COMMENTARY

# Relationship Between Magnitude of Damage to the Hippocampus and Impaired Recognition Memory in Monkeys

Stuart M. Zola[1–3]* and Larry R. Squire[1–4]

[1]*Veterans Affairs Medical Center, San Diego, California*
[2]*Department of Psychiatry, University of California, La Jolla, California*
[3]*Department of Neurosciences, University of California, La Jolla, California*
[4]*Department of Psychology, University of California, La Jolla, California*

**ABSTRACT: Two recent meta-analyses, drawing on data from many of the same studies with monkeys, reached different conclusions about the relationship between hippocampal damage and recognition memory performance. Both studies found evidence of recognition memory impairment following hippocampal damage. However, Zola et al. (J Neurosci 2000;20:451–463) found no significant correlation between extent of hippocampal damage and recognition memory performance, whereas Baxter and Murray (Hippocampus 2001;11:61–71) concluded that the extent of hippocampal damage in monkeys was inversely correlated with impaired performance. Here, we first consider the requirements for carrying out a valid meta-analysis, and point out that the analysis carried out by Baxter and Murray (Hippocampus 2001;11:61–71) is invalid on simple statistical grounds. We then adopt the appropriate statistical procedures (multiple regression analyses rather than simple correlational analysis) to assess the relationship between extent of hippocampal damage and recognition performance across different studies. None of these analyses, including a reanalysis of the data of Baxter and Murray (Hippocampus 2001;11:61–71), revealed a significant inverse relationship between lesion size and behavioral impairment. Most of the variance was explained by differences between the studies that contributed to the meta-analysis, not by lesion size itself. Indeed, analysis of covariance indicated that there were differences among the studies beyond lesion size that significantly affected performance. Finally, we consider what relationship might hold between lesion size and memory performance in the monkey. *Hippocampus 2001;11:92–98.* © 2001 Wiley-Liss, Inc.†**

**KEY WORDS: recognition memory; ibotenic acid; radio frequency lesions; declarative memory; correlation; meta-analysis; multiple regression**

## INTRODUCTION

In mammals, the formation of declarative memory is impaired following damage to a system of anatomically related structures in the medial temporal lobe (Squire, 1992; Eichenbaum, 1997). The important structures include the CA fields of the hippocampus proper, the dentate gyrus, and the subicular complex (referred to collectively here as the hippocampal region, e.g., Zola et al., 2000), and adjacent cortical areas that are anatomically linked to the hippocampal region, i.e., the entorhinal, perirhinal, and parahippocampal cortices (Zola-Morgan and Squire, 1993).

Studies investigating the relationship between extent of damage to the medial temporal lobe system and severity of memory impairment have consistently found that severity of memory impairment increases as more components of the medial temporal lobe memory system are damaged. This is true for rodents (e.g., Jarrard, 1986; Morris et al., 1990), monkeys (e.g., Zola-Morgan et al., 1994), and humans (Rempel-Clower et al., 1996).

Less is known about the relationship between extent of damage and behavioral performance in the case of the individual components of the medial temporal lobe memory system. Perhaps the most straightforward scenario is that, within each structure, impairment grows progressively more severe as extent of tissue damage increases. At the same time, it is possible to imagine the opposite scenario: that a smaller lesion would be more disruptive than a larger lesion.

Recently, information on this matter became available in the case of the hippocampus. Studies of spatial memory in the rat indicated that behavioral impairment is

Grant sponsor: Medical Research Service, Department of Veterans Affairs; Grant sponsor: National Institutes of Health; Grant numbers: MH58933, MH24600.
*Correspondence to: Stuart Zola, Department of Psychiatry 0603, UCSD School of Medicine, La Jolla, CA 92093.
E-mail: szola@ucsd.edu
Accepted for publication 15 January 2001

correlated with the extent of hippocampal lesions (Moser et al., 1993): large lesions produce more severe impairment than small lesions. Importantly, a large number of animals were necessary to demonstrate this relationship between lesion size and impaired performance. For example, in one study, 104 rats with hippocampal lesions and identical test histories were used to establish the relationship between degree of memory impairment and volume of hippocampal damage (Moser et al., 1995).

During the past several years, work with monkeys has also begun to address this issue. Specifically, findings are now available from five studies that investigated the effects of damage to the hippocampal region on recognition memory, and the same test (the delayed nonmatching to sample task) was used in all five studies (Zola-Morgan et al., 1992; Alvarez et al., 1995; Murray and Mishkin, 1998; Beason-Held et al., 1999; Zola et al., 2000). Four of the studies (all but Murray and Mishkin, 1998) found impaired performance. Because the extent of damage to the hippocampal region was variable both within and between studies, one can assess the relationship between extent of hippocampal damage and recognition memory performance in an unusually large group of monkeys.

Two recent studies carried out meta-analyses based on these data, but the findings pointed to different conclusions. In the first study (Zola et al., 2000), a series of correlational analyses was carried out to assess the relationship between locus and extent of hippocampal damage and recognition memory performance, using data from 22 monkeys that had been tested in the same laboratory under the same testing conditions (the data were from three of the five studies cited above: Zola-Morgan et al., 1992, n = 4; Alvarez et al., 1995, n = 4; and Zola et al., 2000, n = 14, their groups RF2, IBO1, and IBO2). Impaired recognition memory was clearly evident, and there was no significant correlation between extent of damage to the hippocampal region and recognition memory performance (as measured by combined scores from two recognition memory tests, delayed nonmatching to sample, and visual paired-comparison task). It was suggested that to establish a correlation between extent of damage to the hippocampus and behavioral performance, it might be necessary to study even larger numbers of animals and to design the study for this specific purpose (e.g., systematically vary lesion size), as had been done in work with rats.

In the second study, Baxter and Murray (2001) used data from 26 monkeys with hippocampal lesions (Murray and Mishkin, 1998, n = 7; Beason-Held et al., 1999, n = 5; Zola et al., 2000, n = 14, their groups RF1, RF2, and IBO1)[1] to study the relationship between hippocampal damage and recognition memory performance as measured by the delayed nonmatching to sample task. From these data, they reported that extent of hippocampal damage was negatively correlated with impairment on the delayed nonmatching task: greater hippocampal damage appeared to be associated with better performance (Fig. 1). Baxter and Murray (2001) further reported that the inverse correlation was unique to the hippocampus and the delayed nonmatching to sample task. Thus, a weak *positive* correlation was reported between extent of hippocampal damage and impaired performance on another test of recognition memory, the visual paired-comparison task, and a significant *positive* correlation was reported between extent of damage to the perirhinal cortex and impaired scores on the delayed nonmatching to sample task.

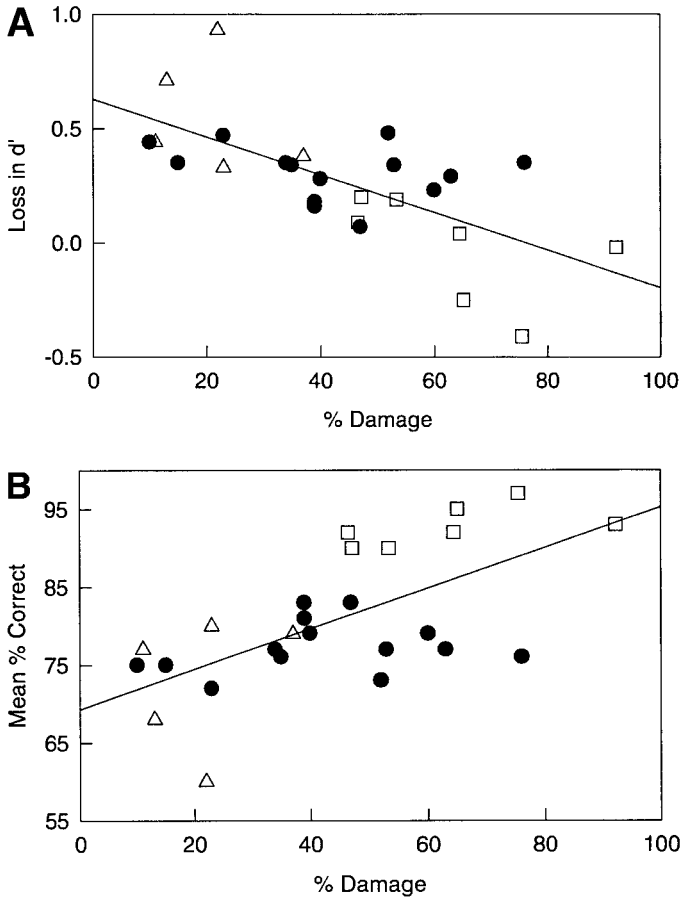


**FIGURE 1. A: Baxter and Murray (2001) combined data from three studies involving 26 monkeys with lesions of the hippocampus. We (as adapted from Fig. 3A of Baxter and Murray, 2001) illustrate an inverse correlation between performance on the delayed nonmatching to sample task (measured by loss-in-d′ scores) and percent damage to the hippocampus (including the cell fields of the hippocampus, the dentate gyrus, and the subiculum). Individual data points are shown for seven monkeys from Murray and Mishkin (1998) (squares), five monkeys from Beason-Held et al. (1999) (triangles), and 14 monkeys from Zola et al. (2000) (circles). B: Data from the same 26 monkeys as in A, illustrating an inverse correlation when percent correct scores are used as the performance measure, as in Murray and Mishkin (1998). These analyses are not valid on statistical grounds. (See text.)**

While the two meta-analyses are not directly comparable (e.g., in Zola et al., 2000, behavioral performance was measured by combining scores from two recognition tasks, and the main analysis in Baxter and Murray, 2001, used data from only one of these two tasks), they do share several features. Thus, data from the same

[1] Baxter and Murray (2001) did not use the data from the IBO-2 group in Zola et al. (2000) because this group was not tested on the delayed nonmatching task. They also did not use the ISC group in Zola et al. (2000) because the available estimates of hippocampal damage in the monkeys with ischemic lesions were limited to the CA1 field and subiculum, and did not include the other cell fields of the hippocampus. Data for the RF1 group (Zola et al., 2000) were originally reported in Alvarez et al. (1995).

14 monkeys with hippocampal lesions, were a part of both analyses. How then can we come to understand the different outcomes?

## HIPPOCAMPAL DAMAGE AND RECOGNITION MEMORY: META-ANALYSES

The overall goal of meta-analysis is to combine the results of earlier studies to arrive at summary conclusions about a body of research (Petitti, 2000). In some cases, the term meta-analysis refers to a synthesis of the outcomes of studies; in other cases, meta-analysis has been used to refer to the process of combining and reanalyzing original data from multiple studies (more appropriately called secondary analysis, e.g., Hedges and Olkin, 1985). This second approach is the one used by Zola et al. (2000), Baxter and Murray (2001), and in the present report.

Meta-analysis is deceptively simple, and on the surface involves simply combining and analyzing original data from multiple studies. However, there are important statistical and methodological considerations that guide whether and how individual studies can be combined and analyzed (Hedges and Olkin, 1985; Petitti, 2000).

In the present report, we first discuss the requirements for carrying out a valid meta-analysis. Next, working within the statistical guidelines for meta-analysis, we carry out several analyses of the relationship between delayed nonmatching to sample scores and extent of damage to the hippocampal region, including a reanalysis of the data used by Baxter and Murray (2001).

Our main analyses involved the data from three different studies involving 26 monkeys that examined the effects on delayed nonmatching to sample performance of lesions limited to the hippocampus (Murray and Mishkin, 1998, n = 7; Beason-Held et al., 1999, n = 5; Zola et al., 2000, n = 14).[2] Thus, these analyses used the same data that were used in the meta-analysis reported by Baxter and Murray (2001). Additional analyses involved the data from the three combinations of 2 of the 3 studies.

We used two different performance measures, one based on the d′ statistic (to allow a direct comparison with the analysis by Baxter and Murray, 2001), and another based on percent correct scores (to allow a direct comparison with the analysis by Murray and Mishkin, 1998). The d′ statistic is sometimes viewed as a more appropriate measure of performance than percent correct, because percent correct scores may not provide a ratio or interval scale (Ringo, 1991). This analysis involved converting each percent correct score on the delayed nonmatching to sample task to a d′ score and then, within each study, taking the difference between the performance of each monkey with hippocampal damage and the mean performance of control animals.

## THE ISSUE OF COMBINING ORIGINAL DATA FOR META-ANALYSIS

Testing the significance of the relationship between two variables in a sample drawn from a single population is a standard statistical procedure. In the present case, the question of interest is whether there is a discernible relationship across different studies between extent of damage to the hippocampus and performance on the delayed nonmatching to sample task. The correlational analysis used by Baxter and Murray (2000) to address this question assumes that the data can be viewed as arising from a single population.

In the present case, however, the studies differ both in performance scores on the delayed nonmatching to sample task and in the extent of hippocampal damage. Specifically, analysis of variance revealed significant differences in performance scores between studies, whether the performance measure was based on loss-in-d′ scores or percent correct scores ($P < 0.0001$), and there were also differences in percent damage ($P < 0.001$). The same differences emerged with nonparametric analyses (Kruskal-Wallis; all $P < 0.005$. Post hoc pairwise comparisons among the studies revealed significant study differences in almost every comparison (Table 1). Finally, performance scores between studies differed significantly even when percent damage was controlled for as a covariate (ANCOVA, $P = 0.014$ for loss-in-d′ scores; $P < 0.001$ for percent correct scores).

Accordingly, the three studies do not constitute the same population. When several studies that have significantly different means are combined (whether the differences are in the predictor scores or outcome scores), the mean differences can substantially influence the correlation of the pooled data, leading to paradoxical and even misleading results (Rosenthal, 1987). This problem was described nearly 100 years ago as Yule's paradox (Yule, 1903) and it is now well-understood that ". . .the size of the obtained correlation coefficient must be evaluated in terms of the heterogeneity of the groups measured" (Hammond and Householder, 1962). In the present case, there were substantial methodological differences among the studies, and these differences likely contributed to the observed differences in performance among studies as well as to the results of the correlational analysis by Baxter and Murray (2001). We return to this point in the Discussion.

## META-ANALYSIS BASED ON MULTIPLE REGRESSION

Multiple regression analysis, like correlation, can be used to determine the relationship between dependent and independent

[2] There has been a long-standing imprecision in the literature with respect to hippocampal terminology. In particular, different terminology was used in describing the intended lesions in each of the three studies evaluated in the meta-analysis. Murray and Mishkin (1998) used the term "hippocampus," Beason-Held et al. (1999) used the term "hippocampal formation," and Zola et al. (2000) used the term "hippocampal region." The intended lesions in each of the studies were, in fact, the same and involved the cell fields of the hippocampus proper, the dentate gyrus, and the subiculum. Baxter and Murray (2000) used the term 'hippocampus" to refer to these regions collectively. To avoid confusion, that same terminology will also be used in the remainder of this paper.

*TABLE 1.*

*Comparisons Among the Three Studies**

| Studies | % damage *P* | DNMS *P* | Regression coefficient (pooled slope) | Partial correlation | Study $R^2$ | Study plus damage $R^2$ |
|---|---|---|---|---|---|---|
| Murray and Mishkin (1998) | 0.001 | 0.0009 | −0.002 | −0.228 | 34.0 | 37.5 |
| vs. Beason-Held et al. (1999) | | 0.104 | 0.057 | 0.189 | 14.8 | 17.9 |
| Murray and Mishkin (1998) | 0.017 | 0.0001 | −0.003 | −0.368 | 50.9 | 57.6 |
| vs. Zola et al. (2000) | | 0.0001 | 0.053 | 0.308 | 85.6 | 87.0 |
| Beason-Held et al. (1999) | 0.031 | 0.002 | −0.008 | −0.453 | 63.2 | 70.8 |
| vs. Zola et al. (2000) | | 0.0001 | 0.118 | 0.294 | 77.6 | 79.5 |
| Three studies combined | 0.001 | 0.0001 | −0.004 | −0.331 | 57.9 | 62.5 |
| | | 0.0001 | 0.065 | 0.240 | 75.5 | 76.9 |

*The second column (% damage *P*) shows that the studies differed from each other with respect to extent of hippocampal damage. The third column (DNMS *P*) shows that performance scores on the delayed nonmatching to sample task (DNMS) differed among the studies. The fourth column (Regression coefficient) shows the slopes relating extent of damage to behavioral performance. The fifth column (Partial correlation) shows the correlation between extent of damage and behavioral performance after the effect of study differences has been removed. *P* values for the pooled slopes and the partial correlations were not significant (range, 0.11–0.45). The sixth column (Study $R^2$) shows the percent of variability in performance scores that was accounted for when the study from which the data came was the predictive factor. The seventh column (Study plus damage $R^2$) shows the variability in performance scores when the variable of lesion size was included as an additional predictive factor. Thus, extent of damage accounted for only a small amount of variability in performance scores (range, 1.4–7.6%) beyond the variability contributed by the factor of study differences. In columns 3–7, the upper number is based on the loss-in-*d*′ scores, and the lower number is based on percent correct scores.

variables, i.e., how accurately one variable can be predicted from another. In the present case, the multiple regression technique has an important advantage over simple correlational analyses, because it can control for the effects of mean differences in performance scores among the studies and the effects of mean differences in extent of damage scores among studies. The regression technique yields a slope, i.e., a measure of how much one variable changes with changes in the other variable after the effect of study differences has been removed. The regression procedure also yields a partial correlation, i.e., a measure of the strength of the relationship between extent of damage and performance after the effect of study differences has been removed.

We carried out two multiple regression analyses, using percent damage as the predictor. In one analysis, loss-in-d′ scores were used as the performance measure, and in the other analysis percent correct scores were used (Fig. 2). The regression analyses yielded slopes near zero, whether measured by loss-in-d′ (pooled slope regression coefficient = −0.004) or by percent correct (pooled slope regression coefficient = +0.065). The partial correlation coefficient for loss-in-d′ was −0.331, $P = 0.11$, and for the percent correct measure, 0.240, $P = 0.26$. Thus, when the combined data from the three studies were analyzed appropriately, neither the correlation coefficients nor the partial correlations differed significantly from zero.

We also computed coefficients of determination ($R^2$) in order to assess the proportion of variability in the dependent variable that could be explained by the independent variable. $R^2$ for differences in performance scores across studies was 0.579 (using loss-in-d′ scores), indicating that approximately 58% of the variability was accounted for by study differences alone. When the variable of percent damage was included, $R^2$ was 0.625, indicating that lesion size accounted for about an additional 5% of the variability in the data. When percent correct was used, $R^2 = 0.755$ for study differences, indicating that approximately 76% of the variability was accounted for by study differences alone. When the variable of percent damage was included, $R^2 = 0.769$, or 77%, indicating that lesion size accounted for an additional 1% of the variability in the data.

We also analyzed the studies in pairs, i.e., two studies at a time, using both loss-in-d′ and percent correct measures. In each case, the overall pattern of findings was similar to that just described (Table 1). Regression coefficients ranged between −0.002 and +0.118, partial correlation coefficients ranged between −0.453 and +0.308, and *P* values ranged between 0.11–0.45. Thus, when the data from the three studies were considered together, or when any 2 of the 3 studies were considered in combination, there was no significant relationship between extent of damage to the hippocampus and performance on the delayed nonmatching to sample task.

## DISCUSSION

In the present study, we used a meta-analysis combining the data from three studies involving 26 monkeys (Murray and Mishkin, 1998; Beason-Held et al., 1999; Zola et al., 2000) to assess how the extent of damage to the hippocampus might relate to performance scores on the delayed nonmatching to sample task.

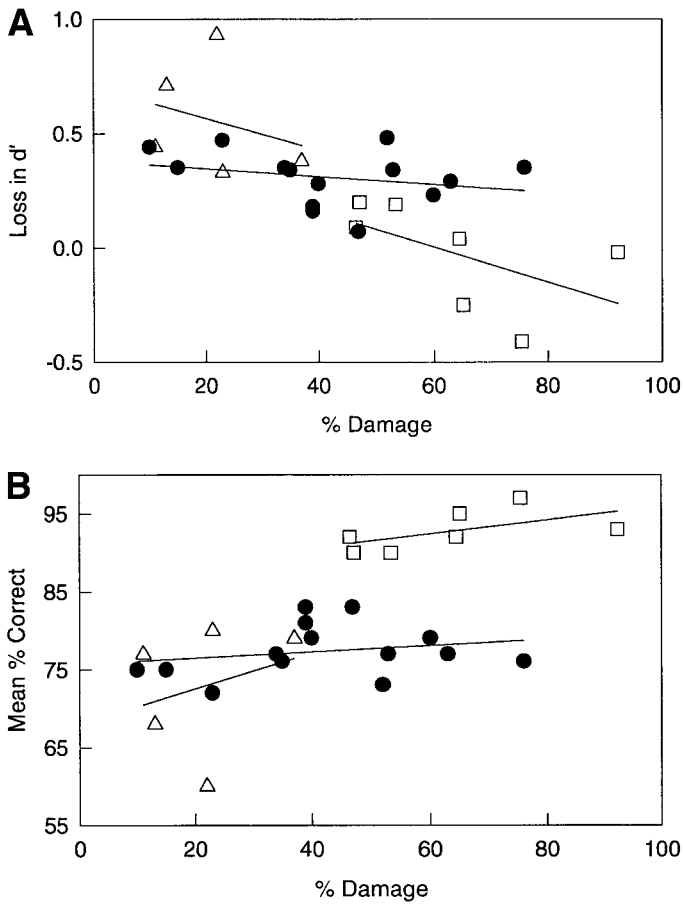


**FIGURE 2. Multiple regression analysis of the data from Baxter and Murray (2001) when performance is measured by loss-in-d′ scores (A) or by percent correct scores (B). Individual slopes are shown for each group of monkeys: Murray and Mishkin (1998) (squares); Beason-Held et al. (1999) (triangles); and Zola et al. (2000) (circles). In A and B, the pooled slopes from the studies are not significant when all three studies are combined or when any 2 of the 3 studies are combined (see Table 1).**

Multiple regression analyses yielded no significant relationship between these two variables (Fig. 2). The same finding emerged for all pairwise combinations of the three studies, whether percent correct scores or loss-in-d′ scores were used. Although in each of the studies under consideration, the slopes relating percent damage to performance went in the same direction (Fig. 2), the important point is that the pooled slope regression coefficients did not differ significantly from zero.

In contrast, Baxter and Murray (2001), using the same data set, reported a significant inverse correlation between loss-in-d′ scores and extent of damage to the hippocampus ($r = -0.673$, $P < 0.0005$; Fig. 1A). Based on these findings, they concluded that smaller hippocampal lesions were associated with greater impairments on the delayed nonmatching to sample task than were larger lesions. However, the conclusion reached by Baxter and Murray (2001) is not supportable. Specifically, their method of analyzing data from multiple studies is not valid because each of the two variables of interest (extent of damage and the performance scores) were significantly different across studies. Accordingly, the data cannot be viewed as arising from the same population, and the studies cannot be meaningfully combined for the analysis carried out by Baxter and Murray (2001).

To illustrate why a simple correlational analysis like that carried out by Baxter and Murray (2001) is not appropriate, consider hypothetical data in which two groups of animals, one with large lesions (80% damage) and one with small lesions (20% damage), have been assessed behaviorally. Additionally, the two populations differ from each other with respect to several other variables (e.g., testing method), suppose that the finding with these data is that the group with large lesions is more impaired than the group with small lesions. A simple correlational analysis of percent damage vs. behavioral impairment would likely yield a significant correlation, but the correlational analysis reveals nothing about the specific relationship between extent of damage and behavioral impairment: having completed the correlational analysis, one knows no more than one did after comparing the two groups with *t*-tests.

The more critical point is that the significant correlation (and the significant difference between groups) could be due to lesion size or to other differences in the groups (e.g., different testing methods). One could say in this case, and in any case when one combines different populations, that what one has are mean differences recast as correlations. For example, in the case of monkeys with hippocampal lesions given the delayed nonmatching to sample task, a correlational analysis reveals no more than what was already known from the original reports. The study by Murray and Mishkin (1998) did not find impairment, whereas the studies by Beason-Held et al. (1999) and Zola et al. (2000) found impairment. One difference between the studies was the size of the lesions, but there were other differences as well.

Indeed, when differences are observed across studies, as in the three studies under consideration here, one focus should be to try to understand the source of the differences (Petetti, 2000). Close analysis reveals several methodological differences among the three studies, as well as differences in the way in which the extent of damage was determined.

## Pretraining vs. No Pretraining

Unlike the monkeys in the studies by Zola et al. (2000) and Beason-Held et al. (1999), the monkeys of Murray and Mishkin (1998) received extensive preoperative training on the nonmatching task. During preoperative training, the nonmatching rule was first trained during several hundred trials using a delay interval of 8–10 s. Training on the rule provides the monkey with extended practice at holding novel objects in memory across short delays, which then might make it easier to hold novel objects in memory across the longer delays from which the performance scores for this task are derived (Bachevalier et al., 1985; Zola-Morgan and Squire, 1986; Ringo, 1988). In the context of the correlational study by Baxter and Murray (2001), it is notable that the group with the largest lesions was also the group that had been given extensive pretraining, and this group did obtain better scores than the other groups.

## One-Stage vs. Two-Stage Surgery

Unlike the monkeys in Zola et al. (2000) and Beason-Held et al. (1999), the monkeys of Murray and Mishkin (1998) were operated on in two stages, separated by at least 2 weeks. A unilateral lesion was made in the first stage, and the lesion on the other side was made in the second stage. Two-stage surgery sometimes results in less functional impairment than one-stage surgery (Finger, 1978; Finger and Stein, 1982). Although the mechanisms underlying this are poorly understood, the effect has been reported after hippocampal lesions (Stein et al., 1969; Isseroff et al., 1976). Indeed, in some cases, deficits associated with one-stage hippocampal lesions were absent altogether after two-stage surgery (Stein et al., 1969). The group that was operated on in two stages would be more likely to obtain better scores than the groups that were operated on in one stage. In the context of the correlational study by Baxter and Murray (2001), it is notable that the group with the largest lesions was also the group given two-stage surgery.

## Delay Intervals Used to Assess Memory

Performance data from the three studies that were combined for the meta-analyses consisted of delay intervals that ranged from 15 s to 40 min. However, different delay intervals were used to assess memory in each of the three studies. Zola et al. (2000) used delays of 15 s, 1 min, 10 min, and 40 min. Beason-Held et al. (1999) used delays of 2 min and 10 min. Murray and Mishkin (1998) used delays of 30 s, 1 min, and 2 min. In the context of the correlational study by Baxter and Murray (2001), it is notable that the group with the largest lesions was also the group for which the overall shortest delays were used.

## Other Differences

There were also differences among the studies with respect to the performance of the control monkeys. First, the initial learning scores of the control monkeys in each of the three studies were not comparable, indicating that the task was more difficult for some of the monkey groups. Specifically, criterion level performance (90% correct in 100 trials) was achieved in 138 trials by the control monkeys in Murray and Mishkin (1998), in 225 trials by the control monkeys in Beason-Held et al. (1999), and in 320 trials by the control monkeys in Zola et al. (2000). The scores obtained by the control monkeys in Murray and Mishkin (1998) were significantly better than the scores from the other two studies, which did not differ from each other. Second, analysis of variance revealed that the delay performance scores of the control monkeys in the three studies were not comparable (for d′, $F = 4.89$, $P = 0.026$; for percent correct, $F = 4.21$, $P = 0.039$). As in the case of initial learning scores, the delay scores obtained by the control monkeys in Murray and Mishkin (1998) were significantly better than the delay scores from the other two studies, which did not differ from each other, Thus, there were differences in control group performance scores between the studies, quite apart from the performance scores of the lesion groups.

Additionally, Murray and Mishkin (1998) and Beason-Held et al. (1999) used rhesus monkeys, while Zola et al. (2000) used cynomolgus monkeys. The age of the monkeys was approximately the same for the studies by Murray and Mishkin (1998) and Zola et al. (2000) (approximately 3–5 years of age, based on reported weights), but in the study by Beason-Held et al. (1999), ages were reported to range between 4–11 years at the start of testing.

In addition to these methodological differences that could have affected the performance scores, there were differences across studies in the way the extent of damage to the hippocampus was determined. The techniques used to assess extent of hippocampal damage were different in each laboratory, so that it is not clear how reliable the measurements of hippocampal damage were across laboratories. In particular, Beason-Held et al. (1998) reported that they likely underestimated the extent of hippocampal damage in their lesion monkeys because lesion size was calculated as a proportion of the spared hippocampus for each animal. This method results in a minimum estimate of damage because of the probable decrease in volume of the lesion resulting from shrinkage of damaged tissue (Beason-Held et al., 1999).

In summary, there were substantial differences among the studies, and these differences likely contributed to the observed study differences in performance scores and, in turn, to the impression of an inverse correlation between performance and extent of hippocampal damage. When the differences in studies were controlled for in multiple regression analyses, there was no evidence of a significant relationship between performance and damage. Even in the case of the two studies that were most similar methodologically (Beason-Held et al., 1999; Zola et al., 2000), methodological differences were nevertheless substantial (e.g., different delays, different methods of lesions analysis, and different species of monkeys).

An earlier report involving monkeys (Bachevalier and Mishkin, 1989) also hinted at the possibility of an inverse relationship between extent of damage and performance scores on the delayed nonmatching to sample task. Specifically, this study compared the effects of damage to the medial temporal lobe made by surgical lesions that involved the hippocampal formation as well as the parahippocampal cortex with lesions produced by bilateral permanent blockage of the posterior cerebral artery. The posterior cerebral artery group appeared to have more severe memory impairment than the surgical group, even though the damage identified in the posterior cerebral artery group was reported to be substantially less than the damage identified in the surgical group. A closer examination of this study, however, revealed that when only monkeys with similar testing histories were compared, the results did not support the idea that monkeys with less damage had more severe memory impairment (Squire and Zola, 1996).

It is notable that the clearest information available to date about the relationship between performance on memory tasks and extent of hippocampal damage has come from work in rats (e.g., Moser et al., 1993, 1995), where large numbers of animals were used, and where the studies were designed specifically to address this issue. In particular rats were prepared with lesions of varying size, ranging from minimal damage to nearly complete damage to the hippocampus The three studies used in the correlational study by Baxter and Murray (2001) and in the present paper were not designed to address the issue of the relationship between memory performance and extent of hippocampal damage. Indeed, the in-

tent in each study was to make substantial lesions of the hippocampus and to do so uniformly across animals.

What then is the relationship between hippocampal lesion size and memory performance in the monkey? While more data are needed across the full range of lesion size, one possibility is that the deficit increases with lesion size until the lesion is large enough to fully disable the structure. Increased lesion size beyond this point would not further increase the deficit. If so, Figure 2 suggests that the threshold for obtaining a maximum effect is about 20% damage. It is of interest that data from rats suggest a similar conclusion. In the study by Moser et al. (1993), the correlation between dorsal hippocampal lesions and water maze performance depended on data from rats with <20% damage to the hippocampus. Damage in excess of 20% did not increase the deficit.

## Acknowledgments

We thank Paul Clopton and Mark Appelbaum for statistical advice.

## REFERENCES

Alvarez P, Zola-Morgan S, Squire LR. 1995. Damage limited to the hippocampal region produces long-lasting memory impairment in monkeys. J Neurosci 15:3796–3807.

Bachevalier J, Mishkin MM. 1989. Mnemonic and neuropathological effects of occluding the posterior cerebral artery in *Macaca mulatta*. Neuropsychology 27:83–105.

Bachevalier J, Saunders RC, Mishkin M. 1985. Visual recognition in monkeys: effects of transection of fornix. Exp Brain Res 57:547–553.

Baxter MG, Murray EA. 2001. Opposite relationship of hippocampal and rhinal cortex damage to delayed nonmatching-to-sample deficits in monkeys. Hippocampus 11:61–71.

Beason-Held LL, Rosene DL, Killiany RJ, Moss MB. 1999. Hippocampal formation lesions produce memory impairment in the rhesus monkey. Hippocampus 9:562–574.

Eichenbaum H. 1997. Declarative memory: insights from cognitive psychology. Annu Rev Psychol 48:547–572.

Finger S. 1978. Recovery from brain damage: research and theory. New York: Plenum Press.

Finger S, Stein DG. 1982. Brain damage and recovery: research and clinical perspectives. New York: Academic Press.

Friedman H. 1972. Introduction to statistics. New York: Random House.

Glass GV, Hopkins KD. 1984. Statistical methods in education and psychology. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Hammond KR, Householder JE. 1962. Introduction to the statistical method. New York: Alfred A. Knopf.

Hedges L, Olkin I. 1985. Statistical methods for meta-analysis. Orlando, FL: Academic Press.

Isseroff A, Leveton L, Freeman G, Lewis ME, Stein DG. 1976. Differences in the behavioral effects of single-stage and serial lesions of the hippocampus. Exp Neurol 53:339–354.

Jarrard LE. 1986. Selective hippocampal lesions and behavior: implications for current research and theorizing. In: Isaacson RL, Pribram KH, editors. The hippocampus, volume 4. New York: Plenum Press. p 93–126.

Morris RGM, Schenk F, Tweedie F, Jarrard LE. 1990. Ibotenate lesions of the hippocampus/or subiculum: dissociating components of allocentric spatial learning. Eur J Neurosci 2:1016–1028.

Moser E, Moser M-B, Andersen P. 1993. Spatial learning impairment parallels the magnitude of dorsal hippocampal lesions, but is hardly present following ventral lesions. J Neurosci 13:3916–3925.

Moser M-B, Moser E, Forrest E, Andersen P, Morris RGM. 1995. Spatial learning with a minislab in the dorsal hippocampus. Proc Natl Acad Sci USA 92:9697–9701.

Murray EA, Mishkin M. 1998. Object recognition and location memory in monkeys with excitotoxic lesions of the amygdala and hippocampus. J Neurosci 18:6568–6582.

Petetti DB. 2000. Meta-analysis, decision analysis, and cost-effectiveness analysis. New York: Oxford University Press.

Rempel-Clower NL, Zola SM, Squire LR, Amaral DG. 1996. Three cases of enduring memory impairment after bilateral damage limited to the hippocampal formation. J Neurosci 16:5233–5255.

Ringo JL. 1988. Seemingly discrepant data from hippocampectomized macaques are reconciled by detectability analyses. Behavioral Neurosci 102:173–177.

Ringo JL. 1991. Memory decays at the same rate in macaques with and without brain lesions when expressed in d′ or arcsine terms. Behav Brain Res 42:123–134.

Rosenthal R. 1987. Judgement studies design, analysis, and meta-analysis. Cambridge: Cambridge University Press.

Squire LR. 1992. Memory and the hippocampus: a synthesis from findings with rats, monkeys and humans. Psychol Rev 99:195–231.

Squire LR, Zola SM. 1996. Ischemic brain damage and memory impairment: a commentary. Hippocampus 6:546–552.

Stein DG, Rosen JJ, Graziadei J, Mishkin D, Brink JJ. 1969. Central nervous system: recovery and function. Science 166:528–530.

Yule GU. 1903. Notes on the theory of association of attributes in statistics. Biometrika 2:121–134.

Zola SM, Squire LR, Teng E, Stefanacci L, Buffalo EA, Clark RE. 2000. Impaired recognition memory in monkeys after damage limited to the hippocampal region. J Neurosci 20:451–463.

Zola-Morgan S, Squire LR. 1986. Memory impairment in monkeys following lesions limited to the hippocampus. Behavioral Neurosci 100: 155–160.

Zola-Morgan S, Squire LR, Rempel NL, Clower RP, Amaral DG. 1992. Enduring memory impairment in monkeys after ischemic damage to the hippocampus. J Neurosci 9:4355–4370.

Zola-Morgan S, Squire LR. 1993. Neuroanatomy of memory. Ann Rev Neurosci 16:547–563.

Zola-Morgan S, Squire LR, Ramus SJ. 1994. Severity of memory impairment in monkeys as a function of locus and extent of damage within the medial temporal lobe memory system. Hippocampus 4:483–495.