Discussion

Hugh A. Chipman, University of Waterloo
Edward I. George, University of Pennsylvania
Robert E. McCulloch, University of Chicago

We congratulate Viele, Kass, Tarr, Behrmann and Gautier (VKTBG hereafter) for an interesting and thoughtful analysis of this dataset. The basic problem is to determine how the two Prosopagnosia subjects, CR and SM, compare to the controls in their ability to distinguish between similar objects. In the experiment considered, this ability is represented by a contrast function of each subject's hit and false alarm probabilities, namely $P(R = 1|A = 1)$ and $P(R = 0|A = 1)$, where the actual value A and subject response R take values 1 for "different" and 0 for "same". Both the hit and alarm probabilities vary across subjects S = CR/SM/Control, image type I=Greeble/Object/Face, image closeness C=Easy/Difficult, and decision time T=Brief/Long. The challenge of the analysis is to understand what aspect, if any, of this variation is attributable to Prosopagnosia. In this discussion, we first suggest a simple Bayesian analysis that sheds light on what the data may tell us, and then consider some of the difficulties that arise in modeling the response probabilities.

A Simple Bayesian Analysis

The most basic unknown quantities in this problem are the response probabilities

$$P(R = 1|A, I, C, T, S)$$

for each of the $2\times3\times2\times2\times3 = 72$ different combinations of A, I, C, T, S. All of the comparison quantities of interest are functions of these probabilities. For example, each subject's ability to distinguish images can be summarized by the simple contrast

$$\delta(I, C, T, S) = P(R = 1|A = 1, I, C, T, S) - P(R = 1|A = 0, I, C, T, S)$$

Values of δ larger than 0 indicate ability better than random guessing, and ability increases with δ.

A comparison of the abilities of the two Prosopagnosia subjects with the abilities of controls is then captured by the differences

$$\Delta CR(I, C, T) = \delta(I, C, T, S = CR) - \delta(I, C, T, S = \text{control})$$

and

$$\Delta SM(I, C, T) = \delta(I, C, T, S = SM) - \delta(I, C, T, S = \text{control}).$$

Negative values of $\Delta CR(I, C, T)$ and $\Delta SM(I, C, T)$ would indicate that CR and SM suffer from inferior visual discrimination abilities. Note that

$\Delta CR(I,C,T)$ and $\Delta SM(I,C,T)$ only vary over the 12 possible combinations of the remaining variables (I,C,T), a substantial decrease from the 72 probabilities from which these quantities were formed.

To learn about the unknown values of $\Delta CR(I,C,T)$ and $\Delta SM(I,C,T)$, we take a simple Bayesian approach. By putting independent uniform priors on each of the 72 unknown response probabilities, we use the data to induce posterior distributions on $\Delta CR(I,C,T)$ and $\Delta SM(I,C,T)$. For calculation, it is trivial to simulate these posteriors by simulating draws from the 72 response probability posteriors.

To get some feel for what the data can tell us about $\Delta CR(I,C,T)$ and $\Delta SM(I,C,T)$, we simulated boxplots of their posterior distributions in Figure 1. The induced prior on these quantities is also plotted and seems relatively noninformative by comparison. The posterior boxplots are organized into four groups corresponding to the four combinations of image closeness C=Easy/Difficult and decision time T=Brief/Long. Within each group, the three image types I=Greeble/Object/Face can then be compared.

¿From the boxplots, we see immediately that, with one exception, there is strong evidence that $\Delta CR(I,C,T)$ and $\Delta SM(I,C,T)$ are all negative indicating that CR and SM have inferior visual discrimination abilities. The one exception is SM in the (Easy, Long, Greebles) setting, where SM appears to be even better than the controls. This is all consistent with the findings of VKTBG. However, it hard to see that much more can concluded. For Brief durations, CR is worst on Greebles, but SM is worst on Faces. There seem to be some similar patterns within the Long and Brief groups, suggesting a large main effect for time. Unfortunately, such conclusions are tentative at best. If there were more Prosopagnosia subjects, it would be natural to consider using a random effects model for inference. However, given the limitation of only two Prosopagnosia subjects, it seems unreasonable to extrapolate any conclusions to the Prosopagnosia population.

It would be straightforward to modify the above analysis using a different contrast scale, replacing $\delta(I,C,T,S)$ above by

$$\delta_g(I,C,T,S) = g(P(R=1|A=1,I,C,T,S)) - g(P(R=1|A=0,I,C,T,S))$$

for some monotonic $g:[0,1]\to R$. Indeed, VKTBG use just such a contrast with the probit transform $g=\Phi^{-1}$, and the logit transform might be reasonable too. Use of the probit or logit scale allows for finer comparisons between probabilities near 0 or 1 by exaggerating small differences in those regions. Because so many of the response probabilities seem to be near 0 or 1, see Figure 2 of VKTBG, such rescaled contrasts could make a substantial difference in our above analysis, and it might be interesting to explore this further. In some contexts, such as disease modeling, such differences between very small probabilities can be quite relevant, although it is less clear to us why they are important here. In any case, an advantage of using the untransformed probability scale with $\delta(I,C,T,S)$ is that it is

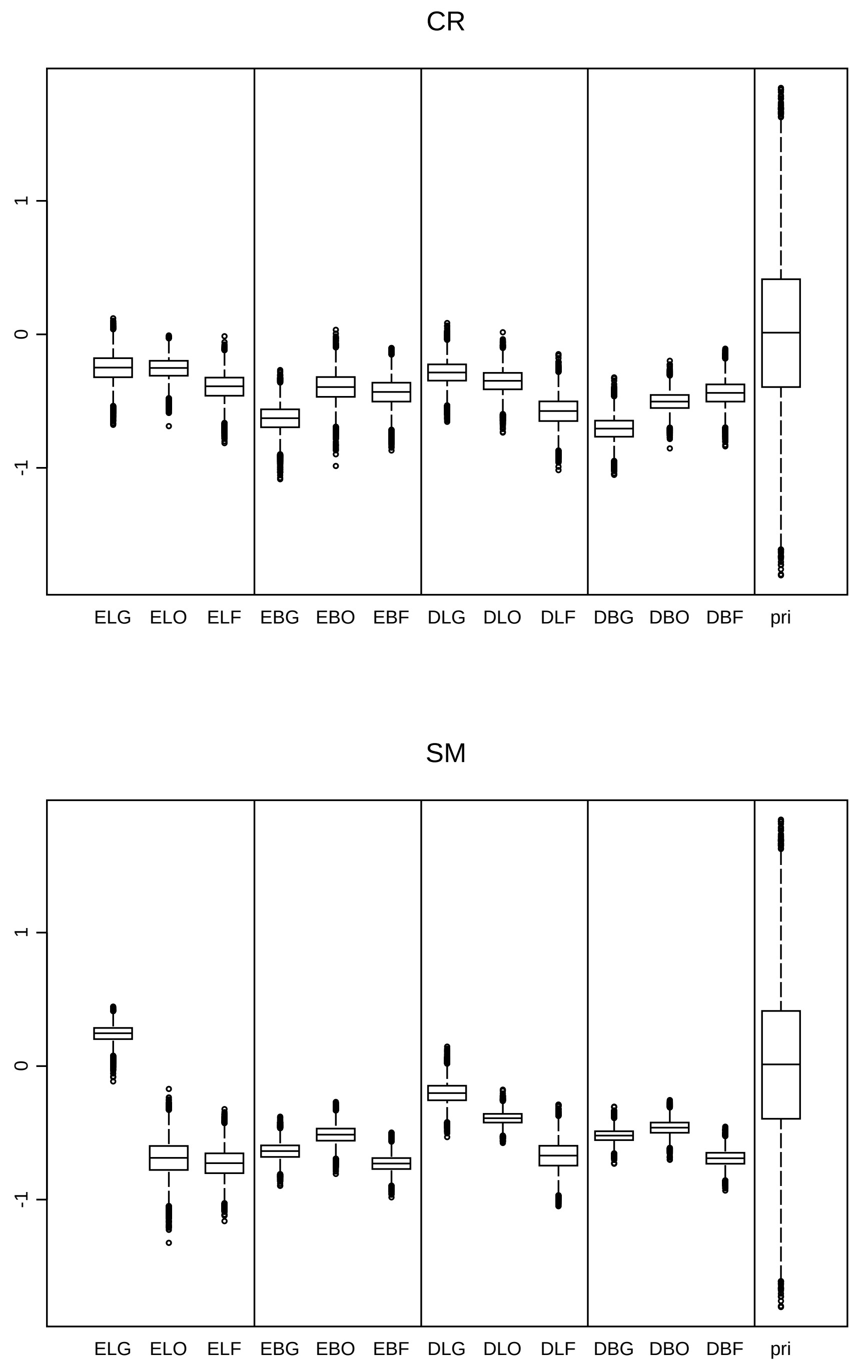

FIGURE 1. Posterior boxplots for $\Delta CR(I, C, T)$ and $\Delta SM(I, C, T)$. Cases grouped by C=Easy/Difficult, T=Brief/Long and I=Greeble/Object/Face. Prior distribution is on the right.

so easy to understand. It is interesting to note that VKTBG used the untransformed probability scale on the axes of their Figure 2, implicitly indicating its intuitive value.

Regression Modeling

To exploit the possibility of some underlying systematic structure across the 72 response probabilities, $P(R = 1|A, I, C, T, S)$, VKTBG consider selection of a probit regression model with R as the response, and A, I, C, T, S and all their interactions as potential predictors. If a large number of these potential predictors could justifiably be dropped from the model, then such a regression would provide improved contrast estimates. For this purpose, VKTBG use a hybrid stepwise-MCMC algorithm in conjunction with BIC to explore the model space and select models. Although the hybrid search algorithm seems to be a worthwhile strategy, we are skeptical that BIC is at all meaningful here. As is exemplified by the absence of a clear value for the sample size n, their use of BIC is effectively arbitrary. Furthermore, relying on an automatic approximation like BIC rather than an interpretable prior yields an uninterpretable posterior. This cannot be recommended when the goal is serious scientific inquiry.

However, coming up with a reasonable prior for this application is by no means straightforward, and we can only suggest some practical considerations, see also Chipman, George and McCulloch (2001). A possible starting point for the coefficient priors of each model M might be the normal $\beta_M \sim N(0, \tau^2 I)$ or $\beta_M \sim N(0, \tau^2 (X_M' X_M)^{-1})$, although this is not entirely realistic. For example, we would expect $P(R = 1|A = 1)$ to decrease as difficulty increases, suggesting a negative prior mean for the coefficient of difficulty. Unfortunately, incorporating such considerations across all interactions in all possible models is simply not feasible. The choice of τ^2 for this prior could be guided by studying the variation of $\Phi(a + b)$ when $\Phi(a)$ is a "ball park" choice for the response probability at a "typical" set of predictor values and $b \sim N(0, \tau^2)$ represents a coefficient in the model. One could choose a sufficiently large value for τ^2 to ensure that $\Phi(a + b)$ varies over a substantial portion of (0,1). For choosing the model space component of the prior, we would be inclined to put decreasing prior probability on higher order interactions. This could be accomplished with VKTBG's approach of restricting attention to graphical models or with the flexible approach of Chipman (1996).

Beyond the issue of variable selection, it seems that other sources of model uncertainty might be even more important. For example, one might want to consider link functions other than the probit, one might want to consider different models for $P(R = 1|A, I, C, T, S)$ when $A = 0$ or 1, and so on. Another possibility would be alternative nonlinear models such as CART, which would segment the 72 response probabilities into homogeneous groups. This could be effective if there were but a few strong interactions, and other interactions were negligible. In fact, when we first examined the data in this paper, we were hopeful that Bayesian CART (Chipman, George and McCulloch 1998) would uncover such simple structure and yield improved contrast estimates. Unfortunately, preliminary cross validation explorations indicated that very large trees (over 40 nodes out of a possible 72) do not overfit this data, suggesting a complicated interaction structure. Yet another alternative that we still plan to explore on this data is treed modeling (Chipman, George and McCulloch 2002). This would entail segmenting the response probabilities into different groups that follow different models.

References

Chipman, H. (1996), "Bayesian Variable Selection with Related Predictors", *Canadian Journal of Statistics*, 24, 17–36.

Chipman, H. A., George, E. I., and McCulloch, R. E. (1998), "Bayesian CART Model Search (with discussion)", *Journal of the American Statistical Association*, 93, 935–960.

Chipman, H., George, E.I. and McCulloch, R.E. (2001), "The Practical Implementation of Bayesian Model Selection (with discussion)", in *Model Selection* (P. Lahiri, ed.), IMS Lecture Notes – Monograph Series, Volume 38, 65-134.

Chipman, H. A., George, E. I, and McCulloch, R. E. (2002), "Bayesian Treed Models", *Machine Learning*, (to appear).