ELSEVIER

Neuropsychologia 44 (2006) 2249–2258

NEUROPSYCHOLOGIA

www.elsevier.com/locate/neuropsychologia

Detecting dissociations in single-case studies: Type I errors, statistical power and the classical versus strong distinction

John R. Crawford [a,*], Paul H. Garthwaite [b]

[a] *School of Psychology, College of Life Sciences and Medicine, King's College, University of Aberdeen, Aberdeen AB24 2UB, UK*
[b] *Department of Statistics, The Open University, Milton Keynes, UK*

Received 6 September 2005; received in revised form 10 May 2006; accepted 16 May 2006
Available online 3 July 2006

Abstract

Dissociations observed in single-case studies play an important role in building and testing theory in neuropsychology; therefore the criteria used to identify their presence should be subjected to empirical scrutiny. Extending work on classical dissociations, Monte Carlo simulation is used to examine the Type I error rate for two methods of detecting strong dissociations. When a Type I error was defined as misclassifying a healthy control, error rates were low for both methods. When Type I errors were defined as misclassifying patients with strictly equivalent deficits on two tasks, error rates for strong dissociations were high for the "conventional" criteria and were very high when cases misclassified as exhibiting either form of dissociation (strong *or* classical) were combined (maximum = 55.1%). The power to detect strong and classical dissociations was generally low-to-moderate, but was moderate-to-high in most scenarios when power was defined as the ability to detect either form of dissociation. In most scenarios patients with strong dissociations were more likely to be classified as exhibiting classical dissociations. The results question the practical utility of the distinction between strong and classical dissociations regardless of the criteria employed to test for their presence.

Keywords: Single-case studies; Statistical methods; Strong dissociations; Classical dissociations; Monte Carlo simulation

1. Introduction

Dissociations observed in single-case studies have come to play an important role in the building and testing of theory in neuropsychology (Capitani & Laiacona, 2000; Ellis & Young, 1996; Shallice, 1988). Shallice (1988) identified three forms of dissociation: classical, strong and trend dissociations. Typically a classical dissociation is defined as occurring when, with reference to the performance of matched healthy controls, a patient is "impaired" or "exhibits a deficit" on task *X* but is "not impaired", "normal" or "within normal limits" on task *Y* (Coltheart, 2001; Ellis & Young, 1996; Shallice, 1988). In the case of a strong dissociation, "neither task is performed at normal level, but task I is performed very much better than task II" (Shallice, 1988, p. 228). Note that in both of the foregoing forms of dissociation, inferences concerning a patient are made by comparing her/his performance to that of a healthy control or normative sample. In contrast, in a trend dissociation, inferences are based solely on an *intra*-individual comparison; namely that the patient's score on one task is simply markedly lower than her/his score on another. This latter form of dissociation is regarded as providing the weakest form of evidence for modularity of function (Shallice, 1988) and is not examined in the present study.

In view of the important status of classical and strong dissociations, it is to be regretted that there has been little attempt to quantify fundamental characteristics of the criteria used to identify them in single-case studies. That is, until recently, relatively little was known about the extent to which criteria for dissociations control Type I errors (i.e., avoid false positives whereby an individual is incorrectly classified as exhibiting a dissociation) and avoid Type II errors (i.e., avoid false negatives and therefore successfully detect individuals with genuine dissociations).

In a recent study Crawford and Garthwaite (2005a) employed Monte Carlo simulation to evaluate criteria for classical dissociations. They compared Type I error rates for the conventional

* Corresponding author.
E-mail address: j.crawford@abdn.ac.uk (J.R. Crawford).

0028-3932/$ – see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.neuropsychologia.2006.05.019

criteria (as defined earlier) with criteria developed by Crawford and Garthwaite (2005b). The principal difference between these criteria is that the latter requires that patients exhibit a statistically significant difference between their performance on the two tasks involved. In contrast, the conventional criteria only require that the patient is impaired on one task and within normal limits on the other (see Crawford, Garthwaite, & Gray, 2003 for a detailed discussion of the problems that stem from this).

A further difference between these two sets of criteria is that, in practice, single-case researchers have commonly used z to test whether the conventional criteria have been met. That is, the patient's score is converted to a standard score based on the mean and standard deviation of the controls and the result referred to a table of the areas under the normal curve; if z is less than -1.645 then the patient's score is significantly lower ($p < 0.05$, one-tailed) than controls and it is concluded they have a deficit (if not, they are considered to be within normal limits). In contrast, Crawford and Garthwaite's (2005b) criteria are tested using methods based on t-distributions rather than z: a modified independent samples t-test is used to test for a deficit (Crawford & Howell, 1998) and the Revised Standardized Difference Test (RSDT; Crawford & Garthwaite, 2005b; Garthwaite & Crawford, 2004) is used to test whether the standardized difference between a patient's scores on tasks X and Y differs significantly from the distribution of standardized differences in the controls. (Note that this latter method examines the *standardized* difference between tasks X and Y because the raw score means and standard deviations for the tasks are usually not comparable.)

Crawford and Garthwaite's (2005a) simulation studies indicated that Type I error rates for the conventional criteria were high in most scenarios examined (i.e., many single-cases would be classified incorrectly as exhibiting classical dissociations) whereas their own criteria were conservative in all scenarios.

1.1. Evaluating criteria for a strong dissociation

Crawford and Garthwaite (2005b) have also provided formal criteria for a strong dissociation: patients must exhibit a significant deficit on both tasks using Crawford and Howell's (1998) t-test and a significant difference between their scores on the two tasks using the RSDT.[1] Unlike the two sets of criteria for a classical dissociation, Crawford and Garthwaite's criteria and the conventional criteria for strong dissociations differ mainly in the degree to which they specify how we should test if the criteria are met. That is, as is implied in Shallice's (1988) definition, and is explicit in the definitions of strong dissociations provided by others (e.g., Ellis & Young, 1996), the conventional criteria for a strong dissociation (unlike the conventional criteria for a classical dissociation) also include a test on the difference between a patient's performance on tasks X and Y. In practice however they also differ in that most single-case researchers still employ z-based methods, rather than the t-based methods advocated by Crawford and colleagues. It is therefore appropriate to compare these two approaches to identifying strong dissociations. In particular, Crawford and Garthwaite's (2005a) suggestion that the use of z-based methods may produce a high number of false positives should be tested.

There are at least three other reasons why strong dissociations should receive attention. The first of these is a general one: strong dissociations are regarded by many neuropsychologists as a legitimate means of testing for modularity (Ellis & Young, 1996) and should therefore be subjected to as much scrutiny as classical dissociations. Second, isolated, highly specific deficits are relatively rare following neurological trauma or disease. Thus, in practice, single-case researchers are liable to encounter more patients with strong dissociations than dissociations of the classical type. Fourth, it can be argued that we should be as concerned, or indeed more concerned, with the *combined* Type I error rate for classical and strong dissociations than with the error rates for either form of dissociation alone.

To illustrate this last point, suppose that the false positive rate for a set of criteria for a classical dissociation is 7.5%. Although this Type I error rate is appreciable, some researchers may be willing to accept this level of risk. However, if the corresponding criteria for a strong dissociation produced the same Type I error rate, then most researchers would feel much less comfortable about inferring the presence of a dissociation using these criteria. That is, it could be argued that it is the *combined* error rate of 15% (the probability of incorrectly identifying cases as exhibiting a dissociation of either form) that should be attended to and this rate is too high for the criteria to be considered at all rigorous.

2. Study 1

2.1. A Monte Carlo simulation study of Type I errors when applying criteria for strong and classical dissociations

In this first study we run a Monte Carlo simulation to quantify and compare control of the Type I error rate when the alternative criteria for a strong dissociation are applied. Crawford and Garthwaite's (2005b) criteria are compared to criteria based on the use of z. We also examine the combined Type I error rate (i.e., the percentage of cases incorrectly classified as exhibiting either a strong *or* classical dissociation). This combined error rate is quantified for both Crawford and Garthwaite's criteria and for the conventional criteria.

Two forms of Type I error can occur in single-case studies. The first is where a member of the healthy, cognitively intact, population is misidentified as exhibiting a dissociation. Quantifying this error rate is fundamental: a reported dissociation in a patient must be viewed with scepticism if the dissociation was identified using a method that misclassifies many healthy individuals. Unfortunately it does not follow from this that we can be confident of the veracity of a dissociation identified by a method that misclassifies few members of the healthy control population. Rather, it can be argued that single-case researchers also need to consider another form of Type I error: namely incor-

[1] Crawford and Garthwaite (2005a, 2005b) provided a computer program that applies the inferential methods required to test whether a single-case meets their criteria for a classical or strong dissociation. This program can be downloaded from http://www.abdn.ac.uk/~psy086/dept/dissociations.htm.

rectly identifying a *patient* as exhibiting a dissociation. That is, a patient may have a strictly equivalent level of acquired impairment on the tasks of interest but be misclassified as exhibiting a strong (or classical) dissociation.

Crawford and Garthwaite (2005a) examined the error rates for this latter form of Type I error for classical dissociations. However, they did not attempt to quantify the error rate for strong dissociations. It follows that neither did they examine the *combined* error rate for dissociations (as noted, it can be argued that it is this latter error rate that should be attended to rather than the error rate for either type alone). In the present study we address both these issues (in addition to quantifying the error rate for healthy individuals).

3. Method

The Monte Carlo simulations were 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).

3.1. Procedure when Type I errors were defined as misclassifying a healthy control case as exhibiting a dissociation

The simulation was run with five different values of N (the sample size of the control sample): for each of these values of N, 1,000,000 samples of $N+1$ were drawn from one of five bivariate normal distributions in which the population correlation (ρ_{XY}) between tasks was set at either 0.0, 0.3, 0.5, 0.7 or 0.8. Thus a total of 25 million individual Monte Carlo trials were run.

On 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 taken as the scores of the individual control case. Crawford and Garthwaite's criteria were then applied to these data. That is, cases were classified as exhibiting a strong dissociation if (a) they obtained a significantly lower score ($p<0.05$, one-tailed) than controls on tasks X and Y using Crawford and Howell's (1998) test, and (b) if the standardized difference between the case's X and Y scores was significantly different ($p<0.05$, two-tailed) from the standardized differences in the control sample using Crawford and Garthwaite's (2005b) RSDT.

In order to examine the combined Type I error rate for dissociations (i.e., to record the percentage of controls incorrectly classified as exhibiting either form of dissociation) it was also necessary to test if cases met the criteria for a classical dissociation: cases were classified as exhibiting a classical dissociation using the same criteria as those set out above, except that the case had to be significantly lower than controls on either task, but not on both (Crawford & Garthwaite, 2005b).

For the z-based criteria for a strong dissociation, z was used to test for a deficit on X and Y and z_D to test the difference between the case's X and Y scores. In this latter test, commonly attributed to Payne and Jones (1957), an individual's scores on tasks X and Y are expressed in z score form and their difference divided by the standard deviation of the difference for controls; this quantity (z_D) is then referred to a table of areas under the normal curve. A control case was recorded as exhibiting a strong dissociation if z was less than -1.645 (i.e., nominal p value = 0.05, one-tailed) on both tasks and if the absolute value of z_D exceeded 1.96 (i.e., nominal p value = 0.05, two-tailed).

For the conventional (z-based) criteria, cases were recorded as exhibiting a *classical* dissociation if z was less than -1.645 on either task (but not both); as previously noted, the conventional criteria for a classical dissociation do not include a test on the difference between a case's X and Y scores.

For both Crawford & Garthwaite's criteria (2005b) and the conventional (z-based) criteria, the percentage of control cases incorrectly classified as exhibiting (a) a strong dissociation and (b) either form of dissociation was recorded.

3.2. Procedure when Type I errors were defined as misclassifying a patient with strictly equivalent deficits as exhibiting a dissociation

The procedure followed to model this scenario was identical to that described in the foregoing section with one crucial difference. As in the first simulation, $N+1$ pairs of observations were drawn from bivariate standard normal distributions, the first N pairs were taken as the scores of control sample and the $N+1$th pair as the scores of the case. However, on each Monte Carlo trial, the case was then "lesioned" by imposing an acquired impairment of two standard deviations on the scores for both tasks. These cases are used to represent patients who have suffered large but *strictly equivalent* deficits on the two tasks; i.e., they do *not* exhibit either form of dissociation. Thereafter the procedure was identical to the previous simulation; i.e., the criteria for dissociations were applied to each case and the percentage of Type I errors recorded and so forth.

4. Results and discussion

When a Type I error was defined as misclassifying a healthy control case as exhibiting a strong dissociation the error rates were uniformly low for both Crawford and Garthwaite's criteria and for the z-based criteria. For Crawford and Garthwaite's criteria the error rates ranged from less than 0.001% (for control sample sizes >10 and a population correlation between tasks of 0) to a high of 0.216% (for a control sample size of 5 and correlation of 0.8). Rates were higher for the z-based criteria but were low in absolute terms: the rates ranged from less than 0.001% (for control sample sizes >20 and a correlation of 0) to a high of 1.49% (for a sample size of 5 and correlation of 0.8).

These results could be taken to indicate that both sets of criteria for a strong dissociation are conservative: if a single-case researcher identifies a case as having a strong dissociation it is very unlikely that they will have incorrectly classified a member of the healthy control population. This then appears to cast doubt on the concern, raised in Section 1, that methods based on z (i.e., methods that treat the control sample statistics as population parameters) may provoke an unacceptably high rate of Type I errors.

It also follows from these results that, for both sets of criteria, the combined Type I error rate (i.e., the rate at which cases were misclassified as exhibiting either form of dissociation) were not appreciably different from the error rates for classical dissociations alone. For Crawford and Garthwaite's criteria the *combined* error rates ranged from 0.80% (for a control sample size of 100 and population correlation between tasks of 0.8) to a high of 2.48% (for a control sample size of 100 and correlation of 0). For the z-based criteria the rates ranged from 5.35% (for a control sample size of 100 and a correlation of 0.8) to a high of 18.69% (for a sample size of 5 and correlation of 0). A full table of results for this first simulation can be obtained from the first author.

We turn now to the results for the second simulation in which the error rates for patients with equivalent deficits on both tasks were examined. The basic pattern of results is clearly demonstrated in Fig. 1. Because the simulation varied both the size of the control sample and the population correlation between tasks, presenting the results for all these combinations would result in an overly busy graph. Therefore, the results presented are those for a correlation of 0.5 only. The full results from the simulation are presented in Table 1.

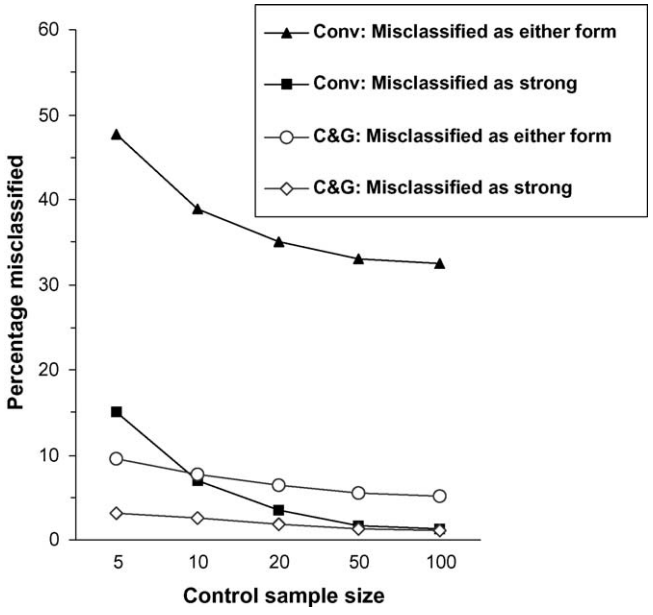

Fig. 1. Percentage of patients with equivalent deficits on X and Y incorrectly classified as exhibiting a strong dissociation or either form of dissociation using conventional (Conv) or Crawford and Garthwaite's (C&G) criteria (ρ_{XY} for these data = 0.5).

It can be seen from Table 1 that, for Crawford and Garthwaite's criteria, Type I errors for a strong dissociation (i.e., wrongly classifying a patient with equivalent deficits on the two tasks as exhibiting a strong dissociation) are appreciably higher than those observed for controls in the first simulation. However, the rates are still low in absolute terms. They range from 0.28% to a high of 5.17%; in 24 of the 25 scenarios examined the error rate is below 5%.

In contrast, the Type I error rates were much higher when the z-based conventional criteria for a strong dissociation were applied; these rates ranged from a low of 0.35% to a high of 20.7%. The error rate is below 5% in only 15 out of the 25 scenarios examined and, for most of these, the control sample size is much larger than is typically encountered in the single-case literature (i.e., ≥50). Thus, in summary, Crawford and Garthwaite's criteria for a strong dissociation are conservative. In contrast, the results obtained when z-based methods are used to test for a strong dissociation demonstrate that the concerns raised in Section 1 over high Type I error rates were well founded. That is, the results from the first simulation (indicating that few healthy controls will be misclassified as exhibiting a strong dissociation) provide little reassurance for researchers given that many false positives will be observed among patients.

The combined error rates (i.e., the percentage of patients with equivalent deficits incorrectly classified as exhibiting either form of dissociation) are also presented in Table 1. Turning first to the results for Crawford and Garthwaite's criteria, it can be seen that the error rates are substantially higher than those observed for healthy control cases in the first simulation. They range from 5.12% to a high of 9.67%. Also, unlike the first simulation, the combined error rate for dissociations is markedly greater than the rate for either type of dissociation alone (recall that in the

Table 1
Results from a Monte Carlo simulation study: percentage of patients with equivalent (two S.D.) deficits on X and Y incorrectly classified as exhibiting a strong dissociation as a function of the criteria applied, values of N of the control sample and correlation between tasks; overall percentage of patients incorrectly classified as exhibiting either a strong or classical dissociation are also presented

N	Percentage misclassified as exhibiting a strong dissociation using Crawford and Garthwaite's criteria				Percentage misclassified as exhibiting a strong dissociation using conventional criteria (z-based)				Percentage misclassified as exhibiting either strong or classical dissociations using Crawford and Garthwaite's criteria				Percentage misclassified as exhibiting either strong or classical dissociations using conventional criteria (z-based)			
	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8
5	1.69	2.45	3.21	4.39	5.17	9.37	12.5	15.1	18.6	20.7	8.79	9.31	9.51	9.67	9.63	55.1	50.9	47.7	43.9	41.4
10	1.11	1.87	2.57	3.59	4.25	3.42	5.29	6.90	9.0	10.2	7.23	7.61	7.79	7.89	7.88	49.5	43.4	38.9	33.6	30.3
20	0.64	1.23	1.85	2.62	3.13	1.37	2.43	3.44	4.68	5.45	6.13	6.36	6.46	6.39	6.28	47.5	40.4	35.0	28.9	25.1
50	0.36	0.83	1.26	1.90	2.29	0.54	1.15	1.72	2.53	3.01	5.45	5.47	5.45	5.29	5.12	46.7	38.8	33.0	26.5	22.4
100	0.28	0.66	1.07	1.65	2.00	0.35	0.80	1.27	1.92	2.33	5.19	5.17	5.08	4.93	4.74	46.4	38.4	32.4	25.7	21.7

first simulation the very low rate for strong dissociations meant that the overall rate did not differ appreciably from the rate for classical dissociations alone: this does not hold here).

Many of these error rates are sufficiently high in absolute terms to cause concern among single-case researchers and would do little to convince those who may already be sceptical observers of the single-case enterprise. However, it can be seen that the high error rates all occur when the control sample is small in size (i.e., $\leq$10); in all other scenarios the overall error rate does not rise above 6.5% (and, in most cases, is well below this figure).

This pattern of results suggests that the control samples used in single-case research should be larger than is currently typical; if the control sample size was $\geq$20 we can have a reasonable degree of confidence that an observed dissociation is not a false positive. Such a suggestion seems eminently reasonable. Given that single-case researchers believe their studies can make as great, or indeed, a greater contribution to advancing theory in neuropsychology than group-based research, they should, in the interest of increasing confidence in their findings, be willing to expend effort to recruit a decent sized sample of controls.

Turning now to the combined error rates for the conventional criteria for dissociations, it can be seen from Fig. 1 and Table 1 that the combined rates are alarmingly high in *all* scenarios studied. They range from a low of 21.7% to a high of 55.1%. Thus, in some scenarios, more than half of patients with strictly equivalent deficits would be spuriously identified as exhibiting one or other form of dissociation.

These results clearly indicate that the conventional criteria for a classical dissociation (which does not require a significance test on the difference between the case's scores), combined with *z*-based methods of inference, produces Type I errors for dissociations that are unacceptable. It should be noted that, although one can draw a distinction between the criteria for dissociations and the method used to test if they are met (e.g., one could test if conventional criteria for a classical dissociation are met using *t*-based methods rather than *z*), in practice they are commonly combined.

Finally, to our knowledge, the present study is only the second to attempt to model the rates at which *impaired* individuals will be incorrectly classified as exhibiting dissociations. This is perhaps surprising, given that this problem will constitute a common threat to validity. That is, in practice, single-case researchers are much more likely to encounter genuinely impaired individuals than individual's who are cognitively intact and, moreover, the misclassification rate for these impaired individuals will be markedly greater than the rate for the latter. Given the importance of this issue we would encourage others to study this question so that a more detailed knowledge base can be built.

5. Study 2

5.1. A Monte Carlo simulation study of the power to detect dissociations

The concern in Study 1 was with Type I errors. Such a focus was appropriate because advocates of the single-case approach have lain great emphasis on the need for methodological rigour (Caramazza & McCloskey, 1988). That is, we consider most would view the importance of avoiding spurious cases of dissociation as paramount. For example, if the broader cognitive neuroscience community, which includes many who are sceptical of the value of single-cases (Kosslyn & Intriligator, 1992; Robertson, Knight, Rafal, & Shimamura, 1993), are to accept that long-standing evidence from group studies of an association between performance on two or more tasks can be overturned by a single-case study, they would want to be confident that the dissociation is genuine.

Thus it could be argued that single-case researchers should be willing to forego detection of many genuine and potentially theoretically important dissociations in order to avoid false positives. Even so, it is surprising that the issue of statistical power in single-case studies has not been investigated empirically until recently (in the present context low power would result in a failure to detect many patients who had suffered genuine dissociations).

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

To our knowledge, the only study to date that has formally examined the power to detect dissociations was conducted by Crawford and Garthwaite (2005a). They reported that, using their criteria, moderate power to detect a classical dissociation could be achieved (maximum = 54.1%) when both the control sample *N* and the population correlation between tasks were at least moderate in size. However, (a) they did not examine the power to detect strong dissociations, and (b) nor did they examine the overall power to detect a dissociation of either form.

This latter issue may be important: just as it was argued that attention should be focussed on the *combined* Type I error rate for criteria for dissociations, we suggest that it is also appropriate to examine whether a dissociation of either type is detected rather than limiting attention to whether the dissociation is of the specified type. It may be, for example, that the power to detect the specific type of dissociation a patient has suffered is low but power to detect that they have a dissociation of some form is moderate. We should also be alert to the possibility that patients with strong dissociations may actually be *more* likely to be detected as having a classical dissociation and vice-versa.

In the present study we examine the percentage of cases correctly classified as exhibiting dissociations for both Crawford and Garthwaite's criteria and the conventional criteria. However, it should be noted at the outset that the conventional criteria generated very high rates of Type I errors in Study 1 (where the combined Type I error rate exceeded 50% in some scenarios). It is not possible to address meaningfully the statistical power of a method unless the Type I error rate is under reasonable control.

As an extreme example, if we applied a rule that all individuals (regardless of their test scores) should be classified as exhibiting a dissociation, then power would be 100% but we would of course have achieved nothing. Thus, the examination of power for the conventional criteria was primarily aimed at determining whether the pattern of results obtained for Crawford and Garthwaite's (2005b) criteria were method specific, as opposed to a reflection of more fundamental factors.

Finally, the bar is set low in these simulations: we take it as a given that power to detect dissociations in cases with mild acquired deficits will be poor. Therefore, we restrict ourselves to examining power to detect dissociations when the deficits on the impaired task (or tasks in the case of strong dissociations) are substantial.

6. Method

To examine the power to detect strong and classical dissociations we followed the same basic simulation procedures used in Study 1 (i.e., five different control sample sizes were studied in combination with five different values for the population correlation between tasks *X* and *Y*). To model patients with a genuine *classical* dissociation, a two standard deviation deficit was imposed on the cases' scores on task *X* only. The percentage of cases correctly classified as exhibiting a classical dissociation using Crawford and Garthwaite's criteria and the conventional criteria was recorded. In addition, we recorded the percentage of such cases that were classified as exhibiting a strong dissociation (i.e., we recorded the number of additional cases that were classified as exhibiting a dissociation, albeit not of the specified type).

To examine the power to detect a strong dissociation the simulation was then repeated but a two standard deviation deficit was imposed on the cases' *Y* scores and a four standard deviation deficit was imposed on the cases' *X* scores. That is, this second simulation models patients who have suffered a large deficit on one task but a substantially larger deficit on another; i.e., they have a strong dissociation. The percentage of cases correctly classified as exhibiting a strong dissociation using Crawford and Garthwaite's criteria and the conventional criteria was recorded. In addition, we recorded the percentage of such cases that were classified as exhibiting a classical dissociation.

7. Results and discussion

7.1. Power to detect dissociations using Crawford and Garthwaite's criteria

The results of the power study for Crawford and Garthwaite's criteria are presented in Table 2. This table records the percentage of cases with strong dissociations that were classified as exhibiting a strong dissociation and the percentage that were recorded as exhibiting either form of dissociation. The second half of the table also records the corresponding results for cases with classical dissociations (i.e., the percentage of cases correctly classified as classical dissociations is recorded, as is the percentage classified as either form of dissociation). These data for strong and classical dissociations are also plotted in Fig. 2a but are limited to results for a population correlation between tasks of 0.5.

It can be seen from Table 2 that the power to detect a strong dissociation is generally low: power ranges from 5.88% to 51.9%. However, it can be seen that the power to detect a case with a strong dissociation as having a dissociation of *some* form is substantially higher; power is moderate (or even high)

Table 2
Simulation study of power for Crawford and Garthwaite's criteria: percentage of patients with a strong dissociation classified as exhibiting a strong dissociation or either form of dissociation; also presented are the percentages of patients with a classical dissociation classified as exhibiting a classical dissociation or either form

N	Patients with strong dissociations								Patients with classical dissociations							
	Percentage classified as exhibiting a strong dissociation				Percentage classified as exhibiting either form of dissociation				Percentage classified as exhibiting a classical dissociation				Percentage classified as exhibiting either form of dissociation			
	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8
5	5.88	7.77	9.85	13.7	17.5	21.3	25.1	29.1	36.5	44.1	14.0	15.4	17.0	20.5	23.9	14.1	15.7	17.5	21.4	25.2
10	7.06	10.9	15.2	23.7	31.9	25.3	31.8	39.0	52.9	65.9	20.2	23.7	27.6	34.5	40.3	20.3	23.9	28.1	35.8	42.5
20	7.70	13.1	19.6	31.5	42.2	27.3	35.6	45.0	62.3	77.0	23.7	28.7	33.8	42.2	48.4	23.7	28.8	34.4	43.9	51.2
50	8.00	14.6	22.7	37.6	49.4	28.5	37.7	48.7	68.6	83.6	25.9	31.5	37.5	46.9	52.7	25.9	31.7	38.1	48.8	56.0
100	8.14	15.2	23.9	39.8	51.9	28.9	38.6	50.1	70.7	85.8	26.7	31.4	38.8	48.3	54.1	26.7	32.8	39.4	50.3	57.6

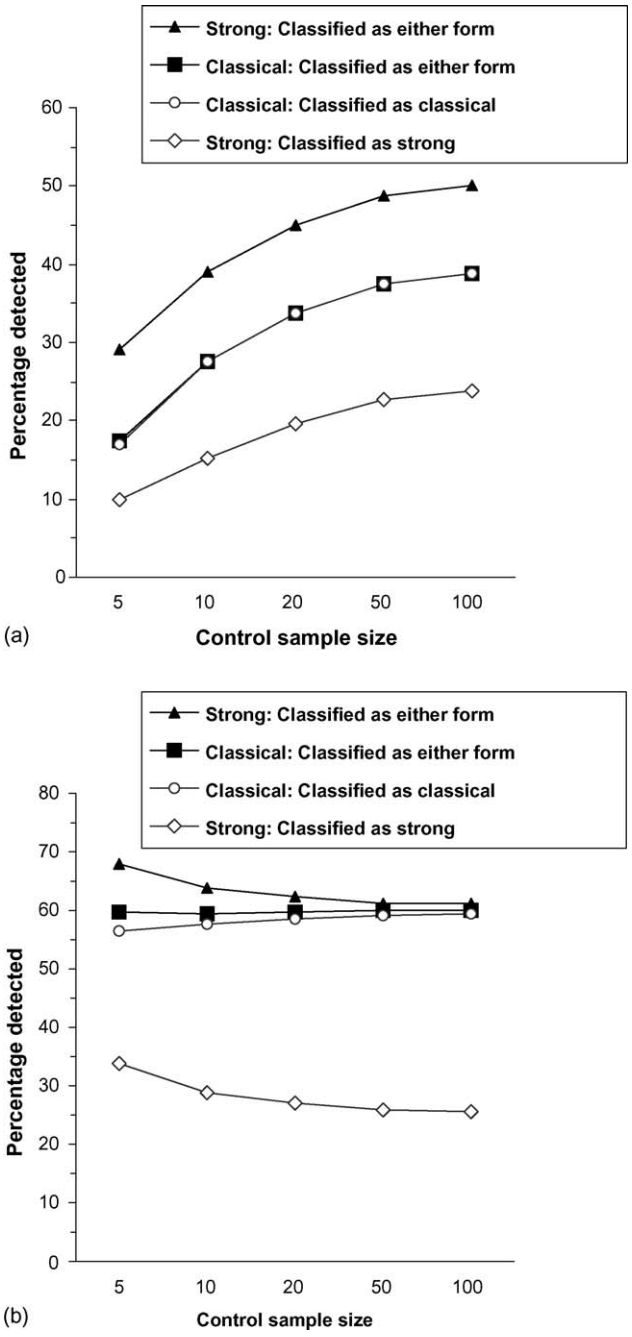

Fig. 2. Power to detect strong and classical dissociations using Crawford and Garthwaite's criteria (a) and the conventional criteria (b): percentage of cases classified correctly and percentage classified as exhibiting either form (ρ_{XY} for these data = 0.5).

when both the population correlation between tasks and control sample size are moderate or larger in magnitude. Power ranges from 21.33% to a high of 85.8%. It follows from this difference in power that, in most scenarios, strong dissociations were more commonly recorded as classical rather than strong dissociations (this is clearly illustrated in Fig. 2a).

Turning to cases with classical dissociations, it can be seen that power to detect these cases as classical dissociations ranges from low-to-moderate. Power ranges from 14% to 54.1%. In contrast to the results for strong dissociations, it can also be seen that power does not increase to any great extent when we include classical cases classified as exhibiting strong dissociations (i.e., when power is defined as identifying cases as either form of dissociation). This striking difference in the pattern of results for strong and classical dissociations is clearly illustrated in Fig. 2a.

As expected, the size of the control sample exerts a strong effect on the power to detect dissociations. It can be seen that for both strong and classical dissociations, power more than doubles in most scenarios in moving from a control sample size of 5 to a sample of 100 (marked exceptions occur for a correlation of 0 but it will be rare in practice to encounter tasks that are uncorrelated, see below for further discussion).

As noted, power will almost inevitably be at best moderate in single-case studies as an individual patient (rather than a sample of patients) is compared to a control sample and, furthermore, this sample will commonly be modest in size. 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 their premorbid competencies (Crawford, 2004).

For example, suppose that two cognitive tests have a population correlation of 0.5 (many pairs of tests used to assess neuropsychological functioning have correlations around this value). With this size of correlation 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 their illness or trauma a patient exhibited a difference of this magnitude in favour of task X over task Y. If their lesion affected the cognitive processes underlying performance on task X but entirely spared any of those underlying Y, then it is still going to be difficult to detect the effect unless it is extremely large. An acquired impairment that reduced performance on task X by one S.D. would only render scores on X and Y equivalent. An impairment of two standard deviations (a substantial decline from the premorbid score) on X would result in a discrepancy that was the mirror opposite of the patient's premorbid pattern but we have already noted that differences of this magnitude will be common in healthy, intact persons.

The method used to study statistical power captures these difficulties. The cases are first drawn randomly from the control population before they are lesioned to impose either a differential deficit on task X (in the case of a strong dissociation) or a selective deficit on task X (in the case of a classical dissociation). Therefore, on average, 50% of the cases will have had a premorbid score on X that exceeds their premorbid Y score; this "premorbid" difference has to be overcome in order for a case to be identified as exhibiting either form of dissociation.

It can be seen from Table 2 that the power to detect dissociations is much higher when the population correlation between tasks is high. For both forms of dissociation (and regardless of whether power was defined as identifying cases as exhibiting the specific form of dissociation or a dissociation of either form) power more than doubled as the population correlation moved from 0 to 0.8 in all but a few cases. 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) points out, 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).

In conclusion, to our knowledge Crawford and Garthwaite's (2005a) study was the first to attempt to quantify power for criteria for dissociations. The present study has extended their results to examine strong dissociations and to quantify the overall power to detect dissociations. If Type I errors are to be kept at an acceptable level, low-to-moderate power will be a common feature of single-case studies in which a patient is compared to a control sample. More encouragingly, 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.

7.2. Detecting dissociations using the conventional criteria

The power results for the conventional criteria for dissociations are presented in Table 3. These data are also plotted in Fig. 2b but are limited to results for a population correlation between tasks of 0.5. As noted in Study 1, when a Type I error was defined as misclassifying a patient with equivalent deficits as exhibiting a dissociation of either form, the error rates for the conventional criteria were very high. Thus the conventional criteria are fatally flawed and so, from one perspective, their power to detect true dissociations is irrelevant.

However, power was examined here in order to determine whether the results for Crawford and Garthwaite's criteria were specific to these latter criteria or were a reflection of a more fundamental factors that influence the detection of dissociations. In particular, it is of interest to examine whether cases with true strong dissociations were frequently identified as exhibiting classical dissociations. It can be seen from Table 3 and Fig. 2b that in this respect the conventional criteria produce a pattern that is very similar to that seen for Crawford and Garthwaite's criteria. A very substantial percentage of cases of strong dissociations were classified as classical dissociations, whereas classical cases were rarely classified as exhibiting strong dissociations.

8. General discussion

8.1. Control sample size in single-case studies

In discussing Type I errors in Study 1, it was suggested that control samples should be larger than is currently the norm in single-case research. This would have the virtue that (provided Crawford and Garthwaite's criteria were used) even the *combined* Type I error rate for a dissociation would be at broadly acceptable levels for both controls and patients with equivalent deficits. The results of the power study (Study 2) indicate that increasing the size of control samples will have the additional benefit of increasing the power to detect true dissociations.

It should be noted that even a moderate increase in sample sizes over those currently typical in single-case studies (i.e., to

Table 3
Simulation study of "power" for the conventional criteria: percentage of patients with a strong dissociation classified as exhibiting a strong dissociation or either form of dissociation; also presented are the percentages of patients with a classical dissociation classified as exhibiting a classical dissociation or either form

N	Patients with strong dissociations								Patients with classical dissociations							
	Percentage classified as exhibiting a strong dissociation				Percentage classified as exhibiting either form of dissociation				Percentage classified as exhibiting a classical dissociation				Percentage classified as exhibiting either form of dissociation			
	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8	0.0	0.3	0.5	0.7	0.8
5	23.6	28.9	33.8	41.5	47.6	59.7	63.9	67.8	74.8	80.7	61.6	58.4	56.4	54.8	54.4	62.9	60.6	59.6	59.9	60.9
10	16.1	22.4	28.9	39.6	48.4	52.5	58.0	63.8	74.1	82.8	61.9	59.2	57.7	56.8	56.5	62.3	60.1	59.3	60.1	61.4
20	12.5	19.4	27.2	40.3	50.6	48.8	55.1	62.4	75.2	85.5	62.2	59.8	58.5	57.8	57.7	62.3	60.3	59.7	60.5	61.9
50	9.95	17.4	26.0	41.3	52.7	46.4	53.2	61.3	76.4	87.7	62.3	60.0	59.1	58.6	58.4	62.4	60.3	60.0	60.9	62.3
100	9.15	16.6	25.7	41.7	53.5	45.5	52.5	61.1	76.8	88.7	62.4	60.2	59.3	58.7	58.7	62.4	60.5	60.0	61.0	62.4

sample sizes of 20–30) would have a marked effect on power and would reduce Type I errors. For control samples larger than this (i.e., ≥50) there are diminishing returns on both counts. Thus, to achieve a reasonable balance between statistical and practical considerations, we suggest that control sample sizes of between 20 and 30 should be employed in future single-case studies.

8.2. Broader implications for the classification of single dissociations

The present simulation results indicate that many cases of strong dissociations will be misclassified as classical dissociations, regardless of whether Crawford and Garthwaite's criteria or the conventional criteria are applied. This then raises the question of the utility of the distinction between these two forms of dissociation. We do not deny that, in theory, a clear distinction can be drawn between classical and strong dissociations, nor that there will be many patients who genuinely suffer a classical dissociation following a neurological insult. However we do suggest that, largely because of premorbid variability, in practice there are inherent difficulties in identifying whether a classical dissociation has occurred in a given case. This is liable to be a problem that *any* existing or future method will struggle to overcome.

Moreover, in the present simulations of cases with strong dissociations, the deficit on the less impaired task was still very substantial (i.e., a two standard deviation impairment was imposed on task *Y*). Despite this, the acquired impairment on *Y* was often not detected and thus many cases were classified as exhibiting classical rather than strong dissociations. If the impairment imposed on task *Y* was much less substantial (say 0.2 of a standard deviation) then such cases would be even more likely to be recorded as a classical dissociation. The probability of detecting an impairment of this magnitude is remote; for example, assuming a normal distribution for the cases' premorbid scores, approximately 42% of cases would still score above the mean of the control population. Note that in this scenario the cases still have an acquired impairment on both tasks (albeit a mild impairment on *Y*) and therefore in reality have suffered a strong dissociation rather than a classical dissociation.

In the light of the present results we suggest that the term "strong dissociation" should be retained: few patients with classical dissociations were misclassified as exhibiting strong dissociations (furthermore, few controls or patients with equivalent deficits were so misclassified). In contrast, we suggest that the term "classical dissociation" be modified to "a dissociation (putatively classical)". This proposed terminology captures the fact that (provided Crawford and Garthwaite's criteria are used in combination with a moderately sized control sample) we can have reasonable confidence that a patient identified as exhibiting a classical dissociation has indeed suffered some form of dissociation, but we cannot be confident that it is classical in type. That is, many apparent classical dissociations will in reality be cases whose deficits on the "spared" task have been missed.

The results also suggest that the crucial elements required for a dissociation are that (1) the standardized difference between the patient's scores on the tasks in question differs significantly from the standardized differences in controls (as we have seen, without this requirement false positives will be very common among both healthy controls and patients with equivalent deficits) and (2) the patient is significantly different from controls on at least one of the two tasks (i.e., there is positive evidence of an acquired impairment). In contrast, the requirement that a deficit is not recorded on one of the tasks (which provides one half of the conventional criteria for a classical dissociation and is retained in Crawford and Garthwaite's criteria) is of much more limited value for the reasons outlined above.

8.3. 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. 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 artefacts (Shallice, 1988; Vallar, 2000). The existence of a double dissociation is widely considered to largely (but not entirely) rule out task difficulty as a competing explanation. However, even if one were to accept that the double dissociation was entirely immune to task difficulty artefacts, the present simulation results serve to illustrate that it would not render a single-case study immune to another source of artefact: simple chance variation.

In Study 1, half of the patients (with strictly equivalent deficits) misclassified as exhibiting a dissociation will have exhibited a dissociation in favour 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 very close to it given the number of simulations performed). As a specific example: if the conventional criteria for dissociations were employed for tasks having a population correlation of 0.5 and the control sample size was 10, the simulation results estimate that 38.9% of patients with strictly equivalent deficits will be misclassified as exhibiting a dissociation of one type or the other (see Table 1). Therefore, under these circumstances the expectation is that approximately 19.45% of such patients would be incorrectly classified as exhibiting a dissociation in favour of *X* and an equal number would be incorrectly classified as exhibiting a dissociation in favour of *Y*. Thus, as Crawford and Garthwaite (2005a) have argued previously, although the double dissociation may largely deal with one source of artefact (differing task difficulties), it only halves the likelihood that another source (chance variation) accounts for the results observed.

References

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

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

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

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. (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: Wiley.

Crawford, J. R., & Garthwaite, P. H. (2005a). Evaluation of criteria for classical dissociations in single-case studies by Monte Carlo simulation. *Neuropsychology*, *19*, 664–678.

Crawford, J. R., & Garthwaite, P. H. (2005b). 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., & Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *The Clinical Neuropsychologist*, *12*, 482–486.

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

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.

Kosslyn, S. M., & Intriligator, J. M. (1992). Is cognitive neuropsychology plausible—the perils of sitting on a one-legged stool. *Journal of Cognitive Neuroscience*, *4*, 96–106.

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 University Press: Cambridge.

Robertson, L. C., Knight, R. T., Rafal, R., & Shimamura, A. P. (1993). Cognitive neuropsychology is more than single-case studies. *Journal of Experimental Psychology-Learning Memory and Cognition*, *19*, 710–717.

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, UK: Cambridge University Press.

Vallar, G. (2000). The methodological foundations of human neuropsychology: studies in brain-damaged patients. In F. Boller & J. Grafman (Eds.), *Handbook of neuropsychology: Vol. 1*,(2nd ed., pp. 53–76). Amsterdam: Elsevier.