The Problem of Multiple Comparisons

- Returning to the world of fMRI, let's first briefly consider two kinds of challenges associated with that technique, as they relate to Kanwisher's work on the FFA

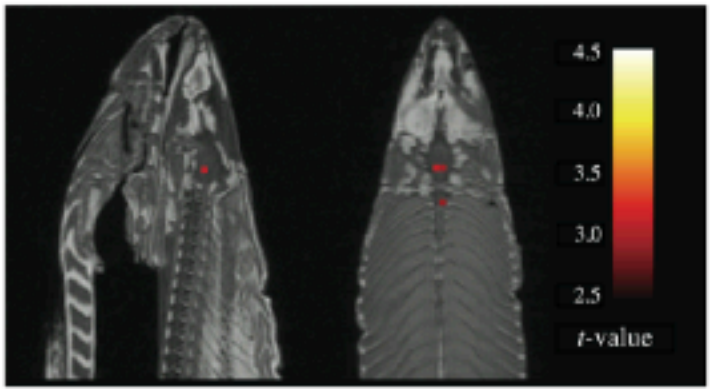
- Recall the problem of multiple comparisons raised on slide 28 of lecture 1: the dead salmon that produced a (definite!) false positive in an fMRI study using analysis methods typical of the time

Neural correlates of interspecies perspective taking in the post-mortem Atlantic Salmon: An argument for multiple comparisons correction

Craig M. Bennett[1], Abigail A. Baird[2], Michael B. Miller[1], and George L. Wolford[3]

[1] Psychology Department, University of California Santa Barbara, Santa Barbara, CA; [2] Department of Psychology, Vassar College, Poughkeepsie, NY; [3] Department of Psychological & Brain Sciences, Dartmouth College, Hanover, NH

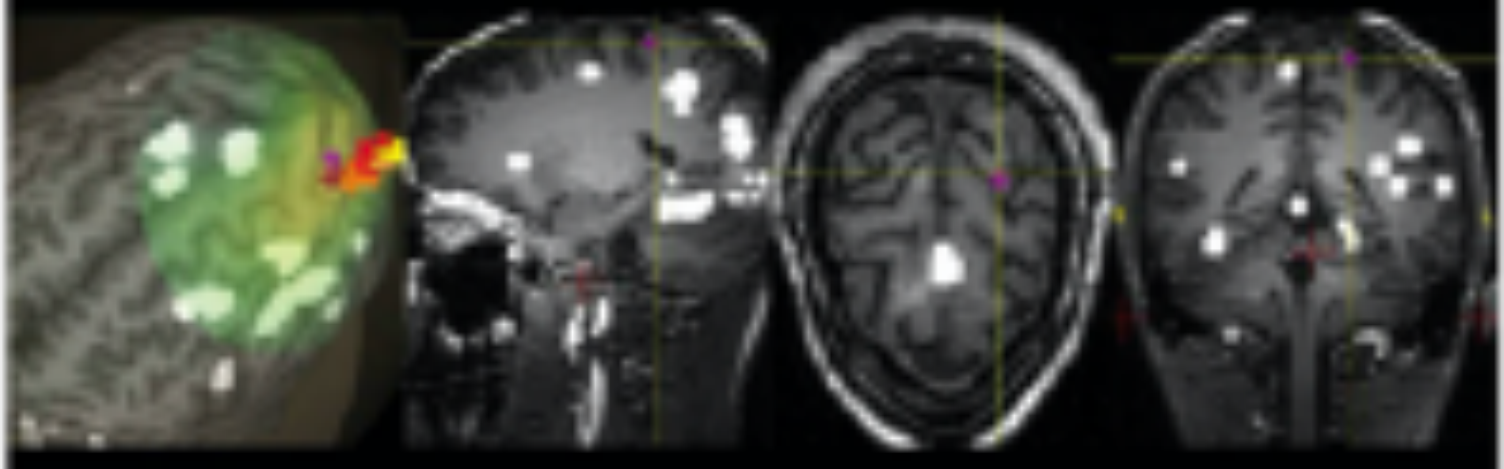
GLM RESULTS



A *t*-contrast was used to test for regions with significant BOLD signal change during the photo condition compared to rest. The parameters for this comparison were $t(131) > 3.15$, p(uncorrected) < 0.001, 3 voxel extent threshold.

We realise that some of our colleagues within the specialties of neuroscience and psychology, who we suspect could be afflicted by the aforementioned bah humbug syndrome, would argue that studies such as the present one overemphasise the importance of localised brain activity and that attempts to localise complex emotions in the brain contribute little to the understanding of these emotions. Citing a paper reporting fMRI evidence of brain activity in frozen salmon,[10] representatives of this view have even coined terms for this practice such as "blobology," "neo-phrenology," "neuro-essentialism," and "neuro-bollocks" (Grinch and colleagues, personal communication). Naturally, in keeping with the good spirit of the holiday, we disagree with these negative perspectives.

Can we conclude from this data that the salmon is engaging in the perspective-taking task? Certainly not. What we can determine is that random noise in the EPI timeseries may yield spurious results if multiple comparisons are not controlled for. Adaptive methods for controlling the FDR and FWER



ESSENTIALS OF COGNITIVE NEUROSCIENCE

Bradley R. Postle

WILEY Blackwell

METHODOLOGY BOX

9.2 Some problems, and solutions, with univariate analyses of fMRI data

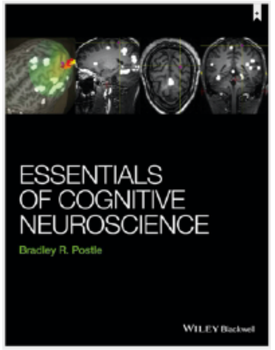
fMRI Methodology

fMRI Methodology

The Problem of Multiple Comparisons

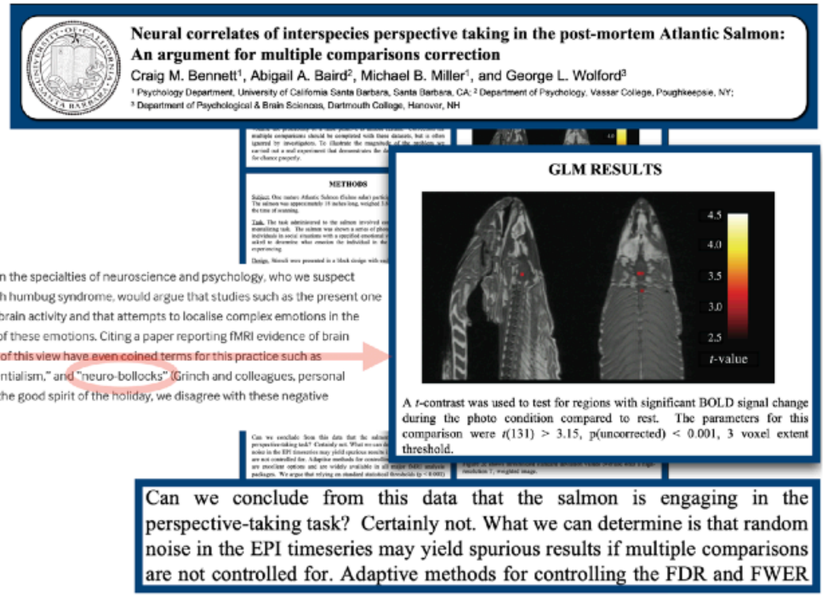
METHODOLOGY BOX

9.2 Some problems, and solutions, with univariate analyses of fMRI data



- Returning to the world of fMRI, let's first briefly consider two kinds of challenges associated with that technique, as they relate to Kanwisher's work on the FFA

- Recall the problem of multiple comparisons raised on slide 28 of lecture 1: the dead salmon that produced a (definite!) false positive in an fMRI study using analysis methods typical of the time

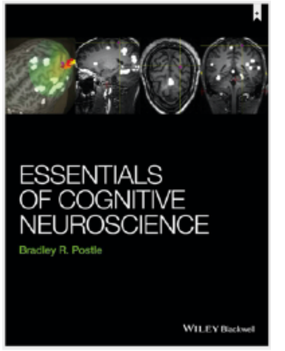


fMRI Methodology

The Problem of Multiple Comparisons

METHODOLOGY BOX

9.2 Some problems, and solutions, with univariate analyses of fMRI data



- While not correcting for multiple comparisons can increase the rate of false positives (**Type 1 error**), correcting for multiple comparisons can also make the threshold so high for achieving statistical significance that it becomes very likely that you will be unable to reject the null hypothesis even if it is not true (**Type 2 error**, or a false negative)
  - "Conventional univariate analyses of fMRI data sets are referred to as "massively uni variate," because a separate statistical test (some variant of the general linear model (GLM), comparable to a t-test) is performed at every voxel. At the level of a whole brain, which typically has > 20,000 voxels, one would expect that 1000 or more voxels would be falsely identified as being "activated" unless one were to apply a correction. The most rigorous is to divide each p value by the number of tests (called **Bonferroni correction)**. For just two voxels, this means that a critical p of .025 needs to be achieved. Not a big deal. But for 20,000, the critical p would drop to .00000025."

    (Postle, 2015)