Comment on Levy et al. “Eye Tracking Dysfunction and Schizophrenia”

by William M. Grove and William G. Iacono

Abstract

We agree with Levy et al.’s conclusion (Schizophrenia Bulletin, Vol. 19, No. 3, 1993) that global quantitative measures of eye tracking dysfunction (ETD) have an important role in studies of ETD in schizophrenia. We clarify some misunderstandings of our own work contained in their review. In particular, an explanation is presented of how we computed a fit of our data to their Mendelian latent structure model, when testing for a fit to their hypothesis of essential genetic homogeneity. We also point out that our mixture analyses, as with our reported abnormality rates obtained by using cutting scores, are not in agreement with their reported rates of ETD in schizophrenia.


We were very favorably impressed with Levy et al.’s (1993) review of eye tracking dysfunction (ETD) in schizophrenia. Written by leading investigators in the field who inaugurated modern studies of eye tracking in schizophrenia (Holzman et al. 1974), the review is remarkably comprehensive and on the whole quite balanced. We were particularly gratified to see that Levy et al. are of the opinion that “no single measure or combination of measures [of ETD] has been definitively established as the most informative indicator” and that “there is ... still an important strategic place for global ... quantitative indices of eye tracking” (p. 494). This means that measures such as root mean square (RMS) error (Iacono and Lykken 1979), with which our group is identified as having used, should not be rejected out of hand, as some other reviewers have tended to do. We are inclined to agree with an anonymous reviewer of this reply, who pointed out that, although global measures (such as RMS error) are weak instruments for evaluating specific brain systems involved in the control of eye movements, as markers of familial risk they are far easier to use, better validated, and at present appear to provide as much useful information for family and genetic modeling as any specific eye movement measure.

Neither Levy et al. nor we have proposed that global measures are useful for all purposes or advocated that the sole measure examined should be a global one based, for example, on subjective ratings or RMS error. However, we believe, as we surmise Levy et al. do, that global measures have been undervalued, as though deemed invalid, useless, or simply out of date.

We wish to take this opportunity to correct a few points Levy et al. made concerning the work of our group. We do this in the spirit of clarification and not controversy, because we are in substantial agreement with most of their views.

Major Depressive Disorder and ETD

On p. 467 of their review, Levy et al. state that, whereas psychotically depressed patients in Iacono and
Beiser's Markers and Predictors of Psychosis (MAP; Iacono et al. 1992) study showed higher average RMS error than normal controls, the proportion having scores over a cutoff score (2 standard deviations [SD] over control mean) was not significantly higher than that evident for the control group (Iacono et al. 1992). Levy et al. call this a "confusing" pattern. We submit that there is nothing confusing here; this result is to be expected. Analyzing dichotomized scores typically leads to a much less sensitive test for group differences than one using continuous scores. Analysis of both continuous and dichotomized scores was performed to see whether the group difference in RMS error, which was significant, was large enough to produce a noteworthy rate of grossly abnormal tracking in psychotic major depressives. Given that neither bipolar patients nor their relatives showed ETD, and that major depressive patients and their relatives did not show excess ETD when Research Diagnostic Criteria (RDC; Spitzer et al. 1978) schizoaffective depressives were removed from the sample, we agree with Levy et al. that our findings are consistent with genetic specificity of ETD to schizophrenia.

RMS Error in the Grove et al. Study

On p. 473, it is stated throughout that "RMSE failed to distinguish [biological relatives of schizophrenia probands] from normal individuals" in a study by Grove et al. (1991). This summary requires qualification. Although the difference between the relatives and controls was not significant, there was a difference in the expected direction of over one-third SD. Levy et al. point out with respect to gain in schizophrenia that results are somewhat mixed. Some show nonsignificantly lower gain; some show significantly lower pursuit gain. We submit that the Grove et al. study gives a similar result for RMS error. It is in fact consistent with other studies on RMS error, all of which have shown that this measure differentiates schizophrenic from normal performance.

Familial Heterogeneity for ETD

On p. 475, the authors cite Clementz et al. (1992), who offered evidence for familial heterogeneity of ETD. They criticize our computation of the probability of observing the values we obtained for a (pooled) sample of schizophrenia probands' family eye tracking data under the Matthyse et al. (1986) Mendelian latent trait model. This model, together with its parameter estimates, implies genetic homogeneity. The authors state that the "details of the calculation are obscure."

If the exposition was unclear, we hope that this explanation clarifies any obscurity and reassures readers that, contrary to any possible implication from Levy et al.'s comments, our procedure was sound. As Levy et al. state, we tested the homogeneity hypothesis "by simulating a population of families with sizes and pedigree structures equivalent to [our obtained sample] and testing for goodness of fit of the simulated distribution of numbers of relatives with bad tracking to the observed distribution" (p. 475). Levy et al. then correctly point out that there has to be a reference distribution (namely, that expected if the hypothesis being tested were correct) to compute such a test.

In fact, we did obtain such a distribution, though Levy et al. claim "No reference distribution is described" (p. 475). We stated in our article: "We bootstrap[ed] the distribution [of likelihoods] to obtain a permutation test" (Clementz et al. 1992, pp. 123–124). A bootstrapped permutation distribution of likelihoods was obtained in the following manner.

For each family, we considered every possible family configuration of genotypes, among those configurations consistent with Mendel's laws, which would yield at least one sibling with a dominant genotype. This is required because, based on Matthyse et al.'s (1986) parameter estimates, one would expect only a minuscule fraction of all schizophrenia patients not to be latent trait-bearers, and all of our families contained at least one schizophrenia sibling—the proband.

The contribution of each family to the overall likelihood (of showing good vs. bad eye tracking) is merely the probability of "observing" that family's (simulated) phenotypes. This likelihood is itself a weighted sum of probabilities. Each probability entering into this sum is the product of two factors. The first factor is itself another probability: that of obtaining a given array of genotypes for mother, father, siblings, and children. It was computed, for a given family's configuration of simulated genotypes, from the allele frequency implied by the latent class base rate given in Matthyse et al. (1986) (We assumed random mating and no mutation, just as Matthyse et al. did.) The second factor is another probability: that of
observing a given set of ETD phenotypes for the family in question, conditional on their having the stated (simulated) genotypes. This was computed from the probabilities that an individual will show bad eye tracking, conditional on latent class membership (in this case tantamount to genotypes), as estimated by Matthyse et al. The weight for each product-of-probabilities entering into the sum is given as a ratio: the probability of obtaining the given configuration of genotypes, divided by the probability that one obtains any configuration of genotypes that yields at least one latent trait-bearer.

To obtain a permutation test, we required that the total number of individuals simulated to have bad eye tracking, in a given simulation sample, add up to the number of bad eye trackers in our real sample. Hence, in our computation, a given number of individuals with bad eye tracking was in effect permuted among families in a manner consistent with the Matthyse et al. (1986) model and parameter estimates. Under their hypothesis, the individuals with ETD should be spread across families; few large families should be found with no bad eye trackers. If heterogeneity prevails, however, the affected individuals can be concentrated in a few families, and one may observe even quite large families with no bad eye trackers. It is important to condition on the number of affected individuals in the sample, as we did. If we had not done so, failing to observe a good fit of our data to the Matthyse et al. hypothesis might have been attributable to our using a different threshold of ETD than theirs.

Having obtained the likelihood of bad eye tracking for a given simulated family, we found the overall likelihood for a set of families as the product of individual family outcomes (since families are observed independently). This overall likelihood was computed repeatedly, once for each of 1,000 simulated sets of families. This estimate of the permutation distribution was then compared to our real sample’s likelihood. In no case out of 1,000 simulations did we obtain a simulated likelihood even close to the observed one. (Figure 1 displays the results of this procedure.)

Sensitivity of RMS Error to Polygenic Effects

In our segregation analysis (discussed by Levy et al. on p. 476 et seq.), we directly tested their dominant gene model and rejected it. However, Levy et al. note that we defined ETD differently from their measure. They used dichotomized four-point qualitative ratings and we used quasi-continuously varying RMS error. They state that our measure is a global one and might therefore be sensitive to polygenic influences, which could account for our rejection of a single-gene-only model in favor of a mixed model (major gene plus polygenes). Although this is possible, their qualitative ratings also are global, and we are puzzled as to why one global measure, but not another, would be sensitive to polygenic effects.

Mixture Analysis of Eye Tracking Score Distributions

On p. 477 and pages that follow, Levy et al. discuss mixture analysis. We have used this methodology to test whether our observed distributions of RMS error and oculomotor gain are better ac-

---

**Figure 1. Bootstrapped permutation test for family heterogeneity**

![Figure 1](image)
counted for by hypothesizing a mixture of latent groups differing on these variables or by a single unmixed distribution. In Figure 2, we graph the results of an admixture analysis so that the reader can easily understand how this works. This analysis was based on all RMS error score data from our New York, Minneapolis, and Vancouver samples, including both probands and relatives. The estimated latent distributions are sketched atop the frequency histogram. The latent distributions were estimated by admixture analysis with SKUMIX (MacLean et al. 1976), as was done in our previous papers. The reader can see that the observed distribution is skewed to the right. The admixture analysis shows that this skewness cannot be accounted for by hypothesizing a single underlying distribution, which just happens to be skewed. Instead, the hypothesis of two distributions, mixed together so that skewing in the overall distribution is produced, is preferred (likehood ratio $\chi^2 = 38.34$, nominal $p < 0.001$—see below for discussion of correcting $p$-values).

We now reply to certain statements made by Levy et al. in their discussion of admixture analysis on p. 477 and pages that follow. First, we cannot agree with the statement on p. 478 that “It is not possible to conclude that … [data on a given eye tracking measure] do not fit a single normal distribution.” Suppose one finds that the fit of a single normal distribution to some data is much poorer than that for a two- or more-distribution mixture (e.g., $p < 0.001$ by likelihood ratio test [LRT]). Then one can say the data do not fit a single normal distribution in precisely the same way that one customarily concludes that data in a one-way analysis of variance do not fit the hypothesis that there are no group differences. In both cases (mixture model $\chi^2$ or F test) the test compares two likelihoods, from two models, one of which is nested in the other (i.e., one model can be obtained from the other by fixing one or more parameters to constant values).

We presume that what the authors meant is that the LRT cannot prove the null hypothesis, that is, the better-fitting of two nested hypotheses is in fact correct, and with this we agree. A given data set may better fit one hypothesis than another, but later turn out to be an even better fit for a third hypothesis. We also concur with Levy et al. that one should never simply assume that the distribution of data is normal. This must be checked with appropriate techniques in every case where there is any serious question about the distribution of a measure.

On p. 481, Levy et al. point out that there are problems with the asymptotic distribution of the LRT $\chi^2$ statistic, such that the significance levels for the test can differ from nominal levels, when sample sizes are small. They recommend a formula for critical values that depends on sample size (Thode et al. 1988). However, all our $\chi^2$ values are so large (whether for RMS error or gain, whether for schizophrenia probands or their relatives—the values ranged from 9.3 to 22.3) that they surpass these sample-dependent critical values, at least at the $p < 0.05$ level. Hence, their comment does not vitiate our report of significant admixture for schizophrenia probands and their relatives, for gain and/or for RMS error, in the Clementz et al. (1992)
and Iacono et al. (1992) reports.

On p. 482, Levy et al. relate that with small samples, mixture analysis may not be very powerful. They point out that Fisher's skewness test can be more powerful than the LRT for normal mixture analysis. Instead of looking for skewness in the data to indicate admixed distributions of eye tracking scores, our group has used skewness-reducing transform fitting during mixture analysis (Clementz et al. 1992; Iacono et al. 1992). In fact, our data on controls are significantly skewed. If we had interpreted skewness alone as an indication of admixture, we might be misinterpreting our data. The presence of skewness is insufficient to infer admixture. Instead, simultaneous fitting of admixed distributions, along with fitting of skewness-reducing data transformations, is required. This is what we did. The result that admixed distributions are required to account for the existing skewness confirms our hypothesis in a way that testing for skewness alone could not.

We agree with Levy et al. that larger samples are required to look at admixture when base rates of ETD are expected to be low. Our control \( n = 168 \) (in Iacono et al. 1992) is one of the largest ever reported but is still not as large as one would like.

Levy et al. comment (p. 483) on how eye tracking scores from our Minneapolis and New York samples (in pooled analyses by Clementz et al. 1992) differed in means and variances. They state that the difference in variances is expectable because in Minneapolis more relatives were sampled per proband. (This effect would be expected because siblings are correlated, whereas parents [assuming random mating for eye tracking proficiency] are not.) However, the biasing effect on the variance of sampling more siblings per family is small, unless the intraclass correlation between siblings is much larger than the size correlation we estimated. We agree that analyses of correlated observations, such as siblings, will tend, in principle, to overestimate the difference between good- and bad-tracking populations. As Levy et al. point out, however, our segregation analysis explicitly models the within-family nonindependence of eye tracking scores. This analysis yielded an estimate of the between-group difference that was over 9 SD (between homozygote means), a value even larger than that obtained from mixture analyses (3.4-4.6 SD, depending on whether one considers RMS error or gain, schizophrenia probands or relatives, and the Clementz et al. [1992] or Iacono et al. [1992] studies). Therefore, we believe that Levy et al.'s point about overestimation of the between-population differences in eye tracking scores, while generally correct, does not apply to our data.

It is stated (p. 489) that if we (Clementz et al. 1992; Iacono et al. 1992) had used a criterion for ETD such that 8 percent of our controls were abnormal (similar to the rates seen in Holzman's lab), our rates of abnormality would be "more meaningful." In fact, our rate of abnormality in controls was 6 percent (using our 2 SD cutoff, which is also near the point of rarity of our bimodal distribution of ETD scores) and close to Holzman's rate. Lowering the cutpoint to make our control abnormal rate 8 percent has little effect on our rates of abnormality in patients or their relatives. It certainly will not make our rates of bad eye tracking approach those reported by Holzman et al. (1974).

We agree with Levy et al. (1993, p. 489) that if there is a mixture in the population of individuals who latently either come from a normal-tracking group or an abnormal-tracking group (rather than just manifesting varying degrees of abnormality within a single latent group), then admixture analysis should be used to estimate the underlying proportion of abnormal trackers. Admixture analysis also yields parameter estimates, which can be used to derive the optimal cutting score for classifying individuals into latent groups, based on observed eye tracking performance. That is, admixture analysis allows statistical rather than visual estimation of the point of rarity. For our data, however, it does not make much difference whether the mixing proportion or the proportion scoring above +2 SD on the controls' distribution is reported. For example, in the Clementz et al. (1992) pooled proband sample, the mixing proportion for probands is 0.31 (gain) or 0.17 (RMS error), compared to 0.37 over -2 SD (gain), 0.21 over +2 SD (RMS error). For the Iacono et al. (1992) MAP sample, this mixing proportion for probands is estimated at 0.24, compared to 0.20 scoring over +2 SD (RMS error). We found it simpler to explain the results to readers in terms of scoring beyond the cutoff. Moreover, in these studies nothing substantive would have changed had we reported mixing proportions.

Sweeney et al. (1993) also have raised questions (different from Levy et al.'s) about our admixture analyses. These are dealt with in our reply (Iacono et al., submitted for publication) to that article, to
which the reader is referred. Their article closely replicates our earlier findings (Clementz et al. 1992; Iacono et al. 1992), thus adding another report to the literature documenting admixture in the RMS error scores of schizophrenia patients. After certain errors of inference by Sweeney et al. (1993) regarding the significance of RMS error admixture are cleared up, our position on global quantitative measures of ETD remains unchanged. Our view is that global measures such as RMS error have been well proven in family genetic studies. Their use is supportable for such investigations and all but mandatory because of their sound empirical basis. This is not to say that global measures should be the sole measures used in any eye tracking investigation. On the contrary, global measures alone do not provide sufficient information. In particular, they do not inform us about potential specific neurophysiological mechanisms underlying ETD. In this regard we think Levy et al. and our group concur.

We hope these remarks assist the readers in understanding Levy et al.’s review with respect to our work. We look forward to seeing results of their own mixture and segregation analyses and wish for a gratifying convergence of findings on major points.

References

Clementz, B.A.; Grove, W.M.; Iacono, W.G.; and Sweeney, J.A.


Iacono, W.G.; Clementz, B.A.; and Grove, W.M. “The Implications of Admixture in the Smooth Pursuit Eye Tracking Performance of Schizophrenics and Their Relatives.” Submitted for publication.


The Authors

William M. Grove, Ph.D., is Associate Professor of Psychology, and William G. Iacono, Ph.D., is Professor of Psychology, Department of Psychology, University of Minnesota, Minneapolis, MN.