

Commentary

Conclusions in Science When Theory and Data Collide

Grant J. Devilly and Joseph Ciorciari

Brain Sciences Institute, Swinburne University of Technology, Hawthorn, Victoria, Australia

We are pleased to be provided this opportunity to further demonstrate our lack of replication of DePrince and Freyd's (2001, 2004) data. DePrince and Freyd predicted and demonstrated that on a divided-attention task, students with high scores on the Dissociative Experiences Scale (DES) would show impaired recall and recognition for trauma compared with neutral words, relative to low DES scorers, and an overall net impairment for trauma words relative to neutral words. They also claimed that "high dissociators reported significantly more traumatic events . . . than low dissociators" (2004, p. 490).

In both of DePrince and Freyd's studies, they predicted, and found, an interaction effect using cell means (plotted with attendant confidence intervals). At no stage until their current Commentary (DePrince, Freyd, & Malle, 2007, this issue) have they made predictions or presented data that utilized residual means. Science requires that data shape theory, not the reverse, and that theories be clear and studies replicable.

Contrast analysis (looking at the order of the magnitude of the means for the different elements) is one useful strategy for looking at a priori hypotheses, yet, obviously, it is a statistical method that requires firm predictions. DePrince and Freyd did not appear to make their predictions explicit and found two different profiles in their two studies (and we found a third profile—twice). However, it appears they now expect that recall of trauma words in the low-DES group and recall of neutral words in the high-DES group will be greater than recall of neutral words in the low-DES group and recall of trauma words in the high-DES group. We did not obtain this contrast profile (Deville et al., 2007) and so question the reason for turning to residuals. Residuals forfeit knowable parameters. They forgo any degree of certainty (variability is not used in the method of calculation) and do not allow for confidence intervals. The Commentary of DePrince et al. leaves us with a nonsignificant interaction effect that has been reduced to residuals to demonstrate a pictorial interaction of differences around which we

have no parameters of certainty. We do not believe that this is a scientifically tenable situation. Instead, we prefer to look at factors on which we can place some degree of confidence.

DePrince et al. claim that our studies were underpowered. We estimated, a priori, g^2 values of 0.8 (requiring $n = 20$ per group) and 1.4 (requiring $n = 7$ per group), respectively, from DePrince and Freyd's (2001, 2004) published data. These effect sizes are slightly larger than those that DePrince supplied and that we presented in our report of our two studies (Deville et al., 2007, this issue); we requested the effect sizes only after the review process for that report had begun (A. DePrince, personal communication, April 7, 2006). Post hoc power analyses are inherently flawed (Hoenig & Heisey, 2001), and we appropriately included effect size confidence intervals for significant analyses that were of interest. Had we conducted our analyses with a cutoff score of 15 for the high-DES group (a cutoff 5 points lower than DePrince & Freyd's, 2001, 2004, but advocated by Steinberg, Rounsaville, & Cicchetti, 1991, as sensitive to detect dissociative disorders), our power would have increased (Study 1: $n = 23$ for the low-DES group and $n = 19$ for the high-DES group; Study 2: $n = 20$ for the low-DES group and $n = 19$ for the high-DES group), yet the interaction effect for the condition of interest would still not have been significant. In fact, looking at difference scores for recall and recognition (trauma words minus neutral words) obtained from all possible participants in our database, one can see (Fig. 1) neither a clear linear nor a systematic stepwise relation between categorized DES score and difference score. It should be noted, in particular, that no DES grouping scored a mean change into the negative, although the high-DES group did display this end result in both studies by DePrince and Freyd.

Regarding our interaction effect sizes, the lower bounds of the confidence intervals in our Studies 1 and 2 ran negative (-0.1 and -0.19 , respectively). These values contrast with those of DePrince and Freyd (0.07 and 0.54, respectively). Let us make this clear: With 95% confidence, one may conclude that an actual interaction could exist, but it may be a small effect in the direction opposite that theorized by DePrince and Freyd. Rosenthal and Rubin's (1982) formula indicates that our Study 2

Address correspondence to Grant J. Devilly, Brain Sciences Institute, Swinburne University, PO Box 218, Hawthorn, Victoria 3122, Australia, e-mail: gdevilly@swin.edu.au.

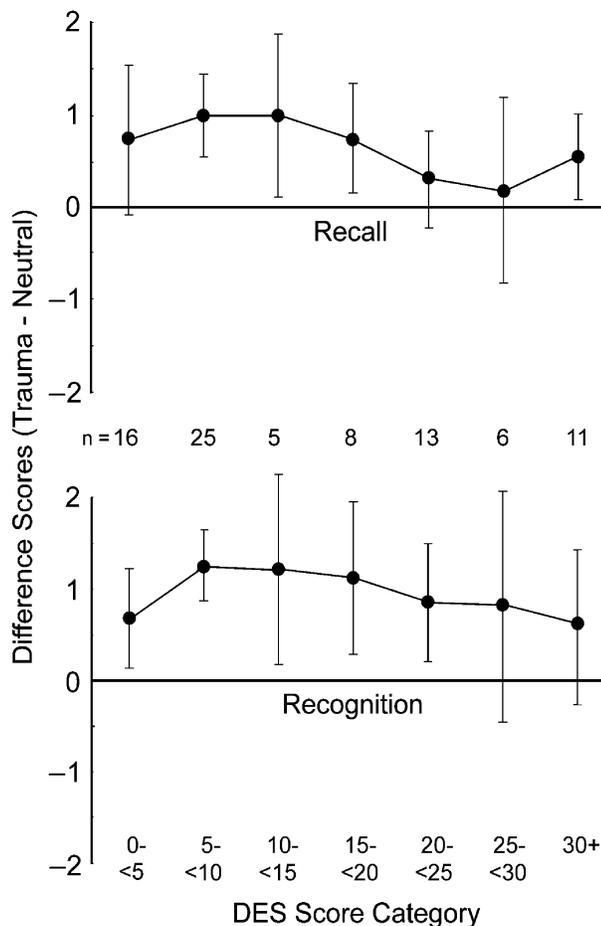


Fig. 1. Difference score (trauma words minus neutral words) in the divided-attention (motor response) condition as a function of Dissociative Experiences Scale (DES) score, for all participants in Studies 1 and 2 ($N = 84$). Results are shown separately for the recall (top) and recognition (bottom) tasks. Vertical bars denote 95% confidence intervals.

effect size is different from DePrince and Freyd's (2004) at a level of $p < .07$, one-tailed.

If DePrince et al. still prefer to see our studies as replications, we should point out that, unlike DePrince and Freyd, we did not find any significant interactions in the recognition data ($F < 0.2$) in either of our studies, and using a validated trauma measure (Posttraumatic Stress Diagnostic Scale; Foa, 1995), we found no significant correlation between number of traumas and DES score, $r(39) = .05$, n.s. In addition, we found a positive correlation between trauma severity and DES score, $r(24) = .64$, $p < .001$, and when we used trauma/no-trauma as the grouping variable (measuring trauma with the PDS or even the unvalidated Brief Betrayal Trauma Survey; Goldberg & Freyd, 2003), we found no main effects of trauma and no interaction effects ($F < 1$) for the recall and recognition tasks in either study.

Conducting a meta-analysis on the simple effects was not our test of the interaction hypothesis, as DePrince et al. erroneously

assume, but rather a demonstration of how even the elements that make up the interaction are at odds across and within research groups. DePrince and Freyd used two methods of word delivery—item and list. We did indeed use only the item method. However, no theory has been proposed as to why different contrasts would be expected using these two methods.

We are, therefore, left with a theory and data in which none of the simple elements that make up the interactions agree. The originally hypothesized interaction effect is not reliably apparent (for either recall or recognition), and our studies did not replicate the finding that people with higher DES scores report more trauma, even though we did find that they have increased memory fallibility. We have also been concerned since starting this line of research why one would present trauma, neutral, and positive words to subjects if one were going to analyze results only for neutral and trauma words (DePrince & Freyd, 2001), and why one would present a selective-attention and two divided-attention conditions if one were going to analyze results only for one divided-attention condition (DePrince & Freyd, 2004). The one area of firm agreement appears to be that people who score low on dissociation remember more trauma words than neutral words.

REFERENCES

- DePrince, A.P., & Freyd, J.J. (2001). Memory and dissociative tendencies: The roles of attentional context and word meaning in a directed forgetting task. *Journal of Trauma & Dissociation*, 2, 67–82.
- DePrince, A.P., & Freyd, J.J. (2004). Forgetting trauma stimuli. *Psychological Science*, 15, 488–492.
- DePrince, A.P., Freyd, J.J., & Malle, B.F. (2007). A replication by another name: A response to Devilly et al. (2007). *Psychological Science*, 18, 218–219.
- Deville, G.J., Ciorciari, J., Piesse, A., Sherwell, S., Zammit, S., Cook, F., & Turton, C. (2007). Dissociative tendencies and memory performance on directed-forgetting tasks. *Psychological Science*, 18, 212–217.
- Foa, E.B. (1995). *Posttraumatic Stress Diagnostic Scale (PDS) manual*. Minneapolis, MN: National Computer Systems.
- Goldberg, L.R., & Freyd, J.J. (2003). *The Brief Betrayal Trauma Survey (BBTS)*. Retrieved October 2004 from <http://dynamic.uoregon.edu/~jjf/bbts/>
- Hoening, J.H., & Heisey, D.M. (2001). The abuse of power: The pervasive fallacy of power calculations for data analysis. *The American Statistician*, 55, 19–24.
- Rosenthal, R., & Rubin, D.B. (1982). Comparing effect sizes of independent studies. *Psychological Bulletin*, 92, 500–504.
- Steinberg, M., Rounsaville, B., & Cicchetti, D. (1991). Detection of dissociative disorders in psychiatric patients by a screening instrument and a structured diagnostic interview. *American Journal of Psychiatry*, 148, 1050–1054.

(RECEIVED 9/6/06; REVISION ACCEPTED 9/10/06;
FINAL MATERIALS RECEIVED 10/3/06)