants that we have used throughout the paper. However, the correlation-based definitions would also include cases in the High/High and Low/Low quadrants – the kinds of cases that are rarely taken into consideration in neuropsychological research. For example, the correlation-based method would classify an individual as a dissociated case, where Y is abnormally low for that patient's value of X, even though this individual is performing close to the group mean on both measures. This corresponds closely to the Fred example that we used in the introduction, an individual who looks relatively normal in terms of their scores, but looks abnormal when the correlation is taken into account (see Fig. 3). We would like to suggest that neuropsychological researchers may want to give such cases a careful look, broadening their definition of "dissociation" to consider these individuals.

On the other hand, we also recognize that many clinical researchers use the term "dissociation" only for cases that are truly deficient along
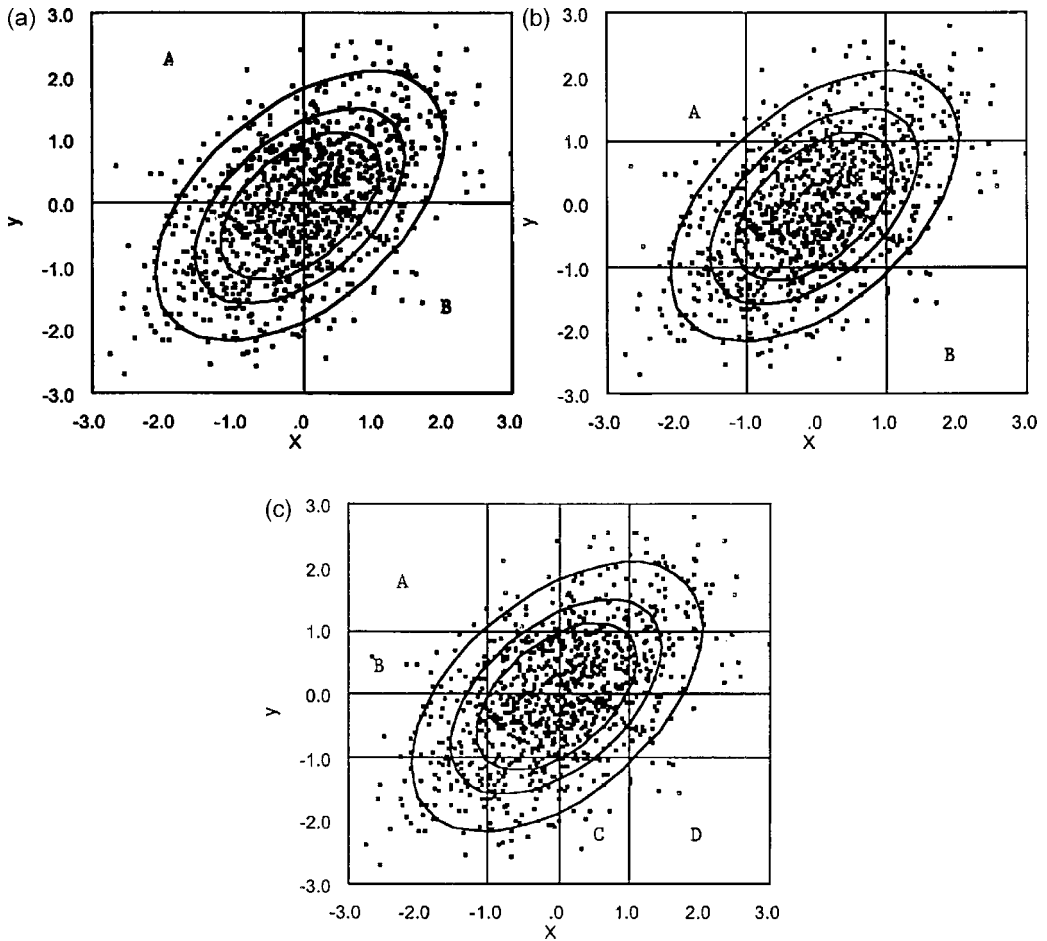
Fig. 5 (a) 90%, 70%, and 50% ellipsoids superimposed upon the quadrant definition of dissociations. (b) 90%, 70%, and 50% ellipsoids superimposed upon the ''Strong'' definition of dissociations. (c) 90%, 70%, and 50% ellipsoids superimposed upon the ''Weak'' definition of dissociations.

at least one dimension. Under this restriction, the term should not be applied to relatively unimpaired cases in the High/High quadrant, and probably would not be applied to globally impaired patients in the Low/Low quadrant. If a clinical researcher holds to this restriction but also wants to take the correlation into account, then the term ''dissociation'' would be applied only to cases that (a) fall outside the specific correlation ellipsoid, but also (b) fall within one of the usual High/Low or Low/High quadrants. Below, we will provide results with and without this restriction, for the respective 90%, 70% and 50% correlation ellipsoids.

Figures 6–9 present scatterplots for the four pairs of variables that we have just reviewed, with the respective 50%, 70% and 90% ellipsoids imposed. Numerical details for all of these exercises are provided in Tables 1–4, to facilitate comparison between methods that do and do not take measure correlation into account. Without repeating all the numbers listed in those tables, we can summarize results briefly in support of three conclusions: (1) the discrepancy between methods is most evident for measures that are highly correlated; as the correlation drops, techniques that take the correlation into account start to yield outcomes numerically similar to those that are
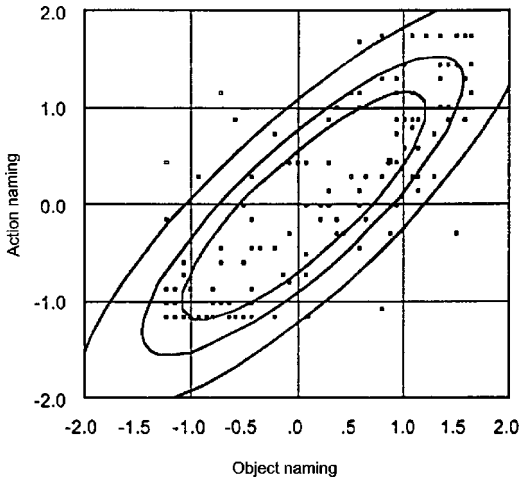
Fig. 6. Empirical distribution of object naming by action naming with regions of dissociation. Scores in standardized units. More than one case may be represented by a single point.
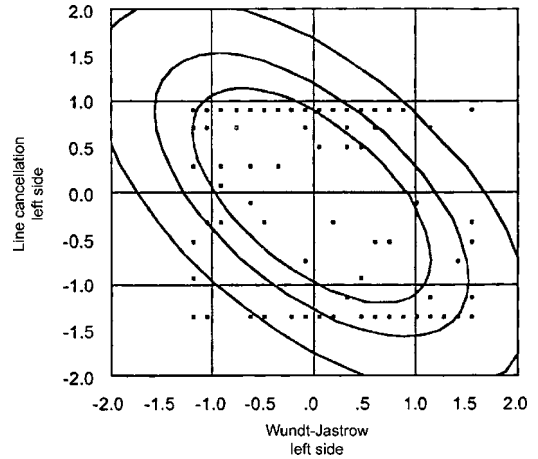

Fig. 8. Empirical distribution of Wundt-Jastrow left side and line cancellation left side with regions of cancellation. Scores in standardized units. More than one case may be represented by a single point.
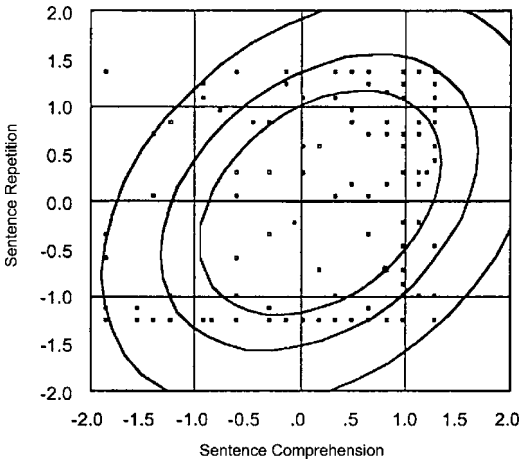

Fig. 7. Empirical distribution of sentence comprehension by sentence repetition with regions of dissociation. Scores in standardized units. More than one case may be represented by a single point.
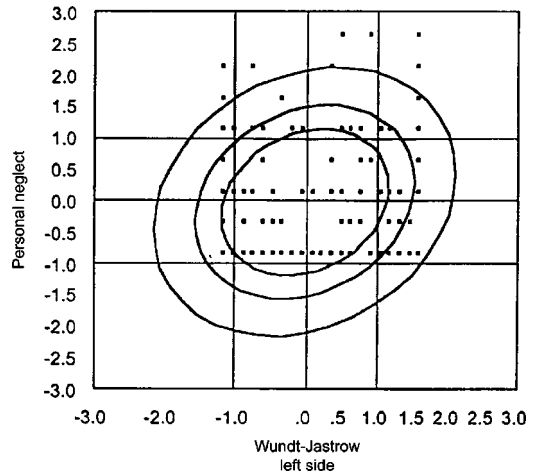

Fig. 9. Empirical distribution of Wundt-Jastrow left-side and personal neglect with regions of dissociation. Scores in standardized units. More than one case may be represented by a single point.

obtained assuming measurement independence, (2) there are also qualitative differences between methods in the *identity* of those individuals who meet each definition, (3) the observed data frequently deviate from the standard assumptions of normality, including the bivariate normal assumption that underlies the correlation-based

methods illustrated here, so that we may want to consider some additional techniques that do not rely on assumptions of bivariate normal.

Size of the Correlation
The disparity between methods is most evident for measures that are highly correlated (Tables 1

and 3). This is particularly evident in the case of noun-verb naming (which had the highest measurement correlation among our four examples, at +.84). Here there were gross disparities between predicted and observed values when we failed to take the correlation into account, under all three definitions. As a result, dissociations between noun and verb naming proved very hard to find. With the various correlation-based criteria, predicted and observed values were much closer, which means in turn that dissociations were much easier to detect. For example, the quadrant method and the 50% ellipsoid are similar in their predictions: 25% of all cases should fall in the Noun > Verb space and another 25% in the Verb > Noun space. Under the quadrant method (assuming measurement independence), we only found 7.6% in Noun > Verb and 5.1% in Verb > Noun, a gross disparity. However, when dissociations are defined by the 50% ellipsoid, 22.3% of all cases fell in the Noun > Verb region and 31.9% fell in the Verb > Noun region. In other words, we find roughly four times more dissociations when the correlation is taken into account, closer to the values that we would expect by the criterion that we chose.

By contrast, the disparity between methods is much smaller for measures with a low correlation. For example, in the case of personal neglect versus the Wundt-Jastrow measure of extrapersonal neglect (which had the lowest correlation among our four pairs of measures, at only +.21), the number of dissociations that we can detect are roughly similar whether or not we take the correlation into account, under virtually any definition. For example, the 70% correlation-based criterion predicts that 15% of the population will fall in the WJ > PN zone and another 15% will fall in the PN > WJ zone. The observed numbers were fairly close: 15.2% and 12.4%, respectively. However, this result (taking the correlation into account) is similar to the numbers observed assuming measurement independence, under the weak criterion in which patients must fall one or more standard deviations below the mean in a single dimension. On this weak criterion, we would expect 12.8% of the population to fall in the WJ > PN region, and another 12.8% in the PN > WJ region. The

observed numbers were 11.7% and 14.5%, respectively.

We have gone into detail only for the two most extreme pairs of measures in our data set. As inspection of Tables 1–4 will show, the other two pairs of measures yield results in between these two extremes – from a large discrepancy between methods that do or do not take the correlation into account, to a relatively small discrepancy. These results suggest that we should think of the correlational issue as a continuum, and worry less about the problem of measurement independence as the correlation between measures approaches the theoretical ideal of zero. However, there is no recommended cut-off here, no correlational threshold below which traditional assumptions of independence can be used. Rather, we recommend that investigators calculate the correlation between measures in some relevant reference population and decide (based on their experimental goals) whether the correlation is high enough to justify adjustments of the kind that we have recommended here.

Who is Dissociated?
As we tried to illustrate in the Fred and Charlie case, the problem of measurement correlation applies not only to the *number* of cases that we can find (i.e., whether we are prone to under- or overestimation), but to the *identity* of the individuals that we designate as "unusual" under a given definition. Even though some correlation-based techniques may yield results numerically similar to those that assume measurement independence, this does not mean that we will always "finger" the same suspects.

To illustrate this point, consider the comparison used above, between the quadrant method (which predicts 25% in each of two dissociative zones, without taking the correlation into account), and the 50% correlation ellipsoid (which predicts 25% on each side of the diagonal, outside of the 50% ellipsoid). As we noted earlier, the quadrant method yields about one-fourth as many dissociations as we would expect between noun and verb naming if the two measures were independent (i.e., 20 instead of 78–80 cases), while the 50% correlation-based method actually yields a slightly larger number than we would

expect by chance when the $+.84$ correlation is taken into account (i.e., 85 instead of 78–80 cases). However, the distribution in Figure 6 shows that a very large number of these "found cases" lie in the High/High and Low/Low quadrants, which are excluded by definition with the quadrant-based method. If we want to preserve the clinical restriction that dissociations should not fall in the relatively spared High/High quadrant or the globally impaired Low/Low quadrant, we can combine the quadrant technique with the 50% correlation ellipsoid (corresponding to the last line in Table 1). This hybrid method is defensible, but it leads to a massive loss in the number of dissociations that are obtained: We can now identify only 14 cases (8.9%), 8 in the Noun > Verb sector and 6 in the Verb > Noun sector, which is actually lower than the yield produced by the quadrant method. This reduction in cases reflects the loss of a handful of patients who are (technically speaking) within the respectively High/Low or Low/High quadrants but also fall within the correlation ellipsoid, relatively close to the means for both measures.

Inspection of Tables 1–4 yields the following generalization: if the correlation between measures is high, then correlation-based techniques are almost always *less conservative* than methods which assume independence (i.e., they identify more cases as "dissociated"), but they also tend to be *more conservative* than methods which assume independence (i.e., they identify fewer cases as "dissociated") when the clinical restriction excluding High/High and Low/Low cases is applied. In other words, the set of individuals whom we identify as "dissociated" can change markedly under these different quantification schemes. If the correlation between measures is low, then these differences tend to disappear, although there are still discrepancies in individual cases (as inspection of Figs. 6–9 will reveal).

Which one of these methods is "right"? Which ones yields "the truth" regarding the individuals who do or do not represent a double dissociation? In order to determine which technique is "right," we need some form of external validation. This might include information about the neural substrates for these profiles (e.g., one criterion maps directly onto lesion type, in

patterns that we would predict based on other kinds of neurological evidence), or it might include some external set of behavioral measures (e.g., the "right" method for classifying patients on any pair of measures is the method that predicts a coherent pattern of behavior on a completely different set of measures). In the absence of such external validation, there is no "right" way to quantify dissociations, no single definition that is appropriate for every research enterprise. Depending upon the investigator's research goals, it may be wise to include as many candidates as possible, or it may be desirable to use relatively stringent criteria so that we only identify cases that would qualify as "dissociations" under virtually any quantification scheme. We will return to this issue later. Our point here is simply to show that quantification decisions matter quite a lot, with individuals moving in and out of the "dissociation zone" depending on the technique that we adopt.

Asymmetries and Deviations from Bivariate Normal

Whether or not we take the correlation into account, all of the techniques that we have described here assume that the data are normally distributed within each measure. We attempted to standardize both distributions, by converting all raw scores (percentages) to z scores to equate the means and variances, but these transformations do not guarantee a normal distribution. If we were successful in assuring a bivariate normal distribution, then we should always obtain symmetrical results (with some degree of sampling error). That is, the number of dissociations should be roughly equivalent on each side of the diagonal, that is, symmetrical. Indeed, if the measures were bivariate normal, then there should always be a close match between the observed and predicted number of dissociations when the correlation is taken into account. Clearly this is not the case. Is this a fatal flaw in the method that we have proposed? At the very least, deviations from the bivariate normal distribution do not invalidate the logic of our argument: in order to determine whether a dissociation is "unusual," we have to have some way of defining the "usual" relationship between two target measures. In pursuit of this goal, there

are several steps that one can take to deal with deviations from bivariate normal. All of them start with visual inspection of the scatterplot between two target measures, with the selected correlation ellipsoids superimposed.

The first remedy applies to deviations that arise from floor and ceiling effects, that is, cases that fall outside the correlation ellipsoid because their performance is "too good" or "too bad" on both of our measures. In research based on experimental methods that are under the investigator's control, floor and ceiling effects can be removed by adding items that are sensitive to individual differences at both extremes of the distribution. If this option is not available (e.g., in research using archival data and/or standardized tests that cannot be altered), the simplest solution is data trimming: remove cases that fall above or below a specified criterion. In the archival examples illustrated above, we repeated all of our analyses with and without those cases that were "true zeroes" (i.e., subjects who were untestable on one or both of our target measures). Results were similar with and without these cases, testimony to the robustness of these results. However, deviations from bivariate normal remained, and visual inspection of our results suggests that these asymmetries are not due to floor and ceiling effects alone. Hence other remedies may be required.

A second remedy applies to systematic variations in the shape of the relationship between two target measures. By definition, correlational techniques test for the fit around a line. And yet there are many situations in which another function is appropriate, for example, the logistic function that governs covariation between vocabulary size and the early development of grammar in young children (Bates & Goodman, 1997). If the relationship is systematic, then it may be possible to "linearize" the relationship by some algebraic transform. Alternatively, the correlation ellipsoids themselves could be tailored to the distribution in question, fitting the space of logistic functions that, for example, govern 90%, 70% or 50% of the cases.

A third remedy acknowledges the possibility that deviations from bivariate normal are real, that is, products of Nature that have been captured faithfully by our measures. Suppose, for example,

that performance on Measure A is mediated by perisylvian regions of the left temporal lobe while performance on Measure B is mediated primarily by regions in the fusiform gyrus. Because of the distribution of the vascular system in the human brain, strokes involving perisylvian cortex are far more likely than strokes restricted to the fusiform gyrus. Hence we should expect to find a significant asymmetry in the distribution of Measures A and B, with more cases of A < B and fewer cases of B < A than we would expect under the bivariate normal assumption. If this is the case, then the asymmetry in our distribution may be viewed as a discovery about Nature rather than a limitation in our method. If we take the third route and decide to live with our distributions as facts of nature, then we can use these asymmetrical distributions as models with which to evaluate data for new individuals or small groups of patients.

SUMMARY AND CONCLUSION

The purpose of this paper was to illustrate some problems that arise in quantifying the important and influential notion of "double dissociations." We have shown that the probability of obtaining a single or double dissociation is strongly influenced by the distribution of scores in the population from which the patients of interest are drawn, including not only the way that "extreme scores" are defined within each measure (i.e., the means and standard deviations) but also the correlation between those measures. We illustrated these points by comparing four pairs of scores, two from a language battery administered to a large sample of patients at risk for aphasia, and two from a nonverbal battery administered to a large sample of patients at risk for one or more forms of neglect. Within each battery, we chose one pair representing one of the highest correlations observed among the available measures, and another representing one of the lowest correlations. We then compared the number of dissociations that were detected under various definitions, with and without taking the correlation into account, with and without the restriction that a dissociated case must always fall in the High/Low or Low/High quadrant.

We found first that the number of dissociations we are able to detect is systematically related to the magnitude of the correlation between measures. If the correlation is low (approaching zero), we are able to find a relatively large number of dissociations. However, this is not necessarily a good thing: low correlations may reflect a true and meaningful form of independence between measures, but they may also reflect high measurement unreliability in one or both of the measures in question. If the correlation is low because measurement reliability is low, then we are at risk for a large number of false positives, reflecting nothing other than chance variation due to measurement error. By contrast, if the correlation between measures is high, then dissociations are much more difficult to detect. Hence we are at risk here for a large number of false negatives, missing dissociations that may be very interesting on theoretical grounds. There is no cure for the problem of false positives other than experimental rigor, detailed information about measurement reliability, and a good theoretical basis for the prediction of extreme scores. But we can defend against some of these false negatives by taking the correlation between measures into account. As we saw, the correlation-based measures illustrated here permit us to find a much larger number of cases if the correlation was high; if the correlation is low, the discrepancy between techniques that assume measurement independence and techniques that take the correlation into account are markedly reduced (though not eliminated).

Our second major conclusion revolves not around the number of dissociations, but the identity of the individuals that we locate with each method. When correlations are high, the use of correlation ellipsoids helps us to detect a larger number of cases, but most of those cases are in the High/High quadrant (patients who are relatively spared on both measures) or the Low/Low quadrant (patients who are globally impaired despite differences in the magnitude of their deficit on each measure). If we adopt a hybrid technique, using a correlation ellipsoid but excluding High/High and Low/Low cases, then we are left only with the High-Low and Low-High quadrants. Compared to methods that do not take the

correlation into account, we will necessarily obtain fewer dissociations than we would detect by assuming measurement independence. This is true because the correlation ellipsoid rules out High-Low and Low-High cases that are technically inside a dissociated quadrant, but are also inside the correlation ellipsoid and thus so close to the means (i.e., the cross-point with z scores of zero on both measures) that the relationship is really not unusual. As we have noted, there is no a priori way to specify whether a given method for identifying dissociations is "right," or whether it identifies too many, too few, or the wrong individuals. This determination can only be made through external validation, using neurological and/or behavioral evidence outside of the immediate data set, with a well-specified set of theoretical goals. Our goal here has simply been to point out how much the correlation matters when one sets out to identify dissociations of interests.

Our third set of conclusions pertains to asymmetries and deviations from normality under any and all of the definitions of dissociation that we have tested here. All of these techniques (including those that fail to take the correlation into account) assume that measures are normally distributed (bivariate normal in the case of the correlation-based techniques). The world in which neuropsychologists carry out their work rarely corresponds to this assumed perfection. Some of these deviations may prove to be important and meaningful, reflecting the structure of a brain that is certainly not random in its construction. However, the deviations from normality and the asymmetrical dissociations that we uncovered in the examples illustrated here were usually small, and they were inconsistent in direction across different quantification schemes. Hence it seems likely (for these examples) that these deviations from bivariate normal reflected sampling error, including the kinds of ceiling and floor effects that often occur in neuropsychological research. Fortunately remedies are available to help to deal with these deviations.

As a final point, we want to underscore the importance of the reference population that is chosen to calculate the correlation between measures. For our purposes here, we used two large samples of patients, one representing all

consecutive cases across a three-year period who were at risk for language disorders, and the other representing all consecutive cases in an equivalent period who were at risk for disorders of spatial attention. These two pools of patients were likely to yield dissociations in multiple directions, among language measures and visuospatial measures, respectively, and hence appeared to us to furnish an excellent starting point for the illustrations provided here. But other reference populations may be more appropriate, depending on the goals of the experimenter. For the measures that were adopted here, healthy adults tend to score so close to ceiling that we would have too little variance to calculate meaningful correlations. But there are many other forms of measurement (e.g., intelligence testing) for which this is certainly not the case. Furthermore, although the correlations among intelligence scales do tend to be high when calculated across large samples of normal controls, many normal individuals can be found who display discrepancies of two standard deviations or more between their "best" and "worst" subscales. Hence it would indeed be valuable in some cases to construct correlation ellipsoids for normal controls as the basis for evaluating the probability of a given dissociation in clinical populations. Alternatively, we may want to use large and truly unselected samples of brain-injured patients as our reference group, pooling, for example, across left- and right-hemisphere cases without making prior assumptions about the pool in which language dissociations or spatial dissociations are likely to occur. All of these approaches could be valid, but each one would lead to different interpretations, and are likely to be used for different purposes.

We also would like to note that the applications of these techniques are not limited to very large archival datasets and that the methods are applicable also to experimental studies with more modest sample sizes. Indeed, we have successfully applied correlation-based outlier analyses of the sort described here to data collected from 30 aphasic patients in an experiment which aimed to examine neural correlates of nonverbal auditory processing in relation to speech comprehension (Saygin, Dick, Wilson, Dronkers, & Bates, 2003). In this sample, we were able to reliably identify

with the techniques depicted here a patient with a rare type of auditory agnosia and further testing confirmed the analyses (Saygin & Moineau, 2002).

To conclude, the particular methods illustrated here for the identification and quantification of double dissociations have implications for at least three aspects of neuropsychological research.

First, because double dissociations are often used to draw inferences about the architecture of the normal mind/brain, we have provided some techniques that investigators may want to use to protect against false negatives as well as false positives, locating potentially interesting cases that might not be evident if the correlation between variables were not taken into account. We are not claiming that any single quantification scheme is right or wrong. Indeed, it is debatable whether there really is such a thing as a "true dissociation" independent of anyone's theoretical assumptions. One might argue, for example, that the "right" definition of a behavioral dissociation is the one that maps most directly and transparently onto the neural substrate, matching predictions regarding lesion type. For example, if it is the case (as some neuroscientists have argued) that verb meanings are mediated by the frontal lobes to a greater extent than noun meanings, then the best method for quantifying noun-verb dissociations might be the one that correlates best with a distinction between anterior and posterior brain injury. Alternatively, the "right" definition for any two behavioral measures may be the one that generalizes reliably and robustly to alternative measures of the same underlying construct. For example, the best metric for identifying noun-verb dissociations on a naming task might be the method that correlates most highly with noun versus verb production in free speech. In the absence of such external validation, the investigator must determine the method of choice based on his or her scientific goals: would it be better to choose a conservative method, or to cast a wider net in search of potentially interesting cases?

Furthermore, these methods may provide additional constraints and new avenues for lesion-symptom mapping. With the advent of new techniques for functional and structural brain imaging, clinical researchers have the opportunity to study lesions in vivo, including structurally

intact regions that are functionally silent (or hypometabolic) during linguistic and cognitive tasks. However, because these applications are expensive, procedures for patient selection become more critical than ever. Instead of basing patient selection solely on extreme scores, a correlation-based method can be used to select candidates whose behavioral profiles are unusual under virtually any definition, increasing the likelihood of a strong positive result in an imaging study.

Third, these correlation-based methods may be a useful adjunct in patient diagnosis, and in the evaluation of recovery (with or without rehabilitation). In pursuit of this goal, we are currently applying these techniques to several large data sets for aphasic patients in English, Italian and Chinese, re-evaluating the relationship between classic aphasia subscales and their use in patient classification when the correlation between scales is taken into account (e.g., see Bates, Saygin, Moineau, Marangolo, & Pizzamiglio, 2002).

ACKNOWLEDGMENTS

REFERENCES

Albert, M.L. (1973). A simple test of visual neglect. *Neurology*, *23*, 658–664.

Appelbaum, M. (2000). *Computations involving the multi-variate normal* (QSL Tech. Report 0400). San Diego, CA.

Bates, E., Appelbaum, M., & Allard, L. (1991). Statistical constraints on the use of single cases in neuropsychological research. *Brain and Language*, *40*, 295–329.

Bates, E., Chen, S., Tzeng, O., Li, P., & Opie, M. (1991). The noun-verb problem in Chinese aphasia. *Brain and Language*, *41*, 203–233.

Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, *12*, 507–586.

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

Bates, E., Saygin, A.P., Moineau, S., Marangolo, P., & Pizzamiglio, L. (2002) Aphasic patients as vectors in a multidimensional symptom space: How to change the bathwater without losing the baby. *Brain and Language*, *83*, 223–224.

Bates, E., Wulfeck, B., & MacWhinney, B. (1991). Crosslinguistic research in aphasia: An overview. *Brain and Language*, *41*, 123–148.

Berndt, R., & Zingeser, L. (1991). Grammatical class effects in word production: Finding the locus of the deficit. *Brain and Language*, *41*, 597–600.

Bishop, D.V.M. (1997). *Uncommon understanding: Development and disorders of comprehension in children.* Hove, UK: Psychology Press.

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

Bouillaud, M.J. (1825). Recherches cliniques propres à démontrer que la part de la parole correspond à la lésion des lobules antérieurs du cerveau, et à confirmer l'opinion de M. Gall, sur le siège de l'organe du langage articulé. *Arch. Générales de Médicine*, *3*, 25–45.

Caplan, D., & Hildebrandt, N. (1988). *Disorders of syntactic comprehension.* Cambridge, MA: MIT Press.

Cappa, S., & Ullman, M.T. (1998). *A neural dissociation in Italian verbal morphology.* Paper presented at the 5th Annual Meeting of the Cognitive Neuroscience Society, San Francisco.

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

Caramazza, A., & Hillis, B. (1991). Lexical organization of nouns and verbs in the brain. *Nature*, *349*, 788–790.

Chen, S., & Bates, E. (1998). The dissociation between nouns and verbs in Broca's and Wernicke's aphasia: Findings from Chinese. Special issue on Chinese aphasia. *Aphasiology*, *12*, 5–36.

Ciurli, P., Marangolo, P., & Basso, A. (1996). *Esame del linguaggio – II.* [Examination of Language – II]. Florence: O.S. Organizzazione Speciali, S.r.l.

Damasio, A., Grabowski, T.J., Tranel, D., Hichwa, R.D., & Damasio, A.R. (1996). A neural basis for lexical retrieval. *Nature*, *380*, 499–505.

De Renzi, E., & Vignolo, L. (1962). The Token Test: A sensitive test to detect receptive disturbances in aphasics. *Brain*, *85*, 665–678.

Diller, L., & Weinberg, J. (1977). Hemi-inattention in rehabilitation: The evaluation of a rational remediation program. In E.A. Wainstain & R.P. Friedland (Eds.), *Hemi-inattention and hemisphere specialization* (pp. 62–82). New York: Raven Press.

Dunn, J., & Kirsner, K. (1988). Discovering functionally independent mental processes: The principle of reversed association. *Psychological Review*, *95*, 91–101.

Dunn, J., & Kirsner, K. (2003). Editorial: What can we infer from double dissociations? with discussion by A. Baddeley, C.T. Kello, P. Juola, D.C. Plaut, J.A. Bullinaria, J.C. Lyons, G.D.A. Brown, K. Lamberts, W.B. Barr, E. Goldberg, D. Andrews, A. Karmiloff-Smith, G. Scerif, D. Ansari, G. Van Orden, H. Kloos, N. Chater, F.L. Bedford, E.M. Reingold, J.C. Dunn, S. Sternberg, H. Kadlec, I. van Rooij, M. Coltheart, M. Davis, J.M. Gurd, J.C. Marshall, M. McCloskey. *Cortex 39*, 1–9 and 129–202.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press/Bradford Books.

Galati, G., Lobel, E., Vallar, G., Berthoz, A., Pizzamiglio, L., & Le Bihan, D. (2000). The neural basis of egocentric and allocentric coding of space in humans: A functional magnetic resonance study. *Experimental Brain Research*, *133*, 156–164.

Gall, F.J., & Spurzheim, J.C. (1809). *Recherches sur le système nerveux en général, et sur celui du cerveau en particulier: Mémoire présenté a l'Institut de France, le 14 Mars 1808, suivi d'observations sur le rapport* [Anatomy and physiology of the nervous system in general and of the brain in particular: A memoir presented at the Institute of France on 14 March, 1808, followed by comments on the report]. Paris: F. Schoell, H. Nicolle.

Goodglass, H. (1993). *Understanding aphasia.* San Diego: Academic Press.

Goodglass, H., & Kaplan, H. (1983). *Boston diagnostic aphasia examination* (2nd ed.). Philadelphia: Lea and Febiger.

Guariglia, C., & Antonucci, G. (1992). Personal and extrapersonal space. A case of neglect dissociation. *Neuropsychologia*, *30*, 1001–1009.

Kinsbourne, M. (1971). Cognitive deficit: Experimental analysis. In J.L. McGaugh (Ed.), *Psychobiology*. New York: Academic Press.

Kertesz, A. (1982). *Western aphasia battery.* New York: Grune & Stratton.

Leonard, L.B. (1998). *Children with specific language impairment*. Boston MA: MIT Press.

Luzzatti, C., & Whitaker, H. (2001). Jean-Baptiste Bouillaud, Claude François Lallemand and the role of the frontal lobe: Location and dislocation of language in the early 19th century. *Archives of Neurology*, *58*, 1157–1162.

Marin, O.S.M., Saffran, E.M., & Schwartz, M.F. (1976). Dissociation of language in aphasia: Implications for normal functions. *Annals of the New York Academy of Sciences*, *280*, 868–884.

Massironi, M., Antonucci, G., Pizzamiglio, L., Vitale, M.V., & Zoccolotti, P. (1988). The Wundt-Jastrow illusion in the study of spatial hemi-inattention. *Neuropsychologia*, *26*, 161–166.

Norman, D., & Shallice, T. (1983). Attention to action: Willed and automatic control of behavior. In R.J. Davidson, G.E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (Vol. IV). New York: Plenum Press.

Pizzamiglio, L., Bergego, C., Halligan, P., Homberg, V., Robertson, I., Weber, E., Wilson, B., Zoccolotti, P., & Deloche, G. (1992). Factors affecting the clinical measurement of visuo-spatial neglect. *Behavioural Neurology*, *5*, 233–240.

Saygin, A.P., Dick, F., Wilson, S., Dronkers, N., & Bates, E. (2003). Neural resources for processing language and environmental sounds: Evidence from aphasia. *Brain*, *126*, 928–945.

Saygin, A.P., & Moineau, S. (2002). Auditory agnosia with preserved verbal comprehension after unilateral left hemisphere lesion involving Wernicke's area. *Society for Neuroscience Abstracts*, *28*, 673.7.

Schuell, H., & Jenkins, J.J. (1959). The nature of language deficit in aphasia. *Psychological Review*, *66*, 45–67.

Shallice, T. (1988). *From neuropsychology to mental structure.* New York: Cambridge University Press.

Teuber, H.L. (1955). Physiological psychology. *Annual Review of Psychology*, *6*, 267–297.

Tomblin, J., & Pandich, J. (1999). Lessons from children with specific language impairment. *Trends in Cognitive Sciences*, *3*, 283–285.

Vallar, G. (1999). The methodological foundations of neuropsychology. In F. Denes & L. Pizzamiglio (Eds.), *Handbook of clinical and experimental neuropsychology.* Hove, UK: Psychology Press.

Weiskrantz, L. (1968). Some traps and pontifications. In L. Weiskrantz (Ed.), *Analysis of behavioral change*. New York: Harper & Row.

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

Zoccolotti, P., & Judica, A. (1991). Functional evaluation of hemi-neglect by means of a semistructured scale: Personal-extrapersonal differentiation. *Neuropsychological Rehabilitation*, *1*, 33–44.