ADAMS *et al.* reply—None of the above criticisms substantially alters our conclusions, though some underline the caution appropriate in certain places.

Dr Vetta has to be thanked for showing that assumptions made by Jensen in deriving a formula for heritability estimation are without theoretical justification. It follows that the formula ought not to be perpetuated. As regards the implications of this for our own report we should like to make three points: (1) We calculated heritability both with and without Jensen's false assumptions; (2) the calculation itself (we made a point of saying) was the demonstration of an *erroneous* estimation by incorporation of 'treatment effects'; (3) our main conclusion was based on the comparison of raw correlations.

Carter underestimates the generally accepted level of accuracy of zygosity determination in twins by impressionistic methods. Such methods are routinely adopted, especially under difficult conditions such as the geographic dispersal of subjects. Vandenburg[1] has claimed in a review of the approach, that "it has been shown repeatedly that for large-scale twin studies a few questions about the frequency with which close friends or relatives mistake one twin for the other will provide a sufficiently accurate diagnosis within the limits of the accuracy of the variable under study whether it be a mental test score or a physical illness". Cederlof *et al.*[2], found 98% agreement with blood-typing for responses on one of their items on a mailed questionnaire.

Carter claims, however, that merely asking mothers is too unreliable because they may have been confused by neonatal misinformation. This is quite plausible, but we need evidence about the extent of confusion, especially many years after the births. Cohen *et al.*[13] mailed a questionnaire to 35 mothers of twins in their Louisville twin study. They asked them (among other things) whether the twins were identical or fraternal, and twins were subsequently blood-typed. In only two cases did definite misclassification occur through neonatal misinformation, and there were no other misclassifications. Hence their assertion that (apart from the above) "we did not find any mother who seriously believed that her MZ pair was DZ, nor any mother of DZ twins who believed that her twin pair was MZ". Their general conclusion was that "parental perceptions of identical and fraternal twins were extremely different". The correlations for height at 7 yr in our sample (height being a trait of indisputably high heritability) were: MZ=0.926; DZ_{ss}= 0.546. We thus do not believe our twin classification was too unreliable though, naturally, caution would not be unwarranted.

Carter illustrates how non-correction for assortative mating could underestimate h^2. The illustration depends, however, on the prior assumptions of high h^2 and high genetic correlation of mates. Both assumptions become less tenable as MZ and DZ correlations approach identity, as is the case in our sample. He does not explain, moreover, why Jensen[4] estimate of pDZ=0.55 "seems implausibly low". The simple fact is that values higher than that applied to previous twin correlations yielded impossible h^2 values >1.00. Both forms of special pleading (high Δrs implying low pDZ; low Δrs implying high pDZ) necessary to preserve h^2=0.8 indeed suggests a limitation of the method, but this does not diminish the finding of insignificant Δrs as an empirical contribution. Our sample size was comparable with (if not larger than) previous studies; bigger ones are unlikely to be forthcoming without transgressing very stringent sampling conditions, satisfying which was our study's main strength.

Harvey Goldstein quotes our conclusion, based on comparison of MZ and DZ_{ss} correlations, and then claims that our "heritability estimate of 0.373" is consistent with an upper confidence limit of 0.6. It should be clear from our paper, however, that that estimate is based on MZ and DZ_t correlations, for non-verbal correlations only, calculated purely to illustrate the inflating consequences of introducing 'treatment effects'. It would be quite indiscriminate to transpose that h^2 value and its 'guesstimated' confidence limit to MZ–DZ_{ss} differences which (1) are not significantly different, (2) are quite different as regards verbal and non-verbal measures, and (3) probably themselves contain treatment effects. Our conclusion of "supportive evidence for zero or low heritabilities" seems to us a more circumspect one.

Bagley seems vague about test validity. The test we used has received some validation and a reference was given. It has been administered to two very large nationally representative samples of 11-yr olds so that excellent information is available about its distributional properties and correlations with other measures, teachers' impressions and so on. It should be pointed out that there is no universally validated and standardised intelligence test in this country, which explains the interest in the long-gestating British Intelligence Scale. As well as face-validity, ultimate validation is always by school attainments/teachers' perceptions; the items in our test conformed closely to the half dozen or so types used in conventional intelligence tests[5]. Bagley is quite wrong to uphold the Claridge *et al.* study as a model. For example, vocabulary test is not a usually accepted intelligence test and the Progressive Matrices has been far from satisfactorily validated and is also a group test. Moreover, it is a test for subjects of '11 years plus', whose use on very much older people for the purpose in mind is dubious. It is the problem of testing such a wide age-range of twins (16–55 yr) especially if the ranges for MZ and DZ twins are quite different, that has flawed previous studies and makes ours (of uniform age) unique. Furthermore, the small number of DZ pairs in the Claridge *et al.* study did not allow separation of opposite-sex and same-sex twins so that their (significant) correlation differences could quite easily be explained by the sort of 'treatment effects' which our own study suggests. Finally, the contamination effects of which Bagley speaks, relating to same-sex twins, are as applicable to MZ as to DZ twins, making separation of DZ_{ss} and DZ_{os} correlations all the more desirable.

Francis is wrong to think that we used our data primarily to suggest the dismissal of previous estimates of heritability. It is the doubts about the empirical sufficiency of previous estimates which have suggested their dismissal. Accordingly, it seems quite wrong to treat our findings as co-extensive with previous ones as if they were of equal empirical sufficiency. Again the Newman, Freeman and Holzinger study sample twins of a very wide range of ages and their small numbers of pairs did not allow separation of the DZ_{os} and DZ_{ss} pairs. Similarly, the well known Erlenmeyer-Kimling and Jarvik study includes studies of widely different measures, many with very small samples, wide age ranges, etc. Thus we dispute Francis's inference that our results, from a nationally representative group of the same age, can be explained away by 'extreme sampling'; indeed, speculation of that sort seems a pointl .xercise. Of course a changing 'population characteristic', and a 'labile heritability', is a possibility, but debatable only with comparable studies; the studies of Cohen *et al.*[3] and several others, moreover, suggest that identical twins are still treated much more similarly than are fraternal twins.

B. ADAMS
M. GHODSIAN
K. RICHARDSON

National Children's Bureau,
8 Wakley Street,
London, UK

1 Vandenburg, S. G. in *Genetics* (ed. Glass, D. C.) (Rockefeller University, 1968).
2 Cederlof, R. *et al. Acta genet. basel.* **11**, 338 (1961).
3 Cohen, D. J. *et al. Archs. gen. Psych.* **29**, 465 (1973).
4 Jensen, A. R. *Proc. natn. Acad. Soc. USA* **58**, 149 (1967).
5 Vernon, P. E. *Intelligence and Attainment Tests* (University of London, 1966).