On Having One's Data Shared

Bradley R. Postle, Lawrence A. Shapiro, and Jeremy C. Biesanz

The topic of data sharing contributes prominently to the current zeitgeist in cognitive neuroimaging. This issue of *Journal of Cognitive Neuroscience*, for instance, follows closely on the heels of 5 articles on the topic that appeared back-to-back in *Nature Reviews Neuroscience* (Bjaalie, 2002; Fox & Lancaster, 2002; Koslow, 2002; Toga, 2002; Van Horn & Gazzaniga, 2002). What we offer in this commentary is a perspective "from the trenches" of researchers whose data are among the first to have been shared via the fMRI Data Center.

Lloyd's (2002) use of our data clearly illustrates one of the arguments of proponents of data sharing (e.g., Koslow, 2002; Van Horn & Gazzaniga, 2002), that most neuroscience data sets have the potential to be reanalyzed in ways not conceived of by those who originally generated them. Our data were generated to test a hypothesis about the role of the frontal eye fields and adjacent regions of the superior frontal cortex in spatial working memory function. The sequence and timing of events within each behavioral trial were constructed to permit isolation of signal evoked by particular events within each trial, and comparison of this evoked signal between various of 4 experimental conditions. Lloyd's hypothesis, in contrast, is unrelated to spatial working memory, his unit of analysis is the global signal (and is thus insensitive to the possibility of regional variation), and his analysis is blind to the different event types within each trial (and, indeed, to the distinction between trial and intertrial interval), as well as to the experimental condition of any trial. The editors have asked us to assess "the importance or otherwise of [Lloyd's] new analysis to the field." Before venturing a response, we'll note that this exercise also illustrates that data sharing need not be constrained by the boundaries of a conventionally defined field of inquiry. Indeed, the first author of this commentary (and of the original report of the shared data in question; Postle, Berger, Taich, & D'Esposito, 2000) felt compelled to enlist experts in the philosophy of mind (L. A. S.) and in multivariate statistics (J. C. B.) to participate in this assessment of Lloyd's work.

From the conceptual perspective, Lloyd's work is without question novel and important. Philosophers of mind see themselves as having made progress in solving or at least diminishing puzzles about intentionality and representation, and it is for perhaps this reason that so much philosophical attention has recently focused on the problem of consciousness. Lloyd's (2002) study suggests an approach to consciousness that, to our knowledge, marks a promising departure from the status quo. In the first place, Lloyd takes seriously work in phenomenology with which most philosophers in the analytic tradition are unfamiliar. This, by itself, is no reason to endorse such work, but if, as Lloyd suggests, phenomenology might be usefully integrated with functional brain imaging, then it is time for the analytic philosopher to think more carefully about phenomenological investigations of consciousness. Perhaps even more significantly, Lloyd's work gives the philosopher, whether of phenomenological or analytical bent, a working example of how philosophical speculations about consciousness might usefully be married to functional brain imaging.

From the empirical perspective, it is not as clear to us that Lloyd has been successful. For starters, the linear change in squared Euclidean distance that he interprets as evidence relating to the temporality of consciousness may be observed under trivial circumstances. Consider the following thought experiment. At seven regular intervals, the location of a University of Wisconsin–Madison psychologist is determined. For the first two observations (or "scans"), he is in his office. For the third, he is one block away, for the fourth, he is two blocks away, for the fifth he is three blocks away, and finally, for the sixth and seventh observations (after he has reached his lunch destination on State Street), he is four blocks away from his office. With this example, we can represent our data in a manner similar to Figure 2 from Lloyd (2002) (Figure 1a). Using this data matrix, we can also create a representation of our data analogous to Figure 3 from Lloyd by taking interobservation intervals of 1 to 6 and computing the average distance of the psychologist from his office for different observations based on that interval width (Figure 1b). There is a clear resemblance between the strong linear relationship present in this thought experiment and that illustrated in Figure 3 from Lloyd. However, consideration of the data from our thought experiment does not teach us anything meaningful about the process of going to lunch. Similarly, in the absence of a clearly articulated hypothesis, it is not clear to us how to evaluate the results of Lloyd's analyses such that we

University of Wisconsin–Madison

a.

```
      +6| 16
      +5| 16  16
Future +4|  9  16   9
      +3|  4   9   9   4
      +2|  1   4   4   4   1
      +1|  0   1   1   1   1   0
       0|  1   2   3   4   5   6   7     Scan #
      -1|      0   1   1   1   1   0
Past  -2|          1   4   4   4   1
      -3|              4   9   9   4
      -4|                  9  16   9
      -5|                     16  16
      -6|                         16
         --------------------------
           0   0   1   2   3   4   4     Distance from Office
```

b.

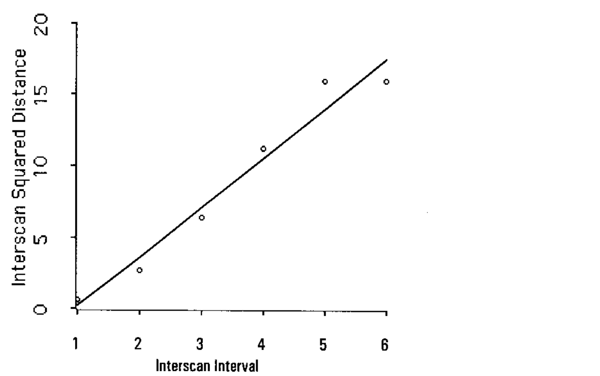

Figure 1. The perambulating psychologist. (a) Time as measured by "scans" is represented by the *x*-axis. Time is also captured on the *y*-axis with positive values indicating the number of scans in the future and negative values indicating the number of scans in the past. Tabled values represent the squared distance from the current scan to future and past scans. (b) The mean relationship between scans and multivariate distance.

can draw any strong conclusions about the nature of consciousness. It may be that the value of these results will be most usefully realized as a hypothesis-generating resource.

A different concern is that Lloyd's results may not even index consciousness-related processes. Potential confounds come from both cognitive/biological and physical sources. From the cognitive/biological perspective, it has long been known that behavior is temporally variable (Cattell, 1966; Fiske & Rice, 1955; Fiske, 1961). In the present context, one specific source of such fluctuations might be habituation to the experience of being in the bore of a scanner and performing a task. Whatever the source, most such cognitive/behavioral fluctuations are oscillatory. Nesselroade (1991) refers to this phenomenon as a "steady state 'hum' that describes the base condition of the individual" (p. 94). One would thus expect the trend of increasing multivariate distance that is characteristic of this natural variability to level off over time, and perhaps, even reverse. However, if behavior is measured on a sufficiently fine temporal scale, it could appear to be systematic change. This may be the case with the time–distance relationship observed by Lloyd (2002). That is, the time–distance relationship that Lloyd describes may simply be the result of the short time intervals at which the data were sampled, without the benefit of observing how this relationship changes over longer time intervals such as hours, days, and weeks.

An additional concern is a possible confound that Lloyd (2002) refers to alternatively as "1/*f* noise" and "scanner drift," but that he fails to address satisfactorily. We take his points in reverse order. Low-frequency structure in fMRI data, far from "elusive," is in fact ubiquitous in BOLD fMRI, regardless of the site from which the data were collected, the manufacturer of the magnet, voxel size, or the other factors that Lloyd lists. Thus, the fact that Lloyd used data generated at 4 disparate sites from 4 disparate experimental protocols does nothing to mitigate the concern about temporal autocorrelation. Secondly, the preprocessing of fMRI data by contributors to the fMRIDC would not necessarily have removed this autocorrelated component from the data, because low-frequency confounds in fMRI data are not typically addressed until inferential statistical analyses are performed. For instance, we included in the design matrices created for the analysis of data from our study a notch filter ("exogenous smoothing"), covariates of no interest that modeled "trial effects," and an empirically derived estimate of the 1/*f* noise. Thus, low-frequency confounds had not been removed from the data from Postle, Berger, et al. (2000) that Lloyd employed in his analyses.

However, having now offered a critique of this study, do we feel that we, as creators of the data that were reanalyzed by Lloyd for his study, should have any special stake or status with respect to the importance or validity of his work? Our view is that we should not. Any concerns that one may have about putatively improper handling and/or interpretation of reanalyzed data that originated in one's laboratory should be of no more or less concern than the improper handling and/or interpretation by other investigators of their own data. These are concerns about the peer review process, and are not specific to the management of shared data per se.

What are the benefits and disadvantages of data sharing? The benefits have been articulated persuasively elsewhere (e.g., Fox & Lancaster, 2002; Koslow, 2002; Toga, 2002; Van Horn & Gazzaniga, 2002). Our concerns about data sharing have primarily to do with the time and resources that are required of the sharers of data in order for the enterprise to be successful. To process and format one's data for submission to a central database exacts a cost that, until now, has not been borne by cognitive neuroimagers in their research activities. As an index of the effort required to submit data to the fMRIDC, at least 30 email messages were sent back-and-forth between B. R. P. and the fMRIDC to effect the submission and documentation of the data from (Postle, Berger, et al., 2000). To be fair, our data were among the first to be submitted and the process is likely smoother now, but our concern remains that data sharing requires a degree of standardization of procedure for data processing and formatting that does not currently exist in the neuroimaging community. To draw on another example, Fox and Lancaster (2002) argue that standardized coordinates permit more precise reporting standards, and note that they require that these standards be met for inclusion in the BrainMap database. This may be sensible from the perspective of the archivist, but not necessarily from that of the scientist. We, for instance, have principled reasons for not smoothing, normalizing, and computing group-averaged statistical maps that would permit extraction of standardized coordinates of statistical hotspots (Postle, Zarahn, & D'Esposito, 2000). (Also see Brett, Johnsrude, & Owen, 2002 for an insightful consideration of problems inherent in localization across brains.) In such an instance, whose responsibility should it be to further process our data so that they would meet the standards of the central database? If the answer is that it is the responsibility of the investigator, then we opine that data sharing will not become standard practice in the cognitive neuroimaging community until systemic change occurs to provide incentives to the investigator to meet this responsibility. This change might take the form of a set of enticements, such as additional funds from granting agencies that are earmarked for data-sharing activities. Or it might take a coercive form, such as the widespread adoption of publication policies like those at *JOCN*, which require submission of data to the fMRIDC as a condition for publication. Regardless of the means by which it is achieved, the ultimate success of data-sharing initiatives in cognitive neuroimaging depends on effecting change in our scientific culture, such that participating in these initiatives becomes a core value.

Reprint requests should be sent to Bradley R. Postle, 1202 West Johnson Street, Madison, WI 53706, USA, or via e-mail: postle@facstaff.wisc.edu.

REFERENCES

Bjaalie, J. G. (2002). Localization in the brain: New solutions emerging. *Nature Reviews Neuroscience, 3,* 322–325.

Brett, M., Johnsrude, I. S., & Owen, A. M. (2002). The problem of functional localization in the human brain. *Nature Reviews Neuroscience, 3,* 243–249.

Cattell, R. B. (1966). Patterns of change: Measurement in relation to state-dimension, trait change, lability, and process concepts. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology* (pp. 355–402). Chicago, IL: Rand McNally.

Fiske, D. W. (1961). The inherent variability in behavior. In D. W. Fiske & S. R. Maddi (Eds.), *Functions of varied experience* (pp. 326–354). Homewood, IL: Dorsey Press.

Fiske, D. W., & Rice, L. (1955). Intra-individual response variability. *Psychological Bulletin, 52,* 217–250.

Fox, P. T., & Lancaster, J. L. (2002). Mapping context and content: The BrainMap model. *Nature Reviews Neuroscience, 3,* 319–321.

Koslow, S. H. (2002). Sharing primary data: Threat or asset to discovery? *Nature Reviews Neuroscience, 3,* 311–313.

Lloyd, D. (2002). Functional MRI and the study of human consciousness. *Journal of Cognitive Neuroscience.*

Nesselroade, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change* (pp. 92–105). Washington, DC: American Psychological Association.

Postle, B. R., Berger, J. S., Taich, A. M., & D'Esposito, M. (2000). Activity in human frontal cortex associated with spatial working memory and saccadic behavior. *Journal of Cognitive Neuroscience, 12, Suppl. 2,* 2–14.

Postle, B. R., Zarahn, E., & D'Esposito, M. (2000). Using event-related fMRI to assess delay-period activity during performance of spatial and nonspatial working memory tasks. *Brain Research Protocols, 5,* 57–66.

Toga, A. W. (2002). Neuroimage databases: The good, the bad and the ugly. *Nature Reviews Neuroscience, 3,* 302–309.

Van Horn, J. D., & Gazzaniga, M. S. (2002). Databasing fMRI studies–Towards a 'discovery science' of brain function. *Nature Reviews Neuroscience, 3,* 314–318.