Neuropsychology
2005, Vol. 19, No. 5, 664–678

0894-4105/05/$12.00 DOI: 10.1037/0894-4105.19.5.664

Evaluation of Criteria for Classical Dissociations in Single-Case Studies by Monte Carlo Simulation

John R. Crawford
University of Aberdeen

Paul H. Garthwaite
The Open University

The conventional criteria for a classical dissociation in single-case studies require that a patient be impaired on one task and within normal limits on another. J. R. Crawford and P. H. Garthwaite (2005) proposed an additional criterion, namely, that the patient's (standardized) difference on the two tasks should differ from the distribution of differences in controls. Monte Carlo simulation was used to evaluate these criteria. When Type I errors were defined as falsely concluding that a control case exhibited a dissociation, error rates were high for the conventional criteria but low for Crawford and Garthwaite's criteria. When Type I error rates were defined as falsely concluding that a patient with equivalent deficits on the two tasks exhibited a dissociation, error rates were very high for the conventional criteria but acceptable for the latter criteria. These latter criteria were robust in the face of nonnormal control data. The power to detect classical dissociations was studied.

Keywords: single-case studies, statistical methods, dissociations

The past few decades have witnessed a significant resurgence of interest in single-case studies. In such case studies, the demonstration of a deficit on a given cognitive task usually becomes of theoretical interest only if it is observed in the context of less impaired or normal performance on other tasks. That is, much of the focus in single-case studies is on establishing dissociations of function (Caramazza & McCloskey, 1988; Ellis & Young, 1996; Shallice, 1988; Vallar, 2000). These dissociations provide evidence of the fractionation of cognitive abilities and thus serve as building blocks in the attempt to develop models of the functional architecture of human cognition.

Typically, a classical dissociation (Shallice, 1988, p. 227) is defined as occurring when, with reference to the performance of matched healthy controls or a healthy normative sample, a patient is "impaired" or shows a "deficit" on Task X but is "not impaired," "normal," or "within normal limits" on Task Y. For example, Ellis and Young (1996) stated, "If patient X is impaired on Task 1 but performs normally on Task 2, then we may claim to have a dissociation between tasks" (p. 5). Similarly, Coltheart (2001) stated that a dissociation is established when a patient "is impaired on Task X but normal on Task Y" (p. 12).

Crawford, Garthwaite, and Gray (2003) identified a number of problems with these conventional definitions. First, one half of the typical definition of a dissociation essentially requires one to prove the null hypothesis (one must demonstrate that a patient is not different from controls or normative samples), whereas, as is well known, one can only fail to reject it. This is particularly germane to single-case studies as the power to reject the null hypothesis will be low: An individual patient (rather than a group of patients) is compared with either a large normative sample or a matched control group. In the latter case, there is a further factor serving to reduce power: namely, that the control sample in such studies is itself often modest in size (Crawford, Garthwaite, & Gray, 2003).

The second related problem with the conventional definitions is that a patient's score on the "impaired" task could lie just below whatever cutoff is used to define impairment, and the performance on the other test could lie just above this cutoff. That is, the difference between the patient's relative standing on the two tasks of interest could be very trivial; in this situation, one would not want to infer the presence of a dissociation, despite the results of any formal statistical tests. Crawford, Garthwaite, and Gray (2003) and Crawford and Garthwaite (2005) argued that one therefore needs to impose an additional criterion for a dissociation that focuses on testing the difference between the patient's relative standing on the two tasks.

This criterion not only deals with the problem of trivial differences, it also provides a positive test for a classical dissociation; otherwise one must rely heavily on what boils down to an attempt to prove the null hypothesis of no deficit on Task Y.

Crawford and Garthwaite's (2005) criteria for classical dissociations are designed to be used in studies in which a patient is compared with a control sample (rather than with normative data). These criteria are based on the pattern of results obtained from the application of three inferential tests. A modified *t* test advocated by Crawford and Howell (1998) is used to test whether the patient has a deficit on Tasks X and Y, and the Revised Standardized Difference Test (RSDT; Crawford & Garthwaite, 2005; Garthwaite & Crawford, 2004) is used to test the difference between the patient's performance on Tasks X and Y (the standardized difference for the patient is compared with the distribution of standardized differences obtained from the controls). Thus, if a patient is significantly different from controls on either Task X or Task Y ($p < .05$,

John R. Crawford, School of Psychology, College of Life Sciences and Medicine, King's College, University of Aberdeen, Aberdeen, United Kingdom; Paul H. Garthwaite, Department of Statistics, The Open University, Milton Keynes, United Kingdom.

We are grateful to Adelchi Azzalini of the University of Padua for his assistance with the methods used to sample from bivariate skew distributions.

Correspondence concerning this article should be addressed to John R. Crawford, School of Psychology, College of Life Sciences and Medicine, King's College, University of Aberdeen, Aberdeen AB24 2UB, United Kingdom. E-mail: j.crawford@abdn.ac.uk

one-tailed), but not both, and his or her standardized difference is significantly different from that of controls ($p < .05$, two-tailed), then he or she is classified as exhibiting a classical dissociation.

One-tailed tests are used to test for a deficit because the interest lies in testing the null hypothesis against the alternative directional hypothesis that the patient's performance is poorer than controls' and because the possibility that neurological damage improves cognitive performance can be discounted except in very unusual circumstances. Two-tailed tests are used to test for a difference between Tasks X and Y because the direction of the difference is not subject to this latter logical constraint; that is, the difference may reflect the effect of neurological damage on either Task X or Task Y (it is not denied that a researcher may have an a priori hypothesis concerning which task will be affected but, unlike when testing for a deficit, he or she cannot rule out of court the possibility that the reverse pattern will occur).

Crawford, Garthwaite, and Gray (2003) ran a Monte Carlo simulation to estimate the percentage of control cases that would be incorrectly classified as exhibiting a dissociation when their criteria were applied. The results were encouraging in that the percentage of control cases classified as exhibiting a classical dissociation was low (well below 5% at all values of N for the control samples and at all values of the population correlation between Tasks X and Y).

Our first aim in the present study was to subject the conventional criteria for a classical dissociation to similar scrutiny. If the arguments made by Crawford, Garthwaite, and Gray (2003) are sound, then the conventional criteria will not perform well; that is, they should misclassify an inappropriately high proportion of individuals drawn from the healthy population as exhibiting a classical dissociation. However, this issue should be examined empirically.

In attempting to implement a simulation to evaluate the conventional criteria for a classical dissociation, we come up against an additional problem. The typical definitions offer little or no guidance as to how one should determine whether a patient does or does not meet the criteria. That is, they do not specify how one should determine if a patient is either impaired or within normal limits on Task X or Task Y. Therefore, it is necessary to consult the empirical literature on single-case studies in which a patient is compared with a control sample.

By far the most common method of forming inferences about the presence of a deficit in such studies is to convert the patient's score on a given task to a z score based on the mean and standard deviation of the controls and then to refer this score to a table of the areas under the normal curve. As these studies often use a one-tailed test, z scores that fall below -1.645 are regarded as statistically significant ($p < .05$) and are taken as evidence that the patient has a deficit on a given task. Scores higher than this critical value are taken as an indication that a patient's score is "not impaired," "normal," or "within normal limits."

Among the many examples of this approach are a series of influential studies that have examined recognition of emotions in patients with bilateral amygdala damage (e.g., Calder, Young, Rowland, Perret, & Hodges, 1996; Scott et al., 1997). These studies used z to compare an individual patient's performance with that of a healthy control sample on tasks assessing the ability to recognize emotions from facial expressions or tone of voice. For example, in the Calder et al. (1996) study, the performance of a patient, D.R., was classified as being "unimpaired" (p. 707) or "normal" (p. 708) for recognition of happiness, disgust, and so forth but was classified as "impaired" (p. 707) for recognition of fear on the basis of whether z was less than -1.65.

A difficulty with this approach to detecting deficits lies in the use of z to form inferences as this involves treating the control sample as if it was a population; that is, the mean and standard deviation are used as if they were population parameters rather than sample statistics. This would not be too serious a problem if the control sample was large, as it should yield sufficiently accurate estimates of the parameters. However, the control samples in single-case studies in cognitive neuroscience typically have modest Ns; $N < 10$ is not unusual and Ns < 20 are common (Crawford & Garthwaite, 2002; Crawford & Howell, 1998).

The practical effect of using z with modestly sized control samples will be to exaggerate the rarity or abnormality of a patient's score and to inflate the Type I error rate. In this context, a Type I error occurs when a case that is drawn from the control population is incorrectly classified as not being a member of this population; that is, the case is incorrectly classified as exhibiting a deficit. Crawford and Garthwaite's (2005) solution to this problem was to use the modified t test referred to earlier (Crawford & Howell, 1998) because this method treats the control sample statistics as sample statistics. The method uses the t distribution (with $n - 1$ degrees of freedom), rather than the standard normal distribution, to test whether the patient's scores are significantly lower than the scores of the control sample.

In view of the foregoing, one should make a clear distinction between the conventional criteria for a classical dissociation and the method used to determine whether these criteria have been met. That is, suppose that empirical examination of these criteria suggests that they are inadequate; one should seek to determine whether this stems either from a fundamental problem with the criteria themselves or from the use of z as an inferential method (of course both aspects could be problematic).

Thus, although z is normally used to test whether the conventional criteria are met, they can also be tested using Crawford and Howell's (1998) method. This allows one to examine the performance of the conventional criteria free of the potentially detrimental effect of using z.

Study 1: Monte Carlo Simulation of Type I Errors When Applying Criteria for a Classical Dissociation

In Study 1, we ran a Monte Carlo simulation to quantify and compare control of the Type I error rate when the alternative criteria for a classical dissociation were applied. As noted, a Type I error will occur if we misclassify a member of the control population as exhibiting a classical dissociation. Three sets of criteria were evaluated: Crawford and Garthwaite's (2005) criteria, the conventional criteria using z to identify the presence or absence of deficits on Tasks X and Y, and the conventional criteria using Crawford and Howell's (1998) modified t test in place of z.

Method

The Monte Carlo simulation was run on a PC and implemented in Borland Delphi (Version 4). The algorithm ran3.pas (Press, Flannery, Teukolsky, & Vetterling, 1989) was used to generate uniform random numbers (between 0 and 1), and these were transformed by the polar variant of the Box–Muller method (Box & Muller, 1958). The Box–Muller transformation generates pairs of normally distributed observations, and by further transforming the second of these, it is possible to generate observations from a bivariate normal distribution with a specified correlation (e.g., see Kennedy & Gentle, 1980); more detail is provided in the Appendix.

The simulation was run with five different values of N (the sample size of the control sample): For each of these values of N, 1 million samples of $N + 1$ were drawn from one of five bivariate normal distributions in which the population correlation (ρ_{XY}) was set at .0, .3, .5, .7, or .8. Thus, a total of 25 million individual Monte Carlo trials were run.

In each trial, the first N pairs of observations were taken as the control sample's scores on Tasks X and Y and the $N + 1$th pair was taken as the scores of the individual control case. Crawford and Garthwaite's (2005) criteria were then applied to these data. That is, cases were classified as exhibiting a dissociation if (a) they obtained a significantly lower score ($p < .05$, one-tailed) than controls on either Task X or Task Y (but not both) using Crawford and Howell's (1998) test and (b) their standardized difference between Tasks X and Y was significantly different ($p < .05$, two-tailed) from the standardized differences in the control sample using Crawford and Garthwaite's (2005) RSDT.

For the conventional criteria using z, a control case was recorded as exhibiting a dissociation if z was less than -1.645 (i.e., nominal $p = .05$, one-tailed) on either Task X or Task Y (but not both). For the conventional criteria using Crawford and Howell's (1998) test, a control case was recorded as exhibiting a dissociation if he or she obtained a significantly lower score ($p < .05$, one-tailed) than controls on either Task X or Task Y (but not both). Note that, in the latter case, these criteria differ from Crawford and Garthwaite's (2005) criteria only in that they do not require a significant difference between a patient's scores on Tasks X and Y. For all three sets of criteria, the percentage of control cases incorrectly classified as exhibiting a dissociation was recorded.

Results and Discussion

The basic pattern of results for the three sets of criteria can be clearly appreciated by examining Figure 1. Because the simulations varied both the size of the control sample (N) and the population correlation between tasks (ρ_{XY}), representing all of these results in a single figure would be messy. We have therefore presented the results for $\rho_{XY} = .5$ only; the pattern in Figure 1 would be more exaggerated for $\rho_{XY} < .5$ and less extreme for $\rho_{XY} > .5$. The full results from the simulation are presented in Table 1.

It can be seen from Table 1 that, replicating the findings of Crawford and Garthwaite (2005), Type I errors (i.e., wrongly classifying control cases as exhibiting dissociations) are low when their criteria for a classical dissociation are applied. The error rate ranges from a low of 0.84% (for a control sample N of 100 and a correlation between Tasks X and Y of .8) to a high of 2.32% (for a control sample N of 5 and a ρ_{XY} of .0).

In contrast, the Type I error rates were much higher when the conventional criteria for a classical dissociation were applied. When, as is common in practice, z was used to test for a significant difference between a case and controls, the Type I error rate ranged from a low of 5.31% (for an N of 100 and a ρ_{XY} of .8) to a high of 18.55% (for an N of 5 and a ρ_{XY} of .0). The Type I error rates for the conventional criteria were still high in absolute terms when Crawford and Howell's (1998) test was used to test for deficits on Tasks X and Y, but they were appreciably lower than the equivalent misclassification rates obtained using z. The Type I error rate ranged from a low of 5.10% (for an N of 100 and a ρ_{XY} of .8) to a high of 9.52% (for an N of 10 and a ρ_{XY} of .0). It can also be seen that the disparity in the results for the two sets of conventional criteria diminishes as the N of the control sample increases; this occurs because z and t converge as sample size increases. However, it will be appreciated that control sample Ns of this magnitude (i.e., $\geq$ 50) are rare in single-case studies in cognitive neuroscience.

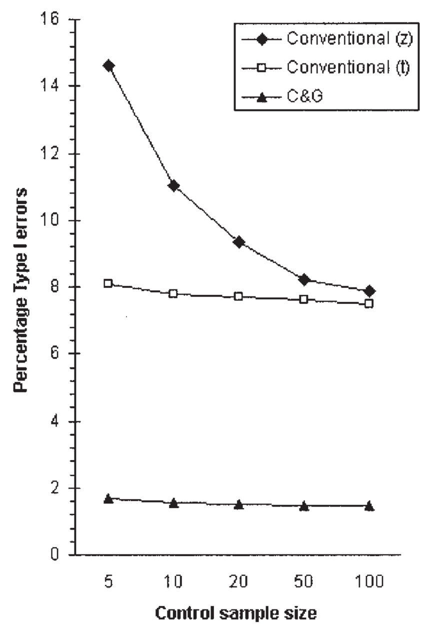

Figure 1. Results from a Monte Carlo simulation: Percentage of Type I errors for three sets of criteria for classical dissociations (for these data $\rho_{XY} = .5$). C&G = Crawford and Garthwaite (2005).

In conclusion, these results indicate that, if Crawford and Garthwaite's (2005) criteria are applied in single-case research, it would be unlikely that a member of the control (i.e., healthy) population would be misclassified as exhibiting a classical dissociation. That is, their criteria are conservative. In contrast, if the conventional criteria are applied (regardless of whether these are operationalized using z or Crawford & Howell's, 1998, test), a substantial minority of normal controls would be misclassified as exhibiting a classical dissociation. Misclassifications will be particularly frequent if z is used with the modest control sample sizes that are typical in single-case studies.

Study 2: Monte Carlo Simulation on the Effects of Nonnormal Control Data on Criteria for Classical Dissociations

The Monte Carlo simulations performed in Study 1 were based on sampling from a normal distribution. Furthermore, the inferential methods used to test whether the conventional criteria and Crawford and Garthwaite's (2005) criteria were met all assume normality; that is, the use of z assumes normality, as does Crawford and Howell's (1998) modified t test and Crawford and Garthwaite's (2005) RSDT.

Ideally, researchers would carefully select the measures they use in single-case studies so as to avoid potential problems arising

Table 1
Results From a Monte Carlo Simulation Study: Percentages of Control Cases Incorrectly Classified as Exhibiting a Classical Dissociation as a Function of the Criteria Applied, the N of the Control Sample, and the Correlation Between Tasks

	Crawford and Garthwaite (2005) criteria					Conventional criteria using t					Conventional criteria using z				
N	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8
5	2.32	1.91	1.67	1.34	1.15	9.51	8.84	8.09	6.80	5.80	18.55	16.56	14.62	11.79	9.83
10	2.41	1.88	1.55	1.21	0.97	9.52	8.69	7.78	6.42	5.39	14.02	12.52	11.01	8.89	7.41
20	2.46	1.89	1.53	1.11	0.90	9.52	8.60	7.72	6.21	5.22	11.74	10.48	9.34	7.45	6.24
50	2.47	1.87	1.47	1.09	0.86	9.50	8.57	7.58	6.15	5.10	10.37	9.33	8.20	6.64	5.50
100	2.48	1.86	1.47	1.07	0.84	9.51	8.57	7.54	6.14	5.10	9.95	8.95	7.85	6.39	5.31

from nonnormal control data. However, for many published single-case studies, it is clear from even a cursory inspection of the control sample means and standard deviations that the control data are negatively skewed. That is, the standard deviations show that, were the data normally distributed, a substantial percentage of scores would lie above the maximum obtainable score on a particular task, yet we know that this is impossible; hence, the data must be heavily skewed.

Skew will be almost inevitable when the tasks used measure abilities that are largely within the competence of most healthy individuals. In this situation, negative skew will occur when the measure of interest is based on the number of items passed (i.e., there will be ceiling effects), and positive skew will occur when the measure is an error rate (i.e., there will be floor effects). Evidence of severely skewed control data can be found in the literature on recognition of facial expression of emotion referred to earlier and also in the extensive single-case literature on category-specific object naming. For example, in a recent review of single-case studies of the living versus nonliving distinction, it was reported that the accuracy of naming in controls was in excess of 95% in the vast majority of these studies (Laws, in press).

In view of the foregoing, it is important to examine the extent to which the criteria for classical dissociations are robust in the face of skewed control data. That is, one should attempt to quantify the level of control over the Type I error rate that will be achieved when these criteria are applied to skewed data. Investigation of this issue necessarily requires more complicated modeling than that used in Study 1. To model the range of scenarios that will arise in single-case research, it is still necessary to vary the sample size and the correlation between tasks (as in Study 1), but in addition, for each of these combinations, it is also necessary to vary the degree of skew. Moreover, in some case studies, the control data for both Tasks X and Y will be skew, whereas, in others, the data for only one of the tasks will be skew. The potential effects on Type I errors are liable to be different in these two scenarios.

Another potential problem that will arise in the conduct of single-case research is that the distribution of control data will be overly peaked and have heavier tails than a normal distribution. That is, in practice, the distribution of the control data may be leptokurtic and thereby inflate the Type I error rate (Garthwaite & Crawford, 2005). The effects of leptokurtic control data on Type I errors for classical dissociations were also examined in Study 2.

Method

Sampling from bivariate skew distributions. The method used to form skew bivariate distributions was based on that of Azzalini and colleagues (Azzalini & Capitanio, 1999; Azzalini & Dalla Valle, 1996). As the method is technical and is liable to be of limited interest to most neuropsychologists, we have consigned the description of the procedure followed to the Appendix.

To investigate the effects of skew across a range of scenarios, we formed four sets of skew distributions. In Set 1, Task X was moderately skew ($\gamma_1 = -0.3$) and Task Y was normal; in Set 2, Task X was severely skew ($\gamma_1 = -0.5$) and Task Y was normal; in Set 3, both Tasks X and Y were moderately skew; and in Set 4, both Tasks X and Y were severely skew. For each of these sets, we varied the correlation between Tasks X and Y using the same range as in Study 1 (i.e., .0, .3, .5, .7, and .8). It should be noted, however, that the degree of skew introduced imposes limits on the correlation between Tasks X and Y (Azzalini & Dalla Valle, 1996). For example, if Task X is severely skew and Task Y is normal, then the maximum achievable correlation between Tasks X and Y is .61. Therefore, in this particular scenario, it was not possible to sample from bivariate distributions in which ρ_{XY} was .7 or .8.

Sampling from heavy-tailed (leptokurtic) distributions. The most common approach to modeling the effects of leptokurtic distributions on test statistics is to sample from t distributions rather than from a normal distribution (Lange, Little, & Taylor, 1989). In the present study, we sampled from bivariate t distributions on 7 (moderate leptokurtosis) and 4 (severe leptokurtosis) degrees of freedom. Kurtosis (β_2) is 5 for a t distribution on 7 degrees of freedom; the kurtosis for a t distribution on 4 degrees of freedom is even more extreme but is undefined (because the denominator in the formula for kurtosis requires subtracting 4 from the degrees of freedom and is hence zero).

To sample from these bivariate t distributions, $N + 1$ observations were drawn from bivariate normal distributions (using the same values of N and ρ_{XY} as in Study 1). Each pair of observations was then divided by $\sqrt{\chi^2/7}$ or $\sqrt{\chi^2/4}$, where χ^2 is a random draw from a chi-square distribution on 7 or 4 degrees of freedom, respectively; the resultant vectors are observations from bivariate t distributions on 7 or 4 degrees of freedom.

Results and Discussion

Effects of skew on Type I errors. With regard first to the effects of skew, the basic pattern of results for Crawford and Garthwaite's (2005) criteria can be clearly appreciated by examining Figure 2. This figure presents the results for $\rho_{XY} = .5$ only but includes the results obtained when sampling from the equivalent bivariate normal distribution (Study 1) for comparison purposes. The full set of results for Crawford and Garthwaite's criteria are presented in Table 2.

It can be seen that the Type I error rates are relatively modest even when skew is severe; rates range from 0.79% to 3.20%, and the error rate is well below 3% in the vast majority of cases. It can also be seen that the effects of skew are dependent on whether the distribution of only one of the tasks is skew or whether both are. When both tasks are skew, the error rates are lower, albeit mar-

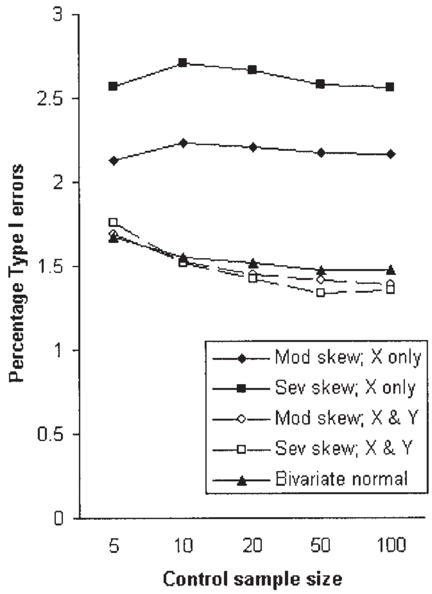

Figure 2. Results from a Monte Carlo simulation: Effect of skew on Type I errors for Crawford and Garthwaite's (2005) criteria for classical dissociations (for these data, ρ_{XY} = .5). mod = moderate; sev = severe.

ginally, than the rates obtained in Study 1 when sampling from a bivariate normal distribution (given the number of simulations performed, this pattern is not a chance finding arising from Monte Carlo variation). In contrast, when only one of the tasks is skew, many of the error rates are close to double those obtained when sampling from a bivariate normal distribution. This effect is most apparent when the tasks are highly correlated (as can be seen by comparing the results in Table 1 with those for skew on Task X in Table 2).

The Type I error rates obtained when the conventional criteria for a classical dissociation were applied (using z to test for deficits) are presented in Table 3. The error rates range from a low of 5.04% to a high of 19.70%. It can be seen that, as was the case when sampling from a bivariate normal distribution, all error rates are much higher than the equivalent results for Crawford and Garthwaite's (2005) criteria. However, the basic pattern of error rates matches these former criteria in that skew exerted a marked effect only when it was present in only one of the tasks. When Crawford and Howell's (1998) test was used to determine whether the conventional criteria were met, the effects of skew mirrored those found using the other two sets of criteria. The absolute values of the error rates, however, are intermediate between these two other sets of criteria; rates range from 4.89% to 10.83%. Full details of these latter results can be obtained on request from John R. Crawford.

Effects of kurtosis on Type I errors. The effects of sampling the control data from leptokurtic (heavy-tailed) distributions are presented in Table 4 (moderately leptokurtic) and Table 5 (severely leptokurtic). These tables present the Type I error rates for Crawford and Garthwaite's (2005) criteria and for the conventional criteria using either z or Crawford and Howell's (1998) method to test for deficits.

For Crawford and Garthwaite's (2005) criteria, it can be seen that the Type I error rates are relatively modest even when sampling from severely leptokurtic distributions; the error rates range from a low of 1.15% (in the case of moderate kurtosis, an N of 100, and a ρ_{XY} of .8) to a high of 3.68% (severe kurtosis, an N of 10, and a ρ_{XY} of 0). The error rate is well below 3% in the vast majority of cases. However, when these results are compared with those from Study 1, it can be seen that the presence of kurtosis has nevertheless markedly raised the error rates over those obtained when sampling from a bivariate normal distribution.

The error rates for either of the two sets of conventional criteria are much higher than those for Crawford and Garthwaite's (2005) criteria, but again, when these results are compared with those obtained when sampling from a normal distribution (Study 1), it can be seen that this stems almost entirely from problems inherent in the conventional criteria rather than from the presence of kurtosis in the control data. Indeed, the rates for the conventional criteria are lower, albeit modestly, than those obtained for these criteria when sampling from a bivariate normal distribution.

In conclusion, the present results suggest that the presence of skew in control data will not grossly inflate the Type I error rate in single-case studies regardless of the criteria applied to detect dissociations. These results are particularly reassuring given that, as noted, heavily skewed control data are a common feature of many existing single-case studies. However, it remains the case that when the distribution of one (but only one) of the tasks is skew, Type I error rates will be raised. That is, the probability of

Table 2
Monte Carlo Evaluation of the Robustness of Crawford and Garthwaite's (2005) Criteria for Classical Dissociations: Percentages of Control Cases Incorrectly Classified as Exhibiting a Classical Dissociation as a Function of the Degree of Skew in Tasks X and Y, the N of the Control Sample, and the Correlation Between Tasks

	Moderate skew (γ_1 = −0.3) Task X only					Severe skew (γ_1 = −0.5) Task X only					Moderate skew (γ_1 = −0.3) Both tasks					Severe skew (γ_1 = −0.5) Both tasks				
N	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8
5	2.59	2.33	2.13	2.02		2.83	2.67	2.57			2.32	1.96	1.69	1.44	1.19		1.97	1.76	1.43	1.24
10	2.82	2.42	2.23	2.05		3.10	2.79	2.71			2.33	1.86	1.53	1.18	0.96		1.84	1.52	1.20	0.97
20	2.89	2.45	2.21	2.00		3.18	2.82	2.66			2.28	1.78	1.45	1.07	0.89		1.75	1.42	1.07	0.86
50	2.93	2.44	2.17	1.93		3.20	2.78	2.58			2.27	1.72	1.41	1.02	0.82		1.67	1.34	1.01	0.80
100	2.91	2.43	2.16	1.87		3.18	2.78	2.56			2.23	1.73	1.39	1.01	0.81		1.64	1.35	0.98	0.79

Note. Because the degree of skew introduced imposes limits on the correlation between X and Y, some columns are blank.

Table 3

Monte Carlo Evaluation of the Robustness of Conventional Criteria (Using z) for Classical Dissociations: Percentages of Control Cases Incorrectly Classified as Exhibiting a Classical Dissociation as a Function of the Degree of Skew in Tasks X and Y, the N of the Control Sample, and the Correlation Between Tasks

	Moderate skew ($\gamma_1 = -0.3$) Task X only					Severe skew ($\gamma_1 = -0.5$) Task X only					Moderate skew ($\gamma_1 = -0.3$) Both tasks					Severe skew ($\gamma_1 = -0.5$) Both tasks				
N	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8
5	19.24	17.38	15.46	12.83		19.72	17.89	16.04			18.18	16.19	14.16	11.52	9.56		15.89	13.98	11.31	9.46
10	14.70	13.35	11.95	10.04		15.20	13.88	12.54			13.88	12.23	10.69	8.57	7.15		12.17	10.62	8.45	7.08
20	12.49	11.41	10.24	8.71		12.97	11.93	10.86			11.86	10.42	9.06	7.19	6.03		10.42	9.04	7.13	5.87
50	11.15	10.24	9.22	7.84		11.58	10.67	9.78			10.61	9.24	8.07	6.37	5.26		9.31	8.03	6.38	5.18
100	10.68	9.22	8.88	7.57		11.13	10.27	9.44			10.12	8.88	7.75	6.12	5.09		8.94	7.72	6.11	5.04

Note. Because the degree of skew introduced imposes limits on the correlation between *X* and *Y*, some columns are blank.

misclassifying a member of the control population as exhibiting a dissociation will be higher. Furthermore, leptokurtic (heavy-tailed) control data will also raise the Type I error rate when Crawford and Garthwaite's (2005) criteria for dissociations are applied. A similar effect was not observed for the conventional criteria, but the error rates were unacceptably high for these criteria regardless of whether the data were normal or otherwise.

It would be prudent for researchers either to select their measures so as to avoid obtaining grossly nonnormal control data or to transform their data prior to applying inferential statistics. For example, the examination of dissociations between the naming of living and nonliving things is often conducted using Snodgrass and Vanderwart's (1980) line drawings. This stimulus set typically yields scores for both living and nonliving things that are very close to ceiling in controls (and hence produces severely skewed data), but with a little effort, it is possible to construct stimulus subsets that yield normally distributed control data (Laws, Gale, Leeson, & Crawford, 2005).

Study 3: Monte Carlo Simulation of Type I Errors When Such Errors Are Defined as Falsely Identifying a Patient as Exhibiting a Classical Dissociation

In Studies 1 and 2, Type I errors were defined as falsely identifying a control case as exhibiting a dissociation. However, an alternative conceptualization of Type I errors in single-case studies is possible. In Study 3 we defined and examined another form of Type I error, namely, incorrectly identifying a patient as exhibiting a dissociation. That is, a patient may have a strictly equivalent level of acquired impairment on both tasks of interest (Tasks X and Y) but be misclassified as exhibiting a dissociation.

To model this scenario, we proceeded as in Study 1; that is, we drew $N + 1$ observations from bivariate normal distributions and took the N observations as the control sample and the $N + 1$th observation as the case. However, on each Monte Carlo trial, we then "lesioned" the case by imposing an acquired impairment of two standard deviations on both Tasks X and Y. As the observations were sampled from a standard normal bivariate distribution, the standard deviation was 1.0 for both Tasks X and Y; therefore, this required simply that 2.0 be subtracted from each case's scores on Tasks X and Y. These cases were then used to represent patients who had suffered equivalent deficits on Tasks X and Y; that is, they did not exhibit a dissociation.

Note that although this procedure is designed to model patients with equivalent acquired deficits, it does not produce cases with equivalent scores on Tasks X and Y. Rather, the method recognizes that (a) patients are initially members of the healthy control population until the onset of their lesion, (b) there will be premorbid differences in competencies on Tasks X and Y, and (c) the magnitude of premorbid differences between Tasks X and Y will be a function of the population correlation between the two tasks (i.e., the magnitude of such differences will, on average, be smaller when the population correlation is high than when it is low).

Table 4

Monte Carlo Evaluation of the Effects of Moderately Leptokurtic Control Data on Criteria for Classical Dissociations: Percentages of Controls Incorrectly Classified as Exhibiting a Dissociation as a Function of the Criteria Applied, the N of the Control Sample, and the Correlation Between Tasks

	Crawford and Garthwaite (2005) criteria					Conventional criteria using *t*					Conventional criteria using *z*				
N	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8
5	2.95	2.42	2.07	1.65	1.39	9.97	9.29	8.28	6.90	5.89	18.18	16.13	14.19	11.51	9.54
10	3.17	2.55	2.14	1.64	1.36	9.63	8.86	7.70	6.28	5.28	13.42	11.90	10.45	8.41	7.02
20	3.12	2.52	2.10	1.58	1.26	9.21	8.43	7.28	5.85	4.86	11.02	9.78	8.61	6.87	5.70
50	3.03	2.37	1.96	1.50	1.20	8.82	8.07	6.86	5.52	4.60	9.49	8.38	7.37	5.89	4.91
100	2.96	2.32	1.91	1.45	1.15	8.61	7.90	6.67	5.42	4.47	8.94	7.98	6.91	5.61	4.63

Table 5
Monte Carlo Evaluation of the Effects of Severely Leptokurtic Control Data on Criteria for Classical Dissociations: Percentages of Controls Incorrectly Classified as Exhibiting a Dissociation as a Function of the Criteria Applied, the N of the Control Sample, and the Correlation Between Tasks

	Crawford and Garthwaite (2005) criteria					Conventional criteria using t					Conventional criteria using z				
N	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8
5	3.44	2.87	2.47	1.96	1.64	10.17	9.27	8.32	6.93	5.90	17.73	15.72	13.81	11.18	9.36
10	3.68	3.02	2.51	1.97	1.63	9.54	8.52	7.53	6.14	5.16	12.89	11.34	9.96	8.07	6.70
20	3.52	2.87	2.42	1.84	1.51	8.82	7.77	6.86	5.51	4.63	10.31	9.05	7.98	6.39	5.31
50	3.26	2.61	2.17	1.69	1.36	8.06	7.11	6.21	5.00	4.15	8.62	7.58	6.61	5.31	4.40
100	3.12	2.49	2.08	1.57	1.27	7.71	6.78	5.95	4.72	3.89	7.97	7.01	6.15	4.87	4.02

Method

The simulations were identical in all but one respect to those used in Study 1 (i.e., 1 million trials were run for each combination of five values of N and five values of ρ_{XY}, giving a total of 25 million trials in all). The crucial difference was that, on each Monte Carlo trial, after drawing the $N + 1$th observation to represent an individual case, 2.0 was subtracted from the case's scores on Tasks X and Y. Thereafter, the simulation proceeded as in Study 1; that is, the calculations were performed to determine if each of the cases met each of the three sets of criteria for a classical dissociation.

Results and Discussion

Following the format adopted in Study 1, the basic pattern of results for the three sets of criteria are presented in Figure 3 (again, these represent the results for $\rho_{XY} = .5$ only). As in Study 1, the pattern in Figure 3 would be more exaggerated for $\rho_{XY} < .5$ and less extreme for $\rho_{XY} > .5$. The full results from the simulation are presented in Table 6. It can be seen from Table 6 that Type I errors (i.e., wrongly classifying a patient with equivalent deficits on Tasks X and Y as exhibiting dissociations) are generally low when Crawford and Garthwaite's (2005) criteria for a classical dissociation are applied. The error rate ranges from a low of 2.74% (for a control sample N of 100 and a correlation between Tasks X and Y of .8) to a high of 7.11% (for a control sample N of 5 and a ρ_{XY} of 0).

In contrast, the Type I error rates were much higher when the conventional criteria for a classical dissociation were applied. When z was used to test for a classical dissociation, the Type I error rate ranged from a low of 19.33% (for an N of 100 and a ρ_{XY} of .8) to a high of 46.13% (for an N of 50 and a ρ_{XY} of 0). In contrast to Study 1, the Type I error rates were even higher (although not dramatically so) when Crawford and Howell's (1998) test was used in place of z. The error rate ranged from a low of 19.49% (for an N of 100 and a ρ_{XY} of .8) to a high of 49.57% (for an N of 10 and a ρ_{XY} of 0).

This difference between the two sets of conventional criteria is worthy of discussion. It is known on theoretical grounds, and on the basis of Monte Carlo simulations (Crawford & Garthwaite, 2005), that Crawford and Howell's (1998) test provides a more valid means of testing for a deficit than does z (because Crawford and Howell's test, unlike the test based on z, treats the statistics of the control sample as statistics rather than as population parameters and thereby controls the Type I error rate). Yet, in the present study, the more valid means of testing for a deficit produced higher misclassification rates when used to operationalize the conventional criteria for a dissociation. This may appear paradoxical, as may the fact that the opposite pattern was found in Study 1 (i.e., the conventional criteria based on z misclassified more healthy control cases than did the use of Crawford and Howell's test). However, there is a relatively simple explanation for these patterns of results.

In Study 1, for all control cases misclassified as exhibiting a classical dissociation when z was used, but not misclassified using Crawford and Howell's (1998) test, both tests were in agreement that there was not a deficit on one of the tasks. The disagreement arose because z recorded a deficit on the remaining task, and this reflects the fact that z is inappropriately liberal. In contrast, in Study 3, the misclassification rate was higher when Crawford and Howell's test was used. This occurred because for all cases mis-

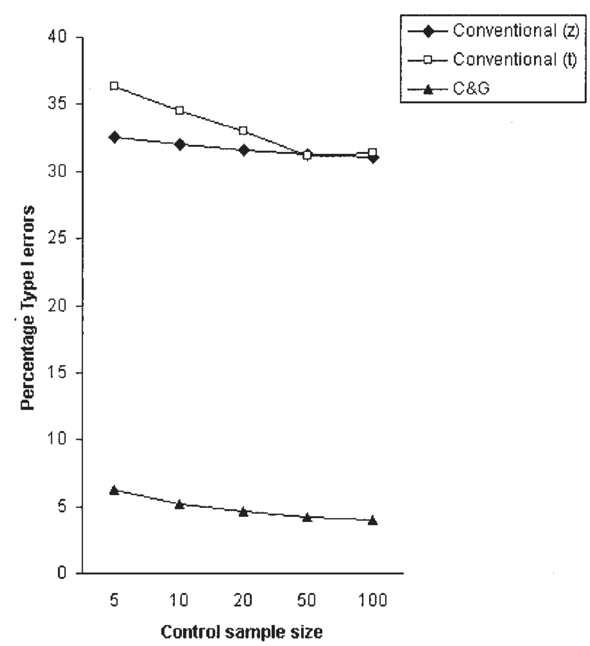

Figure 3. Results from a Monte Carlo simulation: Percentage of Type I errors for three sets of criteria for classical dissociations when Type I errors are defined as misclassifying a patient with equivalent deficits on Tasks X and Y as exhibiting a dissociation (for these data, $\rho_{XY} = .5$). C&G = Crawford and Garthwaite (2005).

Table 6
Results From a Monte Carlo Simulation Study: Percentages of Patients with Equivalent Deficits on Tasks X and Y Misclassified as Exhibiting a Classical Dissociation as a Function of the Criteria Applied, the N of the Control Sample, and the Correlation Between Tasks

	Crawford and Garthwaite (2005) criteria					Conventional criteria using t					Conventional criteria using z				
N	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8	.0	.3	.5	.7	.8
5	7.11	6.87	6.30	5.28	4.46	49.46	42.39	36.26	28.42	23.41	45.76	38.47	32.59	25.35	20.71
10	6.12	5.74	5.22	4.30	3.64	49.57	40.99	34.52	26.59	21.70	46.11	38.09	32.02	24.63	20.08
20	5.49	5.13	4.61	3.78	3.15	48.22	39.60	32.96	25.25	20.46	46.09	37.96	31.59	24.24	19.65
50	5.08	4.65	4.19	3.39	2.83	47.10	38.30	31.88	24.33	19.66	46.13	37.61	31.28	23.93	19.36
100	4.91	4.51	4.00	3.28	2.74	46.53	37.94	31.45	24.02	19.49	46.07	37.59	31.14	23.80	19.33

classified as exhibiting a dissociation using Crawford and Howell's test but not misclassified using z, both tests were in agreement that there was a deficit on one of the tasks. The disagreement arose because z is too liberal and therefore also recorded a deficit on the remaining task (thus, the criteria for a classical dissociation were not met and so the control was not misclassified), whereas the more conservative test of Crawford and Howell did not record a deficit on the remaining task: As a result, the criteria for a classical dissociation were met and the case was incorrectly classified.

The pattern in Study 3, then, is a reflection of the problems with the conventional criteria for classical dissociations rather than stemming from a problem with Crawford and Howell's (1998) test; that is, a test that controls the Type I error rate when applied individually to Task X or Task Y can produce a higher misclassification rate than a test with poor control when used to operationalize the conventional criteria for a classical dissociation.

In conclusion, the results of Study 3 suggest that if Crawford and Garthwaite's (2005) criteria are applied in single-case research, a relatively low proportion of patients who have suffered equivalent impairment on two tasks would be misclassified as exhibiting a classical dissociation. In contrast, if the conventional criteria are applied (regardless of whether these are operationalized using z or Crawford & Howell's, 1998, test), a high proportion of such patients would be misclassified. Indeed, for some of the scenarios examined, it was estimated that close to half of all patients with equivalent deficits would fulfill the conventional criteria for a classical dissociation. These results, therefore, provide further evidence of the need to include a test on the difference between Tasks X and Y when attempting to identify the presence of classical dissociations.

Finally, to our knowledge, the present study is the first to attempt to model the rates at which impaired individuals will be incorrectly classified as exhibiting a classical dissociation. This is surprising perhaps given that this problem will constitute a common threat to validity. That is, in practice, single-case researchers are more likely to encounter genuinely impaired individuals than individuals from the control population, and, moreover, the misclassification rate is markedly greater for such individuals than for individuals drawn from the control population. Given the importance of this issue, we would encourage others to study this question so that a more detailed knowledge base can be built.

Study 4: A Monte Carlo Simulation Study of the Power to Detect a Classical Dissociation

The issue of statistical power in single-case studies has been largely neglected. In Study 4, we examined the power of Crawford and Garthwaite's (2005) criteria to detect a classical dissociation. The power of a statistical test or method refers to its ability to avoid incorrectly accepting the null hypothesis (e.g., to avoid making what are termed Type II errors). In the present context, low power would therefore result in a failure to detect many patients who had suffered genuine classical dissociations.

As sample size is an important determinant of statistical power, power will almost inevitably be low in single-case studies (Crawford, 2004; Crawford, Garthwaite, & Gray, 2003). An individual patient (rather than a sample of patients) is compared with a control sample, and, furthermore, this sample will commonly have a modest N. Dissociations are uncovered only because the effect sizes in this area of inquiry can be very large; neurological damage can have catastrophic effects on the functioning of some cognitive processes while leaving others spared.

We did not examine power for the conventional criteria for a classical dissociation, as these criteria had extremely poor control of Type I errors in Studies 1 and 2 and particularly in Study 3 (in which Type I errors were close to 50% in some cases). It is not possible to address meaningfully the statistical power of a method unless the Type I error rate is under reasonable control (Field, 2001). As an extreme example, if we applied a rule that all individuals (regardless of their test scores) should be classified as exhibiting a classical dissociation, then power would be 100%, but we would of course have achieved nothing.

Method

To examine the power to detect a classical dissociation, the same simulation procedures used in Study 3 were used, but on each trial, instead of subtracting 2.0 from the scores of the case on Tasks X and Y, this was done for Task X only. Thus, the case is used to represent a patient whose lesion had produced a 2 standard deviation impairment on Task X but had spared Task Y; that is, the case represents a patient with a classical dissociation between Tasks X and Y.

Results and Discussion

The simulation results for the power study are presented in Table 7. It can be seen that power is generally low when Crawford

Table 7
Monte Carlo Evaluation of Power for Crawford and Garthwaite's (2005) Criteria: Percentages of Patients with Nonequivalent Deficits on Tasks X and Y Correctly Classified as Exhibiting a Classical Dissociation as a Function of the N of the Control Sample and the Correlation Between Tasks

	Crawford and Garthwaite (2005) criteria				
N	.0	.3	.5	.7	.8
5	13.96	15.37	17.03	20.48	23.86
10	20.17	23.67	27.55	34.50	40.31
20	23.70	28.58	33.77	42.24	48.35
50	25.87	31.52	37.51	46.86	52.67
100	26.70	32.59	38.78	48.29	54.11

and Garthwaite's (2005) criteria for a classical dissociation are applied; the percentage of cases correctly identified as exhibiting a classical dissociation ranges from a low of 14% for a control sample *N* of 5 and a correlation between Tasks X and Y of .0 to a high of 54% for a control sample *N* of 100 and a correlation of .8. As expected, the size of the control sample exerts a strong effect on power. Power more than doubles in moving from an *N* of 5 to an *N* of 100 for almost all values of ρ_{XY} (an exception occurs for $\rho_{XY} = 0$, but it is rare in practice to encounter tasks that are uncorrelated; see below for further discussion).

As noted, power will inevitably be low in single-case studies, as an individual patient (rather than a sample of patients) is compared with a control sample, and furthermore, this sample will commonly have a modest *N*. An additional factor that serves to reduce power in single-case studies is the wide variability in cognitive abilities in the general population. A neurological patient's performance on a given cognitive task will reflect not only the effects of any insult but will also be strongly influenced by the patient's premorbid competencies (Crawford, 1992, 2004; Deary, 1995; Lezak, 1995).

For example, suppose that two cognitive tests have a population correlation of .5 (many pairs of tests used to assess neuropsychological functioning have correlations around this value). With this correlation size, it would not be unusual for members of the healthy population to have a difference between their *z* scores on these tasks of 1.0 or greater (assuming a bivariate normal distribution, approximately 32% of the population would be expected to exhibit such differences). Further suppose that, prior to his or her illness or trauma, a patient exhibited a difference of this magnitude in favor of Task X over Task Y. If the patient's lesion affected some of the cognitive processes underlying performance on Task X but entirely spared any of those underlying Task Y, then it would be difficult to detect the effect unless it was extremely large. An acquired impairment that reduced performance on Task X by 1 standard deviation would only render scores on Tasks X and Y equivalent. An impairment of 2 standard deviations (a substantial decline from the premorbid score) on Task X would result in a discrepancy that was the mirror opposite of the patient's premorbid pattern. However, we have already noted that differences of this magnitude will be common in healthy, intact persons.

The method used to study statistical power in the present study captured these difficulties. The cases were first drawn randomly from the control population before they were lesioned to impose a deficit on Task X. Therefore, 50% of the cases should have had a premorbid score on Task X that exceeded their premorbid Task Y score; this "premorbid" difference would have to be overcome in order for a case to be identified as exhibiting a classical dissociation.

It can be seen from Table 7 that the power to detect classical dissociations is much higher when the population correlation between Tasks X and Y is high; generally speaking, power doubled as ρ_{XY} moved from 0 to .8. This is encouraging as, in practice, the tasks of interest in single-case studies will tend to be moderately to highly correlated. As Shallice (1979) pointed out, this occurs because much of the search for dissociations is focused on tasks that are highly correlated in the general population (i.e., tasks for which there is a prima facie case that they tap a unitary function and therefore may not be dissociable). The previously discussed example of naming living and nonliving things provides a good example; provided that the difficulty level of items is appropriate (i.e., floor effects are avoided), then it is to be expected that the naming of living and nonliving things will be highly correlated in the general population (i.e., they will both reflect the size of individuals' general vocabulary).

In conclusion, to our knowledge, the present study is the first attempt to actually quantify power for criteria for classical dissociations. If Type I errors are to be controlled, low power would appear to be an inherent, unavoidable feature of single-case studies in which a patient is compared with a control sample. It is more encouraging that the focus of interest in many single-case studies will be on tasks that are moderately to highly correlated in the control population; statistical power is higher in these circumstances.

General Discussion

There has been a substantial resurgence of interest in single-case studies in neuropsychology. Among the reasons for this is the belief that the averaged performance of a group of patients can be a meaningless statistical artifact and obscure theoretically important differences among patients. The most extreme version of this argument has been made by Caramazza and colleagues (e.g., Caramazza, 1986; Caramazza & McCloskey, 1988), who argued that neuropsychologists should study only single cases. Vallar (2000) provided a pithy summary of this position: "Hence studies in groups of patients which aim at elucidating the neurological and functional architecture of mental processes are useless and harmful, since they provide misleading results. The only appropriate method is to study individual patients" (p. 334). This is a minority (albeit influential) view. However, many other neuropsychologists have stressed the importance of the study of single cases (Capitani & Laiacona, 2000; Coltheart, 2001; Ellis & Young, 1996; Shallice, 1988). For example, Vallar's (2000) own position is more moderate: "Single-case studies have a number of advantages, in comparison with group studies. The probability to produce significant theoretical advances is perhaps higher" (p. 332).

In parallel with the rise of the single-case approach, dissociations have assumed a greater importance in the building and testing of theory in neuropsychology. For example, Dunn and Kirsner (2003) noted that "dissociations play an increasingly crucial role in the methodology of cognitive neuropsychology . . . they have provided critical support for several influential, almost paradigmatic, models in the field" (p. 2). This is partly because of the limitations of alternative strategies (see next section on associations). For example, Vallar (2000) noted that dissociations constitute "the most effective paradigm for investigating the modularity of the mental processes and their neural correlates" (p. 329).

In view of the foregoing, it is to be regretted that there has been little attempt to quantify fundamental characteristics of criteria for dissociations in single-case studies. This neglect may stem from two inherent difficulties. First, there is obviously no infallible standard for the presence of a dissociation against which any criteria can be compared. Second, the rationale of the single-case approach is such that it is difficult, or may even be impossible, to study the effectiveness of criteria for dissociations through conventional replication studies. For example, if a patient is classified as exhibiting a dissociation, the failure to find the same dissociation in patients who otherwise have shared cognitive features would not be regarded as invalidating the previously observed dissociation. As Vallar (2000) noted, a single-case researcher may take the view that "given the complex architecture of the cognitive system and the variability of the site and extent of naturally occurring lesions, it is very unlikely that two patients have similar functional deficits" (p. 334). Thus, Coltheart (2001) asked the rhetorical question, "If every patient is unique, how can you replicate your results?" and concluded that in some cases it may be that "the result is literally unreplicable, no matter how genuine" (p. 19).

Given the lack of an infallible standard and the emphasis on the uniqueness of patients, we suggest that simulation studies, such as those carried out in the present study, offer a useful means of gaining purchase on this problem. That is, they hold out the prospect of introducing greater methodological rigor into the process of both setting and evaluating criteria for dissociations.

Conventional Criteria Versus Crawford and Garthwaite's (2005) Criteria

Crawford and Garthwaite's (2005) criteria for classical dissociations in the single case are more complicated than the conventional criteria and require more in the way of computation. However, this is a small price to pay in order to avoid the very high rates of false positives observed for the latter criteria. Moreover, the statistical methods used to test whether Crawford and Garthwaite's criteria are met have been implemented in a computer program for PCs. The data inputs required by this program are the means and standard deviations for Tasks X and Y and the correlation between Tasks X and Y in controls, the *N* of the control sample, and the patient's scores on Tasks X and Y. The program then applies the statistical methods required to test if the criteria are met and reports the outcome (the intermediate results from the statistical tests are also reported).[1] By using this program, a patient's data can be analyzed in under a minute. For further details, see Crawford and Garthwaite (2005).

The Status of Dissociations Versus Associations

Historically, much of the research effort in neurology and neuropsychology has been aimed at demonstrating associations between cognitive tasks. That is, emphasis was placed on detecting clusters of cognitive symptoms that reliably co-occurred in order to identify neurological or neuropsychological syndromes. Dissatisfaction with the logical status of such evidence served as one of the drivers for the upsurge of interest in single-case studies (Vallar, 2000). Regardless of how many patients demonstrate co-occurrence of deficits on two tasks, the hypothesis that a common cognitive process underlies performance on them can be overturned by a single patient who exhibits a dissociation (Ellis & Young, 1996). As Karl Popper noted famously, the statement "All swans are white" can be disproved by the discovery of one black swan.

However, although the limitations of associational evidence are widely recognized, it is also clear that there are practical difficulties in evaluating evidence based on apparent dissociations. For example, suppose that there is longstanding associational evidence that a particular set of cognitive tasks are indicators of an underlying unitary function. If this evidence is to be overturned by one counterexample, then one would want to be sure that the counterexample was genuine. At the risk of stretching the aforementioned analogy, one needs to be sure that one is indeed faced with a black swan rather than a regular swan that has had a chance encounter with an oil spill. The results of the present study indicate clearly that a case classified as exhibiting a classical dissociation according to the conventional criteria does not provide the required reassurance.

It could even be argued that Crawford and Garthwaite's (2005) criteria are insufficiently conservative given that a modest but nevertheless appreciable percentage of controls and patients with equivalent deficits were misclassified as exhibiting classical dissociations. It would be straightforward to modify these latter criteria to render them more conservative. For example, the test on the standardized difference between a case's scores on Tasks X and Y could be modified to require that the difference achieve significance at a more conservative value of alpha such as the 2.5% level rather than the conventional 5% level adopted by Crawford and Garthwaite.

Although this strategy would reduce Type I errors, it will be appreciated that anything gained thereby has to be paid for by an increase in Type II errors; that is, the power to detect genuine dissociations will be further reduced. Given that power in single-case studies is already low, a further reduction risks missing all but the most extreme examples of dissociations. Our view is that Crawford and Garthwaite's (2005) criteria, as currently specified, strike a reasonable balance between the competing demands of controlling these two types of errors. However, we appreciate that others may take a different view and therefore encourage debate on this topic.

Classical Dissociations Versus Strong Dissociations

The focus of the present study has been on classical dissociations. However, another important form of dissociation is what Shallice (1988) termed the "strong" dissociation and defined as follows: "In a *strong dissociation*, neither task is performed at normal level, but Task I is performed very much better than Task II" (p. 228). Coltheart (2001) provided a similar definition:

[1] This program is freely available over the Internet and can be downloaded from http://www.abdn.ac.uk/~psy086/dept/dissociations.htm

> One can still speak of dissociations between two tasks even if performance is impaired on both tasks. If a patient is impaired at both Task A and Task B, but is significantly more impaired on the second task than on the first, that can be treated as a dissociation. (p. 12)

The decision to limit the present investigation to classical dissociations was based on a number of factors. The first was simply that the study was already complicated enough without the addition of another factor. Second, classical dissociations are generally considered to constitute a stronger source of evidence for modularity, and hence more weight is attached to them when building or testing cognitive models in neuropsychology. The third and most important reason was that we consider that the existing conventional criteria for strong dissociations are far less problematic. That is, as is implied in Shallice's (1988) definition, and is explicit in the definitions of Coltheart (2001) and others (e.g., Ellis & Young, 1996), the conventional criteria for a strong dissociation include a test on the difference between a patient's performance on Tasks X and Y. The simulation results demonstrate that it is the inclusion of a test on this difference that is the crucial component in reducing false positives.

Crawford and Garthwaite's (2005) criteria for a strong dissociation are, in essence, simply fully operational definitions of the conventional criteria (i.e., they provide explicit rules for concluding that impairment is present on both tasks and that there is a difference between tasks). The crucial difference between these two sets of criteria lies in the fact that in the former, the requirement for significant difference between tasks applies equally to strong and classical dissociations. In contrast, for both Coltheart (2001) and Ellis and Young (1996, p. 5), this requirement need only be invoked when the patient is classified as having a deficit on both tasks (i.e., when the criteria for a classical dissociation have not been met and the issue is whether the patient nevertheless exhibits a strong dissociation).

Although strong dissociations have not been examined here, we nevertheless consider that they are worthy of future study using a similar approach. For existing simulation data on strong dissociations, see Crawford and Garthwaite (2005).

Implications of the Present Results for the Status of Double Dissociations

In attempts to uncover the underlying functional architecture of human cognition, great weight is given to double dissociations. To establish a double dissociation requires two patients who have the opposite patterns of spared and impaired functions. As Coltheart (2001) stated, "With double dissociations we need two patients: patient A who is impaired on Task X but normal on Task Y, and patient B who is normal on Task X but is impaired on Task Y" (p. 12).

A single dissociation is not regarded as providing definitive evidence of fractionation of the cognitive system because the two tasks involved may tap a single, common underlying process but simply differ in the extent to which they place demands on this process; that is, single dissociations are prone to task difficulty artifacts (Shallice, 1988; Vallar, 2000; see also the classic article by Chapman & Chapman, 1973).[2] The existence of a double dissociation is widely considered to largely (but not entirely) rule out task difficulty as a competing explanation; although see Dunn and Kirsner (2003) for a closely argued, more pessimistic, view. However, even if one were to accept that the double dissociation was entirely immune to task difficulty artifacts, the present simulation results serve to illustrate that it would not render a single-case study immune to another source of artifact: simple chance variation.

In Studies 1 and 2, half of the control cases misclassified as exhibiting a dissociation will have exhibited a dissociation in favor of Task X, with the opposite occurring in the remaining cases (the figure will not be exactly 50% in each category because of Monte Carlo variation but will be close to it given the number of simulations performed). As a specific example: If the conventional criteria for a dissociation were used for tasks having a population correlation of .5, the control sample used to quantify an individual's performance had an N of 10, and z was used to determine the presence or absence of a deficit, then the simulation results would estimate that 11% of cases would be classified as exhibiting a dissociation (see Table 1). Therefore, under these circumstances, the expectation is that approximately 5.5% of healthy intact persons would be incorrectly classified as exhibiting a dissociation in favor of Task X and an equal number would be incorrectly classified as exhibiting a dissociation in favor of Task Y.

Under the same circumstances but with a patient with equivalent (2 standard deviations) deficits on Tasks X and Y substituted for the control case (i.e., the scenario examined in Study 3), it can be seen from Table 6 that approximately 32% of such patients would be misclassified. Therefore, the expectation is that 16% of such patients would be classified as exhibiting a dissociation in favor of Task X; this leaves a further 16% of cases to provide the double dissociation for any of these foregoing patients. In this latter scenario, there is an embarrassment of apparent double dissociations. Thus, although the double dissociation may largely deal with one source of artifact (differing task difficulties), it only halves the likelihood that another source (chance variation) accounts for the results observed.

The Size of the Control Sample and Its Implications

Very useful and elegant methods have been devised for drawing inferences concerning an individual patient's performance on fully standardized neuropsychological tests, that is, on tests that have been normed on very large, representative samples of the population (e.g., Capitani, 1997; De Renzi, Faglioni, Grossi, & Nicheli, 1997; Willmes, 1985). When these methods are used in single-case research, the patient is compared against normative values rather than against controls. With such approaches, error arising from sampling from the control population is ignored; this is justifiable because the samples are large enough for such error to be minimal. Some of these large-sample methods (see Capitani & Laiacona, 2000) have the other advantage that they are nonparametric; thus, concerns about departures from normality do not arise.

Although these latter approaches have much to commend them, unfortunately they can be used only in fairly circumscribed situations because (a) the questions posed in many single-case studies cannot be fully addressed using existing standardized neuropsy-

[2] An anonymous reviewer suggested that simulations that vary the difficulty (i.e., discriminating power) of the tasks could be conducted. Although this is beyond the scope of the present study, we also think such research would be very informative and encourage research on this topic.

chological tests, (b) new constructs are constantly emerging in neuropsychology, and (c) the collection of large-scale normative data is a time-consuming and arduous process (Crawford, 2004). Furthermore, even when the requisite large-scale normative data are available for a given task, these norms will be applicable only to patients who are sufficiently similar to the normative sample in terms of basic characteristics, that is, language, country of residence, and so forth (Capitani & Laiacona, 2000).

Therefore, there is a continued need for methods that can be used when a patient is compared with a matched and modestly sized control sample. The methods developed by Crawford, Garthwaite, and colleagues (Crawford & Garthwaite, 2002, 2005; Crawford & Howell, 1998) were motivated by the need to deal with this scenario. However, to avoid any potential confusion, it should be noted that the methods can be used with control samples of any size. Indeed, they remain more valid than the commonly used alternatives based on z when N is large; in this situation, the researcher is still dealing with a control sample and not a control population.

It is also important to note that, although these methods achieve good control of Type I errors even at small Ns, this should not be taken to imply that researchers limit themselves to recruiting small control samples. Study 4 demonstrated that if the Type I error rate is to be held to an acceptably low level, then the power to detect a classical dissociation is also low. Given that sample size is the primary determinant of power, this indicates that the control samples in single-case studies should be larger (i.e., $\geq$ 25) than is currently typical.

It is neither unreasonable nor impractical to suggest that more time and effort should be devoted to obtaining a decently sized control sample in single-case studies. If a researcher believes that the single-case approach can advance knowledge more, or at least as much, as group studies, then he or she should be willing to expend an amount of time and resources equivalent to those typically expended in group studies.

The Criteria for Dissociations and the Conduct of Single-Case Studies

The focus of the present study has been on comparing a patient's performance on a pair of tasks to that of controls. However, it should be acknowledged that findings obtained from such comparisons are not interpreted in isolation. Rather, these findings are normally interpreted in the context of results from a prior assessment in which a broad characterization of the patient's strengths and weaknesses will have been achieved through the use of fully or partially standardized tests (see previous section).

Furthermore, many single-case studies use multiple measures of the constructs under investigation (i.e., different but related Tasks X_1, X_2, etc. and Tasks Y_1, Y_2, etc. to measure constructs X and Y). That is, the patient is compared with controls over a series of tasks. This is in keeping with the fact that researchers are ultimately interested in dissociations between functions, not just in dissociations between specific pairs of indirect and imperfect measures of these functions (Crawford, Garthwaite, Howell, & Venneri, 2003; Vallar, 2000). Thus, researchers seek converging evidence of a deficit or dissociation (Shallice, 1979; Vallar, 2000). The upshot of this is that the overall risk of drawing incorrect conclusions in a single-case study will normally be less than that associated with the application of a set of criteria to a single pair of tests.

However, the integration of these multiple sources of information is a complex and formidable task and is beyond the scope of the present work. As Crawford and Garthwaite (2005) observed, there is currently little consistency across studies in how this task is approached, and existing attempts tend to be qualitative rather than quantitative. The development and evaluation of a quantitative system, whereby the probabilities of a dissociation could be combined or updated as different stages of a study are completed, would make a significant contribution to the discipline. The nature of this problem is such that an approach based on Bayesian rather than classical (i.e., frequentist) statistical methods would be the obvious choice.

Relevance of the Present Work to Practice in Clinical Neuropsychology

With the rise of the single-case approach, many academic neuropsychologists now face the same basic problem that has long been faced by clinical neuropsychologists: how to draw valid inferences concerning the pattern of cognitive performance of an individual patient. Although the emphasis of the present article has been on single-case research, the results also therefore have implications for practice in clinical neuropsychology. Clinical neuropsychologists have commonly laid emphasis on the need to supplement normative comparison standards with individual comparison standards (Crawford, 2004; Lezak, Howieson, Loring, Hannay, & Fischer, 2004). That is, when attempting to detect acquired impairment, it is recommended that analysis should not be limited to comparing a patient with an appropriate normative sample but should also incorporate examination of the difference between the patient's performance on the tasks in question.

The present results underline the importance of these individual comparison standards. It can be seen from the simulation results that inferring the presence of a relative weakness on one task over another by examining only whether the tasks in question are significantly below normal performance can be misleading if it is not supplemented by a test on the difference between the two tasks (as noted, one of the tasks may fall just below the criterion for impairment while the other falls just above; in this situation, the difference is trivial and would not be significant when tested).

As previously noted, a feature of the methods recommended here is that they are suitable for use when the control or normative sample is modest in size. In comparison to single-case researchers, this is not as important a consideration for clinical neuropsychologists, as many of the measures they use have been standardized on large normative samples. However, it remains the case that clinical neuropsychologists often need to fall back on normative data that are based on relatively small sample sizes for the reasons outlined in an earlier section (e.g., the lag between the emergence of an important new construct and the provision of large-scale normative data for tests that measure the construct). Moreover, even when the overall N for a normative sample is large, the actual N against which clinical neuropsychologists compare their patients may be much more modest if the normative data have been stratified by age, gender, and so forth. In view of these practical considerations, in clinical practice it will often be prudent to compare a patient's performance with normative data using the methods employed here, that is, Crawford and Howell's (1998) test for testing for a deficit and the RSDT for testing differences rather than methods that use *z*.

This consideration is particularly relevant when the concern is with examining differences. To draw valid conclusions concerning the abnormality of differences between two tasks, it is necessary to know their intercorrelation. As clinical neuropsychologists use tests from diverse sources, it will be relatively common that the tests being compared have not been costandardized. Therefore, if neuropsychologists want to conduct a quantitative comparison of these two tasks, they must fall back on studies in which the tasks have been coadministered (in order to obtain the required intercorrelation). In most cases, the *N* in such studies will be much more modest than the *N* for the individual normative samples. Therefore, methods (such as the RSDT) that treat the sample statistics as sample statistics are to be preferred over alternatives such as the Payne and Jones (1957) formula, which are appropriate only when the sample is large enough to warrant treatment of the sample statistics as population parameters.

Conclusion

The single-case approach in neuropsychology has contributed greatly to the understanding of the functional architecture of human cognition (Ellis & Young, 1996; McCarthy & Warrington, 1990; Shallice, 1988; Vallar, 2000). However, although logical considerations as well as other considerations have led many researchers to abandon group studies in favor of single-case studies, it is undeniable that the latter approach poses significantly more statistical problems. Newcombe and Marshall (1988), in commenting on the decline in single-case studies in the 1920s, observed that "single-case studies, no matter how well-conducted . . . began to be described as merely anecdotal" (p. 549). If a similar fate is not to befall the current resurgence in the use of single-case studies, there is a need to develop standards of statistical practice that approach the rigor of those demanded in group studies. As Caramazza and McCloskey (1988) noted in their commentary on single-case methods, "If advances in theory are to be sustainable they . . . must be based on unimpeachable methodological foundations" (p. 519).

Although there remains much to do, the present study has provided empirical evidence to guide single-case researchers in their decisions about methodology. Some of this evidence is extremely alarming, for example, the indications that very high percentages (close to 50% in some of the scenarios examined) of controls and patients will be incorrectly classified as exhibiting classical dissociations when the conventional criteria are applied. Methods that produce misclassification rates of this magnitude do not provide a sound basis for theory building in neuropsychology. Fortunately, however, criteria that include a test on the difference between tasks reduce the misclassification rates to acceptable levels.

Other evidence is more generally reassuring. For example, nonnormality, which is a common feature of single-case control data, does not appear to produce a serious increase in false positives. The fact that these studies turned up findings that could not easily have been predicted suggests that further use of simulation studies in this topic area should be encouraged. Finally, the issue of the power to detect dissociations has been largely neglected in the single-case literature; it is hoped that the empirical results and discussion of this issue in the present article have begun to redress this. It is argued that if the rate of false positives is to be controlled, then it is inevitable that power will be relatively low in this area of inquiry. However, the power to detect large effects (which are not uncommon in this area) is higher than might be expected given the inherent difficulties.

References

Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society Series B, 61,* 579–602.

Azzalini, A., & Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika, 83,* 715–726.

Box, G. E. P., & Muller, M. E. (1958). A note on the generation of random normal deviates. *Annals of Mathematical Statistics, 28,* 610–611.

Calder, A. J., Young, A. W., Rowland, D., Perret, D. I., & Hodges, J. R. (1996). Facial emotion recognition after bilateral amygdala damage: Differentially severe impairment of fear. *Cognitive Neuropsychology, 13,* 699–745.

Capitani, E. (1997). Normative data and neuropsychological assessment. Common problems in clinical practice and research. *Neuropsychological Rehabilitation, 7,* 295–309.

Capitani, E., & Laiacona, M. (2000). Classification and modelling in neuropsychology: From groups to single cases. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (2nd ed., Vol. 1, pp. 53–76). Amsterdam: Elsevier.

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

Chapman, L. J., & Chapman, J. P. (1973). Problems in the measurement of cognitive deficit. *Psychological Bulletin, 79,* 380–385.

Coltheart, M. (2001). Assumptions and methods in cognitive neuropsychology. In B. Rapp (Ed.), *The handbook of cognitive neuropsychology* (pp. 3–21). Philadelphia: Psychology Press.

Crawford, J. R. (1992). Current and premorbid intelligence measures in neuropsychological assessment. In J. R. Crawford, D. M. Parker, W. W. McKinlay (Eds.), *A handbook of neuropsychological assessment* (pp. 21–49). London: Erlbaum.

Crawford, J. R. (2004). Psychometric foundations of neuropsychological assessment. In L. H. Goldstein & J. E. McNeil (Eds.), *Clinical neuropsychology: A practical guide to assessment and management for clinicians* (pp. 121–140). Chichester, England: Wiley.

Crawford, J. R., & Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia, 40,* 1196–1208.

Crawford, J. R., & Garthwaite, P. H. (2005). Testing for suspected impairments and dissociations in single-case studies in neuropsychology: Evaluation of alternatives using Monte Carlo simulations and revised tests for dissociations. *Neuropsychology, 19,* 318–331.

Crawford, J. R., Garthwaite, P. H., & Gray, C. D. (2003). Wanted: Fully operational definitions of dissociations in single-case studies. *Cortex, 39,* 357–370.

Crawford, J. R., Garthwaite, P. H., Howell, D. C., & Venneri, A. (2003). Intra-individual measures of association in neuropsychology: Inferential methods for comparing a single case with a control or normative sample. *Journal of the International Neuropsychological Society, 9,* 989–1000.

Crawford, J. R., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist, 12,* 482–486.

Deary, I. J. (1995). Age-associated memory impairment: A suitable case for treatment. *Ageing and Society, 15,* 393–406.

De Renzi, E., Faglioni, P., Grossi, D., & Nicheli, P. (1997). Apperceptive and associative forms of prosopagnosia. *Cortex, 27,* 213–221.

Dunn, J. C., & Kirsner, K. (2003). What can we infer from double dissociations? *Cortex,* 39, 1–7.

Ellis, A. W., & Young, A. W. (1996). *Human cognitive neuropsychology: A textbook with readings.* Hove, England: Psychology Press.

Field, A. P. (2001). Meta-analysis of correlation coefficients: A Monte Carlo comparison of fixed- and random-effects methods. *Psychological Methods, 6,* 161–180.

Garthwaite, P. H., & Crawford, J. R. (2004). The distribution of the difference between two *t*-variates. *Biometrika, 91,* 987–994.

Kennedy, W. J., & Gentle, J. E. (1980). *Statistical computing.* New York: Marcel Dekker.

Lange, K. L., Little, R. J. A., & Taylor, J. M. G. (1989). Robust statistical modelling using the *t*-distribution. *Journal of the American Statistical Association, 84,* 881–896.

Laws, K. R. (in press). Illusions of normality: A methodological critique of category-specific naming. *Cortex.*

Laws, K. R., Gale, T. M., Leeson, V. C., & Crawford, J. R. (2005). When is category *specific* in Alzheimer's disease? *Cortex, 41,* 452–463.

Lezak, M. D. (1995). *Neuropsychological assessment* (3rd ed.). New York: Oxford University Press.

Lezak, M. D., Howieson, D. B., Loring, D. W., Hannay, H. J., & Fischer, J. S. (2004). *Neuropsychological assessment* (4th ed.). New York: Oxford University Press.

McCarthy, R. A., & Warrington, E. K. (1990). *Cognitive neuropsychology: A clinical introduction.* San Diego, CA: Academic Press.

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

Payne, R. W., & Jones, G. (1957). Statistics for the investigation of individual cases. *Journal of Clinical Psychology, 13,* 115–121.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1989). *Numerical recipes in Pascal.* Cambridge, England: Cambridge University Press.

Scott, S. K., Young, A. W., Calder, A. J., Hellawell, D. J., Aggleton, J. P., & Johnson, M. H. (1997). Impaired auditory recognition of fear and anger following bilateral amygdala lesions. *Nature, 385,* 254–257.

Shallice, T. (1979). Case study approach in neuropsychological research. *Journal of Clinical Neuropsychology, 3,* 183–211.

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

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures. *Journal of Experimental Psychology: Human Learning and Memory, 6,* 174–215.

Vallar, G. (2000). The methodological foundations of human neuropsychology: Studies in brain-damaged patients. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology* (2nd ed., Vol. 1, pp. 305–344). Amsterdam: Elsevier.

Willmes, K. (1985). An approach to analyzing a single subject's scores obtained in a standardized test with application to the Aachen Aphasia Test (AAT). *Journal of Clinical and Experimental Neuropsychology, 7,* 331–352.

Appendix

Sampling From Bivariate Normal and Bivariate Skew Distributions

Bivariate Normal Distributions

To sample from bivariate normal distributions, first generate two independent standard normal variates u_0, and u_1. If the desired correlation between Tasks X and Y is denoted as ρ_{XY}, then

$$Z_X = u_0$$

$$Z_Y = \rho_{XY}u_0 + ([\sqrt{1 - \rho_{XY}{}^2}]u_1) \tag{A1}$$

are observations from the required standard normal bivariate distribution.

Bivariate Skew Distributions

The method used to sample from skew distributions was based on work by Azzalini and colleagues (Azzalini & Capitanio, 1999; Azzalini & Dalla Valle, 1996). The starting point for this method is the generation of three independent standard normal variates u_0, u_1, and u_2; u_1 and u_2 are used to form the Task X and Task Y observations, and u_0 is used to control the degree of skew in Tasks X and Y. Next, a correlation matrix Ω is specified that will yield distributions with the required degree of skew (γ_1) in Tasks X and Y and the required correlation between Tasks X and Y (ρ_{XY}),

$$\Omega^* = \begin{bmatrix} 1 & \rho_{u_0u_1} & \rho_{u_0u_2} \\ \rho_{u_0u_1} & 1 & \rho_{u_1u_2} \\ \rho_{u_0u_2} & \rho_{u_1u_2} & 1 \end{bmatrix}.$$

The required values of $\rho_{u_0u_1}$ and $\rho_{u_0u_2}$ can be obtained by algebraic manipulation of Azzalini and Dalla Valle's (1996) formulas for γ_1 to solve, in turn, for $\rho_{u_0u_1}$ and $\rho_{u_0u_2}$. That is, put

$$a_X = \left(\frac{2\gamma_{1_X}}{4 - \pi}\right)^{\frac{1}{3}} \tag{A2}$$

and

$$a_Y = \left(\frac{2\gamma_{1_Y}}{4 - \pi}\right)^{\frac{1}{3}}, \tag{A3}$$

where γ_{1_X} and γ_{1_Y} are the desired skewness of *X* and *Y,* respectively; then

$$\rho_{u_0u_1} = a_X\left(\frac{\pi}{2 + 2a_X^2}\right)^{\frac{1}{2}} \tag{A4}$$

and

$$\rho_{u_0u_2} = a_Y\left(\frac{\pi}{2 + 2a_Y^2}\right)^{\frac{1}{2}}. \tag{A5}$$

Note also that $\rho_{u_1u_2}$ is not the required correlation between Tasks X and Y (ρ_{XY}). The value that $\rho_{u_1u_2}$ must take so that ρ_{XY} has the desired value can be determined by manipulating Azzalini and Dalla Valle's (1996) Formula 2.14 to solve for $\rho_{u_1u_2}$. That is, put

$$\rho_{u_1u_2} = \{(1 - 2\pi^{-1}\rho^2_{u_0u_1})(1 - 2\pi^{-1}\rho^2_{u_0u_2})\}^{\frac{1}{2}}\rho_{XY} + 2\pi^{-1}\rho_{u_0u_1}\rho_{u_0u_2}. \tag{A6}$$

As an example, if the desired skewness (γ_1) for Tasks X and Y is −.5 and .0, respectively, and the desired correlation between Tasks X and Y is .5, then the required matrix is

$$\Omega^* = \begin{bmatrix} 1.00000 & -0.90848 & 0.00000 \\ -0.90848 & 1.00000 & 0.34445 \\ 0.00000 & 0.34445 & 1.00000 \end{bmatrix}.$$

Finally, the standard normal variates u_1 and u_2 are multiplied by the elements in rows 2 and 3, respectively, of the lower triangular Cholesky decomposition of Ω, and these products are summed to form a vector

$$[z_X, z_Y]^T.$$

This vector is then modified such that

$$[z_X, z_Y]^T = \begin{Bmatrix} [z_X, z_Y]^T, \text{ if } u_0 \geq 0 \\ [-z_X, -z_Y]^T, \text{otherwise} \end{Bmatrix}$$

and is then a random vector from the required bivariate skew distribution.

Received May 24, 2004
Revision received September 15, 2004
Accepted December 8, 2004 ■