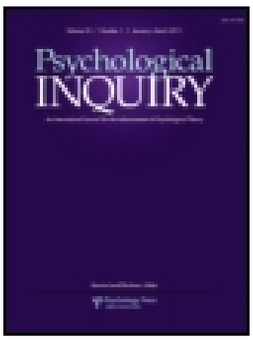

Psychological Inquiry

ISSN: 1047-840X (Print) 1532-7965 (Online) Journal homepage: http://www.tandfonline.com/loi/hpli20

Reply to Criticisms of the Grossarth-Maticek Studies

Hans J. Eysenck

To cite this article: Hans J. Eysenck (1991) Reply to Criticisms of the Grossarth-Maticek Studies, Psychological Inquiry, 2:3, 297-323, DOI: 10.1207/s15327965pli0203_21

To link to this article: http://dx.doi.org/10.1207/s15327965pli0203_21

Psychological Inquiry
1991, Vol. 2, No. 3, 297–323

AUTHOR'S RESPONSE

Reply to Criticisms of the Grossarth-Maticek Studies

Hans J. Eysenck
Institute of Psychiatry
University of London

I found the critiques following my target article very interesting and helpful; they often echoed criticisms I had myself made at various times. In this reply, I divide the argument into three parts. First, I outline very briefly the history of my own association with the Grossarth-Maticek studies; such an account is essential for understanding what happened and for appreciating the nature and import of the criticisms to be answered. Second, I try to deal specifically with all the critics, in as much detail as possible. Third, I try to formulate the conclusion that I think may be drawn from the published studies and the suggestion for future work that they may contain.

History of My Association With the Studies

As recounted elsewhere (Eysenck, 1990), my interest in the topic of cancer-prone personality began when I had shown that cigarette smoking was significantly correlated with personality, specifically with extraversion, and followed a lawful and predictable course. This finding naturally led me to look at the so-called smoking-related diseases, and I was lucky enough to meet Dr. D. Kissen, a British oncologist who ran a chest clinic in Edinburgh and was interested in the topic. We designed and carried out what I believe was one of the first objective tests of the old established theory that cancer was somehow connected with suppression of emotion and with difficulties in expressing emotions.

The degree of originality of such ideas is difficult to assess. Rosch (1979, 1980a, 1980b) has quoted various physicians expressing what seem quite modern ideas, yet who were writing hundreds, sometimes thousands of years ago; indeed, the relevance of stress and personality to the development of cancer and other diseases was taken for granted, and Claude Bernard's concept of the *milieu interne* gave expression to this belief. As Osler (1906) put it in his unique fashion: "It is many times much more important to know what kind of patient has the disease than what kind of disease the patient has (p. 14).

Using a neuroticism scale I had designed, we predicted that low scores, indicative of suppression of emotions, would be predictive of cancer, high scores of absence of cancer. I was somewhat doubtful; the theory seemed based on casual observation, and the test of doubtful relevance—one can get low scores in other ways than by suppression of emotion! The result, however, comparing patients diagnosed as suffering or not suffering from lung cancer after administration of the test (Kissen & Eysenck, 1962), was pretty conclusive in strongly supporting the theory; Kissen went on to replicate the result on other samples. As he summarized the results, "lung cancer mortality rates of those with a poor outlet for emotional discharge may be five times greater than those with a good outlet" (Kissen, 1964, p. 215). Many others have since found similar results (Eysenck, 1985).

The method used by Kissen and Eysenck was primitive but has since received independent support (Kreitler & Kreitler, 1990). For those who distrust questionnaire studies of such complex personality function, there is now also available more direct experimental proof (e.g., Esterling, Antoni, Kumar, & Schneiderman, 1990; Kneier & Temoshok, 1984). Such multimethod, multicriterion verification is particularly welcome in a field where "failure to replicate" is only too frequent.

The Kissen and Eysenck (1962) result is surprising when we remember that suppression of emotion is only one of two suggested risk factors, the other being poor coping when faced with stress, leading to feelings of hopelessness and helplessness. This second variable was also found correlated with (cervical) cancer, giving an even better prediction than the Kissen studies (Schmale & Iker, 1971). There also appeared the suggestion that personality was predictive of cancer generally, not only of one specific type of cancer; many of the replications by others of the original Kissen and Eysenck study worked with cases of mammary carcinoma. Curiously enough, these results were never criticized as being "too good to be true" although together they would appear to predict cancer even better than Grossarth-Maticek claims to do on the basis of questionnaires embracing both theories and custom-built to do so (Eysenck, 1990).

Also on the issue of "too good to be true," consider some data reported by Cohen, Kamarck, and Mermelstein (1983) on their Perceived Stress Scale. In three different groups, they found correlations of .52, .65, and .70 with physical symptomatology; this is as close or closer than most of Grossarth-Maticek's results and was achieved with a much shorter and less comprehensive scale. Studies such as this must be borne in mind in assessing Grossarth-Maticek's results.

Kissen's untimely death caused me to look around for oncologists with whom I could work to continue and improve these early studies. The reaction of the medical establishment was entirely negative; nobody had any interest in working with me, or even in allowing me to approach and test their patients. I also wanted to extend these studies from patients

already complaining of difficulties (although not yet diagnosed) to healthy people showing the cancer-prone personality. This too proved impossible through lack of interest on the part of the establishment, and I had to give up for the time being.

I was delighted when I read in the journals that Dr. Ronald Grossarth-Maticek in Heidelberg had actually carried out such work, first in his native Yugoslavia, then in Germany, with very positive results. I went to see him and was shown mountains of original data, some analyzed, others not. He not only allowed me to look at his original data, he almost forced me to inspect them in great detail. I became interested and decided to follow up the scent. I soon learned that there were many critics who rejected his studies outright, for various reasons that, if true, would leave little of interest standing. I made it my first task to follow up these criticisms and check them out. Let me only mention three objections that seemed particularly damaging.

1. At an international conference where Grossarth-Maticek had delivered a lecture on his work, critics alleged (not in his hearing) that the data must have been manufactured by him because, not being employed at a German university, he would have no access to the death certificates of his probands. This was asserted as a fact and certainly caused many participants who would otherwise have been interested in his work to shy away. An interview with the mayor of Heidelberg disclosed that he had given special permission for Grossarth-Maticek to have access to all the death certificates in question. In addition, Dr. W. D. Heller, an independent assessor from the Statistical Institute of the University of Karlsruhe, checked all the death certificates and saw to it that they were entered according to International Classification of Diseases (ICD) protocol.

2. A high official of the German Institute for Cancer Research tried to get in touch with Dr. M. Jankovic, who had collaborated with Grossarth-Maticek in his Yugoslav study, conducting medical investigations. He phoned the wrong number, decided there was no such person, and immediately (without checking with Grossarth-Maticek) spread the news among his numerous colleagues that Grossarth-Maticek had invented the nonexistent Jankovic! I had no difficulty in meeting this nonperson in Mannheim, where he was working in a hospital.

3. I was told "on good authority" that Grossarth-Maticek was *persona non grata* in his own country and was more or less disowned even by his former co-workers. I was invited, together with him, to give a lecture to the Academy of Sciences in Belgrade; it attracted more people, or so I was told, than the great Tito had done himself, and discussion with the president and leading oncologists established that Grossarth-Maticek was held in the highest esteem. (See Bachman, 1981, for a discussion of these and other untrue allegations.)

It will be clear that these and other quite false accusations created an atmosphere where any sober appreciation and discussion of Grossarth-Maticek's work were quite impossible. This does not prove that Grossarth-Maticek is right in what he asserts; it merely illustrates the existence of a situation in which the accused is asked to prove his innocence, in the teeth of highly emotional, serious charges that have no substance in fact, and are transmitted verbally, thus never giving Grossarth-Maticek a chance to answer them.

Add to this the determined opposition of the establishment to anything Grossarth-Maticek might wish to do. His applications for grants were rejected, his manuscripts were sent back by journals, young oncologists and epidemiologists who showed signs of interest in his work were warned off, and endangered their careers if they persisted, newspapers were "authoritatively" warned against his heresies. It will be appreciated how easy and inviting it would have been for me to depart with vague expressions of interest, rather than to invite the wrath of the "Cancer Mafia" (Bachman, 1981).

In addition, there was the fact that his published work was far from perfect. Many of the criticisms made by commentators of the target article were pretty obvious and certainly detracted from the work. There were many aspects of the work about which I could only say that I would not have done it that way, and I regretted that I had not been there when decisions about methodology or statistical analysis were made.

Should the rat refuse to enter the sinking ship? Grossarth-Maticek told me that indeed the ship was sinking and that he was at the end of his tether. There was no money to continue the research, no prospect of any grants, universal hostility among the experts—what could he do? I thought that, although the negative aspects of his research were real enough, there were countervailing considerations that had considerable power. Clearly, here was a unique set of data, collected in a prospective design of great power, apparently predicting cancer and coronary heart disease (CHD) with an accuracy that matched that achieved by Kissen and by Schmale and Iker, while testing theories similar to theirs. Could nothing be done with these data? Even more important from the social point of view was the evidence suggesting that behavioral methods of stress management, following clear theoretical principles, could prevent cancer and CHD in many predisposed people. The scientific promise, no less than the sociomedical one, suggested to me that an attempt should be made to see what could be rescued from the wreck.

Criticisms might justifiably be made of the data collection. The interviewers, keen young students working for a charismatic leader and imbued with his preconceptions, might easily and unintentionally shift inventory answers and instrument readings in a direction they thought in line with Grossarth-Maticek's theories.

The large volume of data, collected on many different samples, could easily lead to mistakes, particularly when handled by hand. Differential analyses, to test a great variety of theories, could result in confusion where so many probands were concerned. Confusion was made more confounded by the harsh rules of the *Datenschutz*, the German law concerning the protection of research data; I have a lot more to say about this particularly in connection with van der Ploeg's criticism, but let me say here only that the law insists that names of participants, diagnoses, and test results must never be brought together; names must be coded first, then correlated with other data indirectly through code numbers. This complex and indirect manipulation seems to have been intended primarily to discourage research; it certainly makes for transcription errors that are difficult to avoid. Errors such as these occur in every large-scale epidemiological study; they are more likely when the investigator does not have a large university department to help him, but is thrown back on his own resources. I have myself argued that terrible things happened in relation to the large epidemiological liter-

ature on the effects of smoking (Eysenck, 1991); there is no question that they could have happened here. Was it likely that all the results were due to such extraneous causes?

There were reasons to doubt. Grossarth-Maticek had collaborated with dozens of well-known scientists, both in Yugoslavia and in Germany, who would certainly have detected any data manipulation. Some of his colleagues actually worked with him on the collection of data; it would have required a degree of paranoia to imagine that all these noted scientists were in league to produce false data! In addition Grossarth-Maticek had been careful to deposit the names and code numbers of his subjects with independent university departments in Karlsruhe and Zurich in 1982, and previous to that with the mayor of Heidelberg in 1977; as I mention later, agreement between the earlier and later depositions was independently checked.

It was also fortunate that Dr. W. D. Heller, of the Institute for Statistics of Karlsruhe, agreed to check the death certificates that are the backbone of Grossarth-Maticek's work, to check code numbers, names, and data in every case, and to act as repository for the data entrusted to him. Dr. Heller is a completely independent observer who spent a great deal of time on this responsible and unrewarding task; he is now working part time for the scientific secretariat of the Forschungsrat Rauchen und Gesundheit in Bonn; he has both in person and in writing verified all that I am saying concerning the verification of Grossarth-Maticek's data. Dr. Heller also interviewed a random sample of the students who had collected the data for Grossarth-Maticek, to look at methods used, reliability of testimony, and adequacy of training; he found nothing to complain about.

I argued that if the results of the 1972–1982 prospective study were genuine, then a further follow-up covering the next 4 or 5 years should give conclusive evidence of the lack of data manipulation. All the names, codes, and data had been given to Dr. Heller; cancer and CHD mortality after that date could hardly be affected by manipulation of data safely deposited with him and others. I was at the time participating in a series of symposia on the origins of cancer and CHD, organized for Reynolds Tobacco Company by Professor Spielberger, and I used one such occasion to give a talk on Grossarth-Maticek's data and suggested that Reynolds might like to finance a reanalysis of the original data, and a new follow-up from 1982 to 1986. They agreed to do this and appointed Professor Spielberger, Dr. Heller, Dr. van der Ploeg, and myself to undertake these tasks.

The people concerned with this extension of the original study agreed that the result would be crucial; if the new follow-up gave results similar to those of the first 10-year follow-up, then clearly the possibility of data substitution or manipulation would be minimal, or completely absent.

Because of the importance of these 1982–1986 follow-up data, I have asked Dr. Vetter to write up the results of his analyses insofar as these had been completed, and I have translated his report, which appears as Appendix A in this response. The main results can be briefly summarized:

1. Mortality remained at roughly the same level, very significantly higher for the stressed than for the unstressed, normal sample.
2. Predictors such as the scales measuring hopelessness, or anger, remained predictive at a high level of significance.

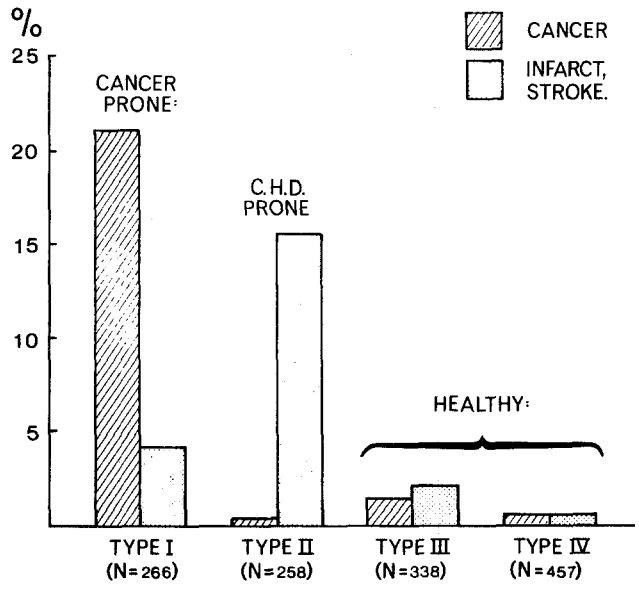

Figure 1. Mortality of different personality types from cancer and CHD, 1982 to 1986.

3. Type scales also remained at a high level of significance; Figure 1 depicts these results.
4. There is again significant interaction between smoking and personality for cancer, indicative of synergy.
5. There is again significant interaction between smoking and personality for CHD, indicative of synergy.

On all these vital points, the second follow-up thus agreed totally and in every detail with the results of the first follow-up concerning which doubts had been raised, doubts that hardly apply to the second follow-up. The results thus justify new confidence in the value of the work described in the target article.

Vetter suggested a further analysis, which to him seemed even more convincing. He suggested taking only cases where the onset of the disease (cancer or CHD) occurred after 1982; in this way any possibility, however unlikely, that the interviewer might have known which subjects were suffering from cancer or CHD, but had not died yet, and could have altered their scores to agree with Grossarth-Maticek's theory, could be avoided. (The whole idea is unlikely because of the practical difficulty of altering figures in the 10-year-old data sheets; anything of the kind would have been only too obvious.) This additional test was carried out by Dr. Vetter, as pointed out by Grossarth-Maticek in his contribution; his report of April 19, 1990, stresses that "the result suggests in the strongest possible way that predictors published in 1982 gave predictions in those who were diagnosed after 1982 no less valid than in those diagnosed before 1982." One would imagine that such a very clear-cut result would establish the correctness of Grossarth-Maticek's demonstration once and for all.

As Grossarth-Maticek points out in his own contribution, he is bitterly disappointed that, although it was universally agreed that the success of the 1982–1986 follow-up would end the criticisms of his work, the critics seem to have paid very little attention to the successful outcome, but to have concentrated instead on minor and comparatively less important problems that in their very nature could not have influ-

enced the successful outcome of the trial. I deal with these criticisms, often concerning earlier rather than later articles, in the next part of my reply. I then, in the succeeding section, indicate what conclusions are possible from all this work.

Before doing so, let me say one final word about the apparent failure of tables of results reporting the same research to agree. Consider my Table 12, which was taken from the Eysenck and Grossarth-Maticek (1991) article. The results do not agree with those of a table given elsewhere (Eysenck, 1991), and neither table agrees with the values given by Vetter in Appendix A. How is this possible? The data in the Eysenck and Grossarth-Maticek (1991) article were based on results ascertained at the end of 1985, with cause of death identified by Grossarth-Maticek himself on the basis of death certificates. The table given by Eysenck (1991) is based on data collected in 1986, and the cause of death was established independently by Dr. Heller on the basis of ICD rules of classification; he allocated fewer deaths to "other causes" than Grossarth-Maticek had done in order always to take the most conservative line. Thus the tables refer to (a) temporally different mortality ascertainments and (b) different diagnostic interpretations of death certificates. There is no contradiction.

But how can the Eysenck (1991) and the Vetter (Appendix A) data differ? A few probands had refused to allow their names to be entered on the list of participants, and the question arose whether they should be included or excluded in the analysis. I followed one rule, Vetter another, and thus we have a (slight) apparent contradiction. This can throw no doubt on the actual data, which were checked by Dr. Heller, and in any case makes no difference to the outcome. Readers may like to check my Table 12 against the values given by Vetter in Appendix A; they will see that the later values, coded by Dr. Vetter, are rather better than Grossarth-Maticek's original values, interpreted quite conservatively.

This case is typical of many others. We often have disagreements because:

1. Death certificates require coding, which has an element of subjectivity. Grossarth-Maticek's original coding has at times been changed by Heller, so that there are subtle differences in tables; the Heller coding is always the one finally adopted, as being independent.
2. Research is ongoing, and updating may slightly alter previous values.
3. Occasionally, slight errors are discovered and corrected in later publications; one or two examples are given in the next section.

All this may lead to apparent disagreements in listings from apparently identical data, but usually there are perfectly good explanations. Seltzer (1989) listed results from the famous Framingham study of smoking and CHD, which show how, for instance, risk ratios may change from 1.8 to 0.8 depending on duration of follow-up, age group considered, sex of probands, and so on. Critics of Grossarth-Maticek rely too much on minor differences of this kind as sources of doubt, rather than being concerned with the major results, and the steps taken to ensure objectivity and validity.

Critics still rightly ask for independent replications. One recent study (Schmitz, in press) is rather reassuring concerning the reliability of Grossarth-Maticek's results. Using the six-type assessment schedule, he found high and predicted correlations with the major Eysenck personality dimensions in two separate samples. Medical diagnoses were recorded for a 5-year period for the older sample, and it was found that, of six subjects with the notation "cancer," all had been allocated to Type 1. Of six subjects with the notation "cardiac infarction," four had been allocated to Type 2. The probability of this happening by chance is well below 1 in 1 million. Other diagnoses also agreed well with the earlier findings of Grossarth-Maticek and Eysenck. Correlations for all disorders, physical and mental, were negative for Type 4 (the healthy, autonomous type) and positive for all others. One swallow does not make a spring, but the study seems well-conducted and analyzed and certainly replicates certain fundamental features of our earlier work. (For another replication study, see Dixon & Dixon, 1991.)

Replies to Individual Commentators

Let me now turn to the replies to individual commentators, saving arguments contained in other parts of this reply, which apply to more than one commentator. In some cases, I have quoted replies by Professor Grossarth-Maticek. These were taken from his original contribution, which now consists of more general comments only; it seemed that this method would make it easier for readers to follow the argument.

Schwarzer

Concerning Dr. Schwarzer's suggestions, I am very much in agreement.

1. The definition of stress clearly involves a problem. What we have really attempted to do has been to identify strain—that is, the result of an interaction between an external stressor and the reaction of the individual. Perhaps this might be labeled *personality/stress,* or simply *strain,* but for the sake of brevity and clarity we have for certain purposes used as a measure of "stress" a person's belonging to Type 1 or Type 2. As long as this process of operational definition is clearly understood, little harm is done, although the purist might readily object.

2. Would a normative system of measurement be better than an ipsative one—traits rather than types? There are advantages and disadvantages to both. Grossarth-Maticek has indeed used both; Figure 2 shows the results of using a seven-trait questionnaire on his Yugoslav sample, with standardized regression coefficients indicating the relative contributions of the variables toward the prediction of cancer incidence (Eysenck, 1988). Also, the scores on the "Type" factors can be treated as normative scores. On the whole, I agree with Dr. Schwarzer that normative measurement is superior, but "types" are more easily understood by the medical profession and by laypersons.

3. The model suggested by Schwarzer is very much in line with my own thinking (Eysenck, 1991) and does incorporate most of the essential elements. To be complete, it would have to incorporate other stressors (smoking, genetic predisposition) that interact synergistically with the stress and personality factors already incorporated; on this point I am sure Dr. Schwarzer would agree. At the moment the

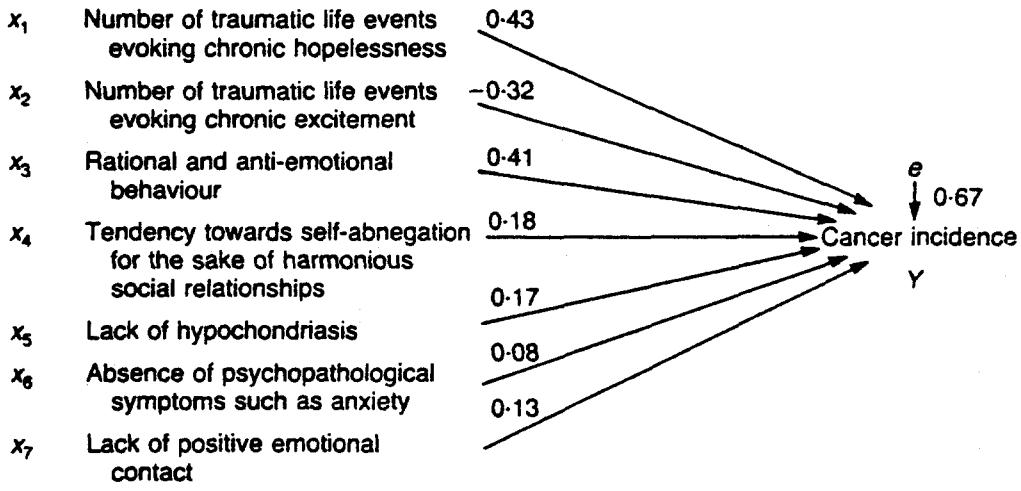

Figure 2. Seven personality inventories as predictors of cancer, using standardized regression coefficients. From "The Respective Importance of Personality, Cigarette Smoking and Interaction Effects for the Genesis of Cancer and Coronary Heart Disease" by H. J. Eysenck, 1988, *Personality and Individual Differences, 9,* p. 458. Copyright 1988 by H. J. Eysenck. Reprinted by permission.

details are still "fuzzy"—inevitably so—but it does suggest the major lines along which research should proceed.

Suinn

Dr. Suinn queries the changing nature of the questionnaires used. There are two points to be made. First, different types of inventory have been used at different times—some normative, others ipsative; the attempt was made to see which form was the most successful, which most acceptable to subjects, and which easiest to administer. Reliabilities, when calculated, were usually acceptable; indeed, they would have to be to obtain valid discrimination. Second, interest in the early work was exclusively in the testing of the original hypotheses relating personality/stress to cancer and CHD; psychometric concerns were relegated to such time as when validity had been established. In recent years, Professor Spielberger has taken over the original seven scales of the normative (trait) questionnaire and has carried out detailed psychometric studies, rewriting some of the items to suit American conditions. He reports (personal communication) that the outcome was surprisingly successful. This was one of the studies commissioned by Reynolds to test Grossarth-Maticek's claims, and the outcome certainly supports his views.

Suinn makes the valid point that other factors, apart from personality/stress, should always be taken into account; he mentions the Yugoslav sample, which included a "high psychological stress" group as well as the major "oldest inhabitant" group. Spielberger analyzed the data separately for the two groups and found that the "oldest inhabitant" group actually showed higher predictability than the "stress" group; this suggests that their inclusion did not serve to improve prediction through the introduction of external variables such as high cholesterol level, but rather lowered it.

A few questions require specific answers. Chi-square values comparing "expected" with "observed" frequencies did not include the variables mentioned, but "expected" data simply refers to distributions that would have occurred had the active variable been inoperative. Correlations as reported are linear, even though Schmidt criticized this assumption. The typology approach derives from a theoretical conception that also led to the trait approach; both reinforce each other.

Regarding the intervention studies, the nature of the impact is difficult to specify; any number of behavioral changes may theoretically act as mediators. Exercise may certainly be one, because it did show a significant increase; cessation of smoking was not involved, because incidence of smoking did not change significantly. But this whole question must remain open.

Suinn recommends profile analysis, rather than simple grading by highest type score. This is precisely what I hope to do when the 1973 follow-up data are finally collected. The method is clearly superior and may lead to better predictions and classification, as well as getting away from the ipsative model, which is probably inferior psychometrically to a normative one.

Concerning our report on the effects of psychoanalysis, Suinn asks very relevant questions many of which do not at present have an answer. Our data demonstrate that psychoanalysis constitutes a risk factor for cancer and CHD; whether the relationship is causal is a matter for further research. The causal implications of smoking for cancer and CHD are often assumed, although there is little direct evidence and many contradictory findings (Eysenck, 1991); I do not intend to make any such assumptions regarding psychoanalysis. The data indicate a correlation; whether this implies causation cannot at present be demonstrated.

Determination of "type" concepts was done largely on the basis of theoretical considerations, themselves based on experience with cases of cancer and CHD, and on information based on the literature. I agree that it would have been extremely useful if more attention had been paid to psychometric concerns in earlier work, and I have pressed for this to be done in more recent work. What I think is worth knowing, however, is the simple fact that the "type" scores seem to give reasonably accurate predictions of cancer and CHD over 14 years; improving the scales psychometrically, as

Spielberger is doing, is likely to improve their validity. The Type A–Type B classification also lacked proper psychometric foundations; medical workers are simply not used to the need for such a discipline.

I also agree with Suinn concerning the unreliability of death certificates; it seems likely that a number of "other causes of death" might very well be due to cancer or CHD. I also acknowledge the importance of detection bias, but this is more likely to operate in relation to smoking as a risk factor than with respect to personality and stress; few physicians, for instance, would be likely to be swayed in their diagnoses by thoughts of the patient's personality, nor would they be competent to assess it (Eysenck, 1991).

Suinn thus raises many very relevant questions, the majority of which do not at present find an answer. This failure is of course not infrequent in epidemiology; I have drawn attention to a dearth of answers to important questions in the field of cigarette smoking and its effects on health, specifically cancer and CHD (Eysenck, 1991). But whereas research on smoking has been carried on for a long time, by literally thousands of people, well funded by government and research agencies, the work of Grossarth-Maticek has been carried out essentially by a very small group of people, underfunded, and in a very novel field.

Hence, this dearth of answers is perhaps understandable, and even excusable, where a similar lack, combined with confident assertions, is not so in the case of smoking. What Suinn has done together with Fox and one or two other critics, has been to define an agenda for research to put Grossarth-Maticek's findings on a more secure footing and to supply answers to many urgent questions, particularly those affecting causality and mode of transmission.

Wills

Dr. Wills raises some important questions, and comes to what I believe is the correct evaluation of Grossarth-Maticek's work. The problem she raises is that of measurement. Clearly, the questionnaires designed by Grossarth-Maticek in the early 1960s can be improved by using modern psychometric methods, and Professor Spielberger among others is doing just that. The problem with prospective studies, after all, is that with hindsight, and making use of the knowledge accumulated in recent years, a much better study could be designed. In our follow-up, however, we are obliged to stick with the original concepts and methods of measurement. This poses a problem for replication; should we use the original methods and instruments, or should we change them, hopefully, for the better? If we do the former, we are in danger of repeating past errors; if we do the latter, we are in danger of abandoning a successful formula for what may be a less successful one.

The problem is confounded by difficulties of translation. Grossarth-Maticek's original inventories were difficult and intricate in their wording; administration by interviewers was almost mandatory. I found translation extremely difficult; should one go for accuracy, at the risk of sounding non-idiomatic, or should one rephrase completely to create an English version, at the risk of misrepresentation? I have chosen the first alternative, but for actual test use, rephrasing will be necessary. (This is an ever-present problem in translation, doubly difficult when personality inventories are concerned.)

Dr. Wills is obviously correct in saying that only independent replication will answer questions such as the importance of the therapist's personality. Some tentative results seem to suggest that students of Grossarth-Maticek are effective but less so than Grossarth-Maticek, their effectiveness being correlated with the duration of their training. The study is not well enough controlled to be publishable, but the results are probably reasonable; we may expect that replication by others will be possible, but perhaps at a lower level of success. Spiegel, Bloom, Kramer, and Gottheil (1989) demonstrated that even better success than that claimed by Grossarth-Maticek can be achieved with quite different methods, so obviously Grossarth-Maticek's charismatic personality is not a vital ingredient. However that may be, replication is vital, and urgently required to put the issue beyond doubt.

Grunberg and Singer

The reply by Drs. Grunberg and Singer raises many important points. Their first point raises the question of definition of personality. It is quite usual to define this in terms of habitual behavioral and cognitive responses to situations (traits), and their intercorrelations (Eysenck & Eysenck, 1985); there is nothing loose about such definitions. Other definitions are possible, of course, but that is the one adopted because it gives rise to testable consequences.

Their second point, relating to the questionable nature of blood pressure and cholesterol as risk factors for CHD, is well taken; I have discussed it at some length elsewhere (Eysenck, 1991). However, when the research was originally undertaken these were almost universally regarded as risk factors, and even now there is considerable argument; as in all prospective studies, the original variables were chosen on the basis of knowledge then available.

Their third point essentially asks some questions concerning the questionnaires used, allocation of subjects, and multiple diagnoses. Subjects are allocated to a given type if they obtain their highest score on that factor. Better prognosis might be obtained if all scores were taken into account; we will attempt to do this when all subjects of our 1973 study have had their mortality and cause of death ascertained. The tests seem to work equally well in German and Serbo-Croat; we do not know about English, although Professor Spielberger has had suggestive proof of their adequacy when transformed into an American version (private communication). Multiple diagnoses were certainly common; the problems raised thereby will be dealt with when the complete 1973 sample has been looked at for mortality—the data given are provisional, and based on small subsamples.

Regarding their fourth point, I meant by risk factors being benign not that they were completely harmless, but rather that each by itself seemed to contribute little to mortality (see Table 11 of my target article, in which each risk factor by itself was not associated with lung cancer). The expression I used can be misleading; it should read that risk factors in violation are relatively harmless. In regard to the evidence on quitting smoking, I have summarized this at great length elsewhere (Eysenck, 1991), and I repeat that there is a "relative failure of quitting smoking to reduce cancer and CHD mortality." The studies mentioned by Grunberg and Singer misrepresent the situation.

On their fifth point, there is an important difference between psychoanalysis and behavior therapy (Eysenck & Mar-

tin, 1987). The evidence shows that there is a dose–response relationship between duration of psychoanalytic treatment and mortality from cancer and CHD (Grossarth-Maticek & Eysenck, 1990d). What is important to note is that, when patients were asked if the analyst tried to increase or decrease their autonomy, greater mortality was only observed for those in the latter group; this is in good agreement with our major hypothesis that autonomy is the essence of a healthy personality (Type 4).

On their sixth point, I agree with Grunberg and Singer that the precise way in which psychological treatment reduces risk of cancer and CHD is unknown; I have suggested a direct influence on the immune system and the factors making for sclerosis, and have found good evidence for this supposition, both in the literature and in our own work (Eysenck, 1991; Grossarth-Maticek, Eysenck, Gallasch, Vetter, & Frentzel-Beyme, 1991). When checked, we found that subjects did not reduce their smoking but did increase their physical exertion (walking, jogging, sport). This is one of the many questions urgently needing research to provide proper answers. We have found that treated subjects show a marked decrease in Type 1 and Type 2 scores, and a similar increase in Type 4 scores; this may give us some insight into the causal process.

Does psychological treatment significantly increase survival? Our data certainly suggest that it does, and the work of Spiegel et al. (1989), using quite different methods, supports such a view. There were unavoidable difficulties of an ethical nature in some of our studies; you cannot randomly assign terminally ill women suffering from mammary cancer to having or not having chemotherapy. I doubt if this seriously affects the issue; after all, the comparison between psychotherapy and no psychotherapy was made within chemotherapy groups.

Finally, considering the last point raised, I would like to disagree with Grunberg and Singer when they say that my general theory "lacks convincing support." I have cited several dozen studies elsewhere that support quite strongly the various aspects of my theory (Eysenck, 1991); I think it is not correct to say that it "is not substantively based enough to merit serious debate at this time."

Schwartz

Dr. Schwartz is right in asking whether our data on interaction between risk factors are capable of being generalized. There are two reasons for believing that the answer is in the affirmative.

1. Many years ago Kissen and I considered the problem of interaction between smoking and personality, as far as lung cancer is concerned (Kissen, 1964; Kissen & Eysenck, 1962). We found that "lung cancer mortality rates of those with a poor outlet for emotional discharge may be five times greater than those with a good outlet" (Kissen, 1964, p. 215). The poorer the outlet for emotional discharge, the less cigarette smoke required to induce lung cancer. This relation appears synergistic when looked at from that point of view.

2. Grossarth-Maticek and I have recently completed a replication study of our earlier works (Eysenck, 1991), using 1,914 subjects from our 1973 Heidelberg study. The results are shown in Table 1 for lung cancer and Table 2 for CHD. "Stress" is defined as belonging to Type 1 in the former and as belonging to Type 2 in the latter. It will be seen that there is a strong synergistic effect in both cases.

Table 1. *Lung Cancer Mortality as a Function of Smoking and Stress: Replication With Heidelberg 1973 Data*

	No Stress	Stress
No Smoking	0.69%	2.09%
Smoking	0.24%	10.59%
Smoking Effect	0.24% − 0.69% =	−0.45%
Stress Effect	2.09% − 0.69% =	1.40%
Real Combined Effect	10.59% − 0.69% =	9.90%
Additive Effect	1.40% − 0.45% =	0.95%
Difference (Synergistic Effect)	=	8.95%

Note: N = 1,914.

Obviously, replication of all this work, and our additional study on cancer of the mouth and pharynx (Grossarth-Maticek & Eysenck, 1990b), is necessary to establish the truth of our hypothesis, but at the moment it does look reasonably healthy.

Derogatis

Dr. Derogatis raises a number of issues and caveats most of which I agree with; we may perhaps disagree on the degree to which they might affect my conclusions. There is no doubt that death certificates are very unreliable assessments of cause of death, although they may perhaps be granted some validity as evidence of death! I have reviewed in detail the studies which have been done in this field, and without doubt reliability and validity (comparison with autopsy results) are low. It is curious that almost the whole literature linking cigarette smoking with cancer and CHD is based on death certificates, without attracting the criticism that should follow (Eysenck, 1991). How does the fact of such unreliability affect my conclusions?

It is claimed that there is a positive correlation between belonging to Type 1 and later death from cancer, between belonging to Type 2 and later death from CHD, and between belonging to Type 3 or 4 and survival. If cause of death is determined with low reliability, then the observed correlations

Table 2. *CHD Mortality as a Function of Smoking and Stress: Replication With Heidelberg 1973 Data*

	No Stress	Stress
No Smoking	1.10%	5.30%
Smoking	3.04%	17.50%
Smoking Effect	3.04% − 1.10% =	1.94%
Stress Effect	5.30% − 1.10% =	4.20%
Real Combined Effect	17.50% − 1.10% =	16.40%
Additive Effect	4.20% − 1.94% =	6.14%
Difference (Synergistic Effect)	=	10.36%

Note: N = 1,914.

will be attenuated; in other words, they would be even greater if corrected for attenuation. The differential mortality of Types 1 and 2 vis-à-vis Types 3 and 4 is not affected; mortality as such is reliably reported. It would of course be true to say that each death should have been investigated in detail, rather than relying on death certificates. The huge number of cases involved, the lack of financial support, and the medical reluctance to discuss individual case histories or carry out autopsies, made anything of the kind impossible. Even traditional enquiries amply funded by the state have usually had to rely on death certificates; if epidemiology had to rely on other sources of information, there would be little to report!

How about susceptibility bias? Derogatis argues that the elevated age of the Yugoslav group would spuriously enhance the relationship observed. It is difficult to understand this criticism. The incidence overall of cancer and CHD is of course higher in this older group than in the 10-years-younger Heidelberg normal group, but this should not affect the correlation within each group between type and mortality, or type and diagnosis. On this point I cannot agree with Derogatis.

Detection bias is another problem which I have discussed in detail in relation to smoking-related diseases (Eysenck, 1991). It is meaningful to question correlations between smoking and death from lung cancer when the diagnosis is itself in part based on the fact that the patient was a smoker, and the medical establishment firmly assumed the relationship it was trying to prove. But this is not so in relation to personality; few physicians would have signed "lung cancer" on the death certificate because they suspected that the patient had been Type 1! The argument might apply to our data for "incidence" but not for mortality, and it is the latter on which Derogatis would lay most stress.

Derogatis is correct in criticizing the various conceptions of "stress" in the work described; certainly the identity of the concept according to which the Heidelberg stressed sample was recruited with the various definitions in terms of belonging to Type 1 or 2, or having a score for Types $1 + 2 + 5 > 3 + 4 + 6$, cannot be assumed. It is a fact that there were more individuals of Type 1 or 2 in the Heidelberg stressed than in the Heidelberg normal sample, and fewer individuals of Type 3 or 4; to that extent there is congruity. Other than that, all one can say with hindsight is that if the studies were repeated, better and more explicit definitions and measures might be found. Yet it is important to note that very significant differences in mortality and cause of death were found in all these groups, compared with the underlying theory; this suggests that the criticisms raised by Derogatis are not insuperable.

I cannot agree with what Derogatis has to say about personality types. I have dealt with the point of comparing normative with ipsative measurement in my reply to Schwarzer, and will not repeat my argument; note simply that Figure 2 shows the traits that have gone into the Type 1 concept and that have received positive replication by Quander-Blaznik (1991) and others (see Eysenck, 1991). Whether to use the multiple trait or the type approach should not (and does not) make much difference. When analyzing our very extensive 1973 data, we hope to use normative multivariate methods rather than type scores, and we hope to compare the efficacy of both methods. This is clearly advisable in view of Derogatis's statements concerning assignment algorithms. Two people having identical Type 1 scores may have very different scores on the other types, and this may crucially affect their prospects as far as contracting cancer and CHD is concerned. But let us not forget that any improvement in the validity of the assignment procedures is likely to be mirrored in the efficacy of the prediction process. Even if suboptimal, the method has been found to give predictions significantly better than chance. This is an important achievement, and indeed Derogatis does not fail to recognize it as such.

Levy

Dr. Levy makes several points that in one sense are completely justified, whereas in another they seem unjust. Thus she criticized "the author's style of review and research presentation." What I have tried to do, in a relatively short article, has been to introduce the reader to a very large body of data, give some examples of the kind of results found and the conclusions I believe can be drawn from them. In no sense was this a scholarly review of all the evidence; a whole book would not be sufficient to do that. Nor was this an attempt to present sufficient detail in connection with each demonstration to allow the reader to follow all the details. Finally, I did not attempt to anticipate and answer all the questions that might arise, justifiably, in the reader's mind. I intended the article to act as a guide to a large body of literature that, I believed, made a significant contribution to our understanding of the body–mind relationship and also had some very important social consequences with respect to preventive medicine. It was hoped that interested readers would turn to the more detailed accounts in the literature.

Now let me turn to the particular points raised by Levy. Her first complaint is that on at least one occasion independent study samples were directly compared on outcome data when the populations themselves were not comparable. She refers to my Table 4, which

> compares lung cancer and other deaths for smokers and nonsmokers from samples drawn from both Yugoslavia and Germany, but the populations themselves were demographically distinct. . . . What effect such demographic, as well as cultural and sampling, differences had on findings is both unclear and unexamined.

The purpose of the table was to illustrate a point on which both samples agreed, namely, that smoking and typology collaborate synergistically to produce lung cancer mortality. The fact that demographic differences exist only makes the demonstration more impressive. In any case, a detailed description of all the data for these and other samples is given in the original publication; it would not have been useful to repeat it in the framework of this presentation.

Dr. Levy also criticizes the description of the nature of the intervention (autonomy training) as showing "lack of methodological rigor," apparently because "not only is the method unspecified in terms of detail . . . but apparently, the precise method differed substantially from group to group and from individual to individual." True; but that is inevitably the nature of psychological therapy. Different individuals have different problems, different social support, different coping abilities and mechanisms; they differ in their responses to particular methods of intervention, and of course in the reactions of their relatives, friends, and employers to their changes in behavior consequent on therapy. It is unre-

alistic to imagine that complex therapeutic interventions can be more precisely characterized than was done by Grossarth-Maticek and Eysenck (1991); the area is inevitably "fuzzy," and to pretend otherwise would give quite the wrong impression.

Levy's criticism of my discussion of synergistic effects in Table 4 is not, I think, valid; compared with individual effects of smoking or stress, the interaction in its synergistic form is greatly superior. I have discussed this in connection with my comments on Dr. Schwartz's critique and will not go into further detail. (See also my reply to Cooper & Faragher.)

Dr. Levy also complains of "erroneous, rather than problematic, interpretations of findings." In Table 1, I have stated that Type 1 probands die "mainly from cancer," and Type 2 probands die mainly from heart disease. She points out that Type 2 individuals die even more frequently from "other causes" than from CHD. But my point was the comparison of cancer and CHD mortality, for which the column "other causes" is irrelevant. In any case, all the relevant data were given for the reader to come to a conclusion. She also suggests that I should have commented on the fact that in Table 11, 31% of those with four risk factors for lung cancer also died from other causes. To have commented on all the interesting findings in our studies would have tripled the length of the article. But the explanation is not difficult, the factors of smoking and stress are also relevant to cancer generally, and in part to CHD, which would make up the majority of deaths from other causes.

Concerning the effects of psychoanalysis, and its effect of increasing the probability of contracting cancer or CHD, our data as shown in Table 15 seem pretty conclusive. Levy refers to Table 14 as showing that

> those who were randomized to a form of psychoanalytic intervention did no worse than controls (in fact, a slightly greater proportion of those receiving psychoanalysis were alive on follow-up compared to the control group: 19% vs. 15.8%). Yet, Professor Eysenck boldly concluded that "psychoanalysis can *increase* significantly the likelihood of cancer and CHD mortality."

But as Table 14 makes clear, we did not administer "psychoanalysis" to the 100 control subjects in question; they were included as a placebo group and received a psychoanalytically slanted text to read, with 4 hr of discussion on how to use this in connection with their own problems and stresses. This is hardly "psychoanalytic treatment" of a serious kind, and no effects one way or another would have been expected, and none were found; the differences between no treatment and placebo treatment were insignificant. We introduced the placebo to control for the simple fact that, compared with the no-treatment controls, the presence of some form of treatment by itself, regardless of content, might have produced an effect. It did not.

In connection with the outcome of our therapy research, Levy states that I "[bolster] the believability of the data by citing a similar outcome shown in a recent report by Spiegel et al. (1989), but the latter demonstrated more modest, but statistically robust findings, rigorously reviewed by peer examination." This is an extraordinary statement. Spiegel et al.'s results are not "more modest," but even more impressive than ours, prolonging life in the treatment group by more than doubling survival duration, whereas in our two studies the improvement was less than double. Our findings too were "statistically robust," and they appeared in a journal where they received equally rigorous peer review. Levy's insinuations are groundless, and counter to easily ascertainable fact.

Kiecolt-Glaser and Chee

The commentary by Drs. Kiecolt-Glaser and Chee raises several questions that certainly deserve answers but for which the evidence is largely lacking. They are correct in saying that the Grossarth-Maticek and Eysenck (1990c) article "provided neither a theoretical rationale for the six personality types and their relation to disease-proneness nor sufficient information regarding the psychometric properties of the inventory to enable evaluation of its reliability and validity." Such information concerning the theoretical rationale was provided in previous articles; it would not have been appropriate to repeat what had been said many times before. Regarding reliability and validity, it was expressly stated that "the data reported in this article are . . . only preliminary" (p. 362); it was thought that the new inventory, and the new method of administration, would be of interest to people working in this field and would repay replication. Many requests for reprints and further information, and statements of intent to use the measure for experimental purposes, suggest that this purpose was fulfilled. But surely the data there published are evidence of validity, and does validity not imply reliability? We also stated that "test–retest correlations are all in excess of 0.80, and so are quite satisfactory" (p. 358). We also give factor analyses of the intercorrelations between the scales. It is not quite clear what more the critics would want. Kiecolt-Glaser and Chee are right in saying that "studies examining the convergent and discriminant validity of the Personality–Stress Inventory with other measures, including measures of depression, psychopathic personality, and proneness to CHD and drug addiction, are needed"; does this amount to more than the time-honored statement at the end of most research articles, that more research is needed? No research would ever be published if all its ramifications had to be explored beforehand. Our article replicated earlier correlations between personality and disease, and improved prediction by a novel method of administration; it did not set out to solve all problems of personality measurement in this field.

The commentators are right in saying that "it is also unclear from the information presented . . . whether the Personality–Stress Inventory is assessing stable personality traits as opposed to the ability to cope with stress that may be influenced by situational circumstances." This is an urgent issue to be settled by research, but it is not the kind of research we have concentrated on. One cannot solve all the problems of stress-related research in one go. We have reported on the prediction of cancer and CHD by means of specially designed personality questionnaires that were created to test certain quite specific hypotheses; these predictions were surprisingly successful, following in the tradition of earlier work (Kissen & Eysenck, 1962; Schmale & Iker, 1971). It is obvious that more questions are raised by their success than are answered; that is not unusual in science, but hardly represents a target for criticism. In any case, there is no either–or answer; all personality test scores are partly

determined by genetic factors, partly by environmental ones (Eaves, Eysenck, & Martin, 1989); they are largely stable but may be changed in specific ways (Eysenck & Eysenck, 1985). We are at the moment analyzing twin data from our studies to answer some of these questions, but a comprehensive answer is still unlikely to be found in the near future.

Kiecolt-Glaser and Chee may be right in doubting the importance of immune-function regulation for cancer, and in doubting the wisdom of concentrating on cancer in general, rather than on specific cancers, and on primary tumors, rather than metastases. In planning research, one must follow one's theories and the existing research findings; we may be entirely wrong in our choices but that is our privilege. Whatever the fate of these particular theories to explain our findings, the findings themselves are what matters at the moment, and these certainly call for an explanation. It is quite likely that alternative theories will be put forward and will have better explanatory value; that we must leave to the future. Possibly "the costs are not justified by the paucity of literature relating immune function to primary tumors"; if we follow that line of argument nothing new would ever be found! We will certainly do what we can, helped by immunologists less certain of the established wisdom than Kiecolt-Glaser and Chee, to put the matter to the test. After all, as T. H. Huxley said: "It is the customary fate of new truths to begin as heresies and to end as superstitions."

Cooper and Faragher

Drs. Cooper and Faragher make two main points. The first is that, of two possible ways of analyzing our data, we have adopted the wrong one. As they say:

> Conventional epidemiological wisdom dictates that study cohorts be grouped by outcome for statistical analyses . . . Professor Eysenck prefers to fly in the face of this convention and to subgroup his study cohorts with respect to preconceived notions of important personality types. By so doing, he runs the considerable risk of overlooking important information produced in his studies.

Well, one man's "preconceived notions" are another man's carefully contrived theories. The theories embodied in our "types" or the original traits (see Figure 2, in my replies to Schwarzer) are derived from longstanding tradition and earlier studies (Kissen & Eysenck, 1962; Schmale & Iker, 1971), to name but a few. Our method followed the universally adopted method of the hypothetico-deductive model; I do not feel that it requires justification.

However, it is certainly possible that the "conventional method" has certain advantages. It is based on the absence of any specific theory, and capitalizes on chance errors, but it may suggest correlations that had not occurred to the investigators. We are planning to carry out such an analysis on our 1973 sample when data collection is complete, but only in addition to a major analysis along the lines of our previous work. As Lewin used to say: "There is nothing as practical as a good theory," and our work seems to have shown that the theory, ancient as it may be, is indeed a good one.

Cooper and Faragher's other criticism refers to the problem of synergistic interaction. The data in the target article, and further data presented in our reply to Schwartz, require a special discussion regarding their statistical analysis here. Although these data are of a kind to show even by visual inspection that there is a strong interaction, and although analysis using an additive model clearly suggests such a view, this is not the only available model (Cox, 1970; Darrock, 1974; Galtung, 1967, p. 415; Grizzle, Starmer, & Koch, 1969; Plackett, 1974). There is also a multiplicative, logistic model (Everitt, 1977), and the two models may give apparently different answers, as Everitt and Smith (1979) pointed out in discussing alternative interpretations of identical data by Brown and Harris (1978) and Tenant and Bedington (1978). Briefly, the essential differences between the two models is that the additive one looks simply at differences between proportions, whereas the multiplicative models work with ratios of proportions, or relative risks. As Everitt and Smith (1979) pointed out, "it is quite possible for the 2 models to lead to seemingly conflicting results when applied to the same set of data" (p. 582). In the case of Table 4 of the target article, the logistic analysis (log-linear contingency-table analysis) gives a result in terms of independent variables not interacting with each other; in other words, there was no significant interaction.

Which model is the correct one? As Everitt and Smith (1979) stated, "unfortunately there is no absolute answer, and in practice the choice between them may depend on rather complex reasoning" (p. 582). Linear, rather than logistic analysis using a log-linear model, is perhaps more direct, and gives us a clear answer to a most practical question: Which of the four groups in Table 4 of the target article is the one we should direct our effort toward when suggesting giving up smoking and learning how to cope with stress? We have proceeded in the discussion on the basis of the linear, additive model, but readers should be warned that there exists an alternative model which may have to be considered in future discussions of this problem. (Other alternatives are the probit and the complementary log-log functions.)

It would take us too far out of the discussion to consider in detail the reasons for choosing an additive rather than a multiplicative model; the references given suggest that for data such as those of Brown and Harris (1978), or those here considered, which posit two separate and largely independent risk factors against a known background of risk enables us to postulate a natural scale on which to search for interaction in effects where the factors can be conceptualized as, say, physical (smoking) or psychological (stress) insults to the organism. The interaction term would then suggest that the effects of our "insult" would be stronger in an organism already subjected to another "insult." The available data do not prove this analysis to be correct; they merely render it likely.

Ultimately, of course, there is no real inconsistency between outcomes. As the names imply, the additive model adds separate effects and finds a huge interaction effect; the multiplicative model multiplies separate effects (on a different scale) and fails to find interaction effects because these have been incorporated into the process of multiplying effects. The main point remains that effects are synergistic, with the interaction in additive models, or with multiplied effects in multiplicative models.

I believe that our data indicate quite clearly a synergistic mode of interaction. There seems to be no question but that the smoking–stress group must be the target group for any intervention. The appropriate statistical model, and its interpretation, is clearly subject to discussion and argument; as

indicated, such discussion has taken place in several different contexts. Readers may also consult Appendix A, where Dr. Vetter analyzes the 1982–1986 follow-up data with respect to synergistic interactions.

Cooper and Faragher are on safer ground when they refer to "a confusion between stress and personality. These two terms appear to be used interchangeably at various points throughout the article—they are distinct entities and should not be confused in this way." This is an error in one way, but difficult to avoid in another. The definition of Types 1 and 2 is in terms of their different ways of dealing with stress; in other words, they are not necessarily types of universal relevance, but personality is involved only in a specific type of situation and our concern is with a specific type of stress management. How these "types" are related to more general variables such as neuroticism, extraversion, and so on is a matter for further research. What is clear is that Type 1 implies (a) the existence of great interpersonal stress and (b) a certain type of reaction—difficulties in coping, feelings of hopelessness and helplessness, denial of anxiety and other emotions. Thus we are dealing with a measure of personality-under-stress, or stress/personality, and we can use the test as either a measure of stress, or a measure of reaction to stress (i.e., personality). This constitutes the essential novelty of Grossarth-Maticek's approach; it presents obvious difficulties, of which the semantic ones mentioned by Cooper and Faragher are only the least onerous. But what is important to note is that this approach has been very successful in predicting cancer and CHD and has opened up the field for further studies to answer the many questions that arise.

A particular problem that arises relates to the obviously neurotic nature of the Type 1 and 2 reactions, combined with the inability of such individuals to express emotion readily. This combination is difficult to assess by questionnaire, and the Kissen and Eysenck (1962) expedient of using low scores on a neuroticism questionnaire as indicating cancer proneness, although successful, is obviously only a *pis-aller*. Methods introduced by Gudjonsson (1981), Kneier and Temoshok (1984), Scott and Thomson (1956, p. 507), Weinberger, Schwartz, and Davidson (1979), and others are probably superior but more expensive and time-consuming. As usual, a theory (repressed emotionality) generates different methods of measuring the concept in question; only time will tell which is the most adequate for the purpose of indexing the nomological network involved.

Lee

Dr. Lee voices a point of view that has found much expression in several contributions—namely, that the results are "too good to be true," and unlike any others published previously. I think there are two points to be made. The first is that a correct theory may make predictions that are much more accurate in practice than predictions made on the basis of less correct theories—or heuristic findings based on no theory at all. I have discussed the point elsewhere (Eysenck, 1990), taking some previous studies using theories similar to those elaborated by Grossarth-Maticek, and I find that the predictions made there are not all that different from those of Grossarth-Maticek in their effectiveness. Some rather small-scale independent replications have already been carried out (see summary in Eysenck, 1991), with some success, but Lee is of course eminently right in asking for further independent replications. We too have ceaselessly asked for such studies to be undertaken and cherish the hope that the publication of this debate will encourage others to begin such a replication.

It is, of course, not true that the material here presented "is all published in the psychological literature." I note at random that Grossarth-Maticek, alone or with others, has published related material in the *Journal of Behavior Therapy and Experimental Psychiatry, Cancer Detection and Prevention, Deutsche Zeitschrift für Onkologie, Der Kassenarzt, Neuropsychobiology, Journal of Psychosomatic Research, Psychotherapy and Psychosomatics, Social and Scientific Medicine,* and *Health Care and Human Behaviour* and in a book, *Primary Prevention of Cancer,* to name but a few. The data are presented in a manner befitting the testing of a psychological theory, although this may differ somewhat from standard epidemiological practice.

But to return to the astonishment often felt at the size of the observed effects. Consider an example. Obsessive–compulsive cleansing rituals have been found almost impossible to treat psychologically, so that the spontaneous remission rate and the treatment success rate is near zero. Yet when Eysenck and Rachman (1965) suggested a theory that would link the disorder with well-established psychological principles, and the method was implemented by Rachman and Hodgson (1980), a success rate of over 90% was reported for quite short-term treatment. Furthermore, the results have been replicated successfully and independently at least twice. Too good to be true? Expectations based on past history do not always make good prognostications.

Temoshok

Dr. Temoshok is correct in stating that a variety of inventories and questionnaires was used at different times and that analyses were carried out on different subsections at different times. This does not seem unreasonable considering the novelty of the theoretical conceptions (certainly at the time when research began; I have commented on this point extensively in connection with other commentators). Also, Grossarth-Maticek tried different ways of allocating subjective types: It could be done by simply scoring the subject's inventory and allocating him or her to the type receiving the highest score, it could be done by allowing the interviewer to make the decision on the basis of the answers and scores, or it could be done on the basis of inventories filled in by relatives. All this work was performed to find the best way of ascertaining personality reactions to stress; it does not seem unreasonable to carry out such a search and to publish results sometimes using one method, sometimes another.

Dr. Temoshok is quite wrong in thinking that I had any influence on Grossarth-Maticek's conception of a cancer-prone personality (Type 1). Although this did emerge from my early work with Kissen (Kissen & Eysenck, 1962), this study was not influential for Grossarth-Maticek's thinking, and the notion of his four (later six) types was as little influenced by me as by Temoshok's or the Morris and Greer notion of Type C.

Dr. Temoshok also seems to be critical of my decision to devote this article to the much-neglected work of Grossarth-Maticek and not to give a lengthy historical introduction detailing the contributions of the many other people who have put forward similar ideas and who have made important empirical contributions. The space available seemed better

devoted to a more detailed presentation of the major results of Grossarth-Maticek's work than to a historical introduction recapitulating things already widely known to people working in this field. I still feel that that was the right decision.

The rest of Dr. Temoshok's remarks I refer to in my final paragraphs, where I also indicate what conclusions we may draw from all the work reported. On one point, however, a few more words may be said. In a very long-term follow-up study, it occurs all the time that new data come in concerning old samples. Thus people who would not be contacted at first may be found after all, preliminary diagnoses based on death certificates may be changed by the independent assessor (death certificates are not always clear-cut and may need interpretation!), a new way of scoring data may be tried (a person who has equal scores on Type 1 and Type 2 may be assigned to an "unscorable" category, given half a score for both types or another, or assigned according to his or her other answers to one type or the other). This may create small differences from article to article, suggesting errors of one kind or another, but note that the differences are small and never alter the statistical significance of the data. Temoshok is right in calling for a more detailed discussion in each article of the precise method of scoring, allocation, and so forth used, and without a doubt this is a reasonable criticism. However, it does not alter the major conclusions drawn which do not depend on slight variations of this kind.

I append Professor Grossarth-Maticek's answers to Dr. Temoshok:

> Dr. Temoshok concentrates on two points. She complains (a) about inconsistency of measuring instruments and (b) that the typology of the Yugoslav study of 1977 was published rather belatedly. In this, she fails to note that all my publications concern partial aspects of an ongoing research program covering 28 years so far. This program contains three large prospective studies, some including intervention attempts, within which both theory, methodology (measuring instruments), and therapy have been developed and improved. As an example, the original four-type theory of 1963 developed into a six-type theory in 1973; the measuring instruments were simplified and thus made easier to use. Some instruments were used only in Yugoslavia, and not in Heidelberg, dependent on my estimation of their usefulness because I did not consider that they would be relevant in Germany; in 1976 we found that the contrary was true, as shown in our 1982 article in *Social Science and Medicine*.
>
> Similarly, the rational-antiemotional scale, used successfully in Yugoslavia, was used in Germany only on a relatively small sample; results showed that we were mistaken in not using this scale in Heidelberg on a larger scale.
>
> Nevertheless, there are several identical measuring instruments used in both countries, such as the inventories concerned with the four-type assignment, and the Ronald Grossarth-Maticek–Personality (RGM–P) questionnaire and the assessment of degree of autonomy; these have remained unchanged since their first use in 1964–1965 until 1972. For the 1973 study, these methods were extended and made more precise.
>
> Regarding theoretical conceptions and the development of the typology, it is important to note that in an original trial investigation in Yugoslavia in 1963–1964, four independent types had been isolated in comparing cancer patients, CHD patients, hysterics, and a group of elderly healthy persons. These comparisons suggested a correlation between personality-behavior and disease. Admittedly the instrument used for typological allocation was relatively complex and difficult to use, as Dr. Temoshok remarks.
>
> My primary interest in predicting cancer and CHD led to the development of a better but still complex inventory which concentrated on elements of Types 1 and 2, while retaining only some elements of Types 3 and 4 with regard to RGM–P and assessment of degree of autonomy. With respect to RGM–P there exist well-documented answers to every item in the Yugoslav and Heidelberg 1972 study, making possible multivariate analyses. With respect to typological assessments, of course these are only available as a classification into types, in accordance with the highest number of points reached for one type, without respect to points in other types. If a person had identical scores for Types 1 and 2, a further highly complicated method of analysis was undertaken which is correctly criticized by Dr. Temoshok.
>
> The methods used have been described in unpublished protocols of the Yugoslav and Heidelberg studies, but in 1975 I decided not to publish the typological inventories and to make public only the six-type classification carried out first with the 1973 samples. In 1985, the decision was made to go back to the earlier typology when the data were given to a group of research workers who, under the leadership of Professor Spielberger, were to undertake a reanalysis of the whole data complex. When it appeared that results were indeed relevant also before the 1986 mortality study, we decided to publish the results in spite of lack of psychometric information at certain points.
>
> If Dr. Temoshok declares that the typology had nothing to do with the original variables related to cancer and CHD, she disregards Table 7 (p. 491) of the article by Grossarth-Maticek, Eysenck, and Vetter (1988). This table illustrates very clearly the correlation between the variables from RGM–P and the typology. Again, Dr. Temoshok turns to the 1973 typology and asserts that Type 5 (rational-antiemotional behavior) suddenly loses all connection with cancer, because she believes that such a connection only exists for Type 1; the results clearly contradict such a view. Our first analyses show a clear interaction between Type 1 and Type 5 for the prediction of cancer, and between Type 2 and Type 5 for the prediction of CHD. This confirms the results of the Yugoslav study which showed the relevance of rational-antiemotional conduct for both cancer and CHD. Closer inspection of our six-type classification will show that there is comparatively little agreement with the Type C behavior she describes. Her belief that Eysenck had used her concept as a source of our typology is thus quite unrealistic.

Spiegel

The comments of Dr. Spiegel are discussed by Grossarth-Maticek in some detail, and I only summarize what he has to say. Spiegel relies exclusively on the inaccurate and biased account of van der Ploeg and has not responded to Grossarth-Maticek's offer to give him detailed answers to the points raised. My reaction to van der Ploeg's critique is given later. His criticisms are factually inaccurate, disregard important information in van der Ploeg's possession, and fail to take

into account important recent developments. They are so unreliable that one wonders why Dr. Spiegel accepted them without probing their veracity or inquiring from Grossarth-Maticek or myself whether the accusations these contained might not have a complete answer.

Grossarth-Maticek's answer to Spiegel is as follows:

> Spiegel relies exclusively on van der Ploeg's argumentation; I offered him the opportunity to answer all his questions regarding our scientific publications, but he never answered this offer. Regarding the therapeutic efforts by Grossarth-Maticek, it may be stated that during a 3-year period altogether 2,631 hr of preventive intervention were carried out; that is an annual duration of 877 hr, or between 2 and 3 hr daily (average 2.4 hr). I received assistance from three co-therapists, who carried out a similar amount of preventive therapy. This may make the number of 5,500 hr of therapy "delivered by Professor Grossarth-Maticek" more plausible.
>
> Spiegel's assertion that the predictive accuracy of our personality data for mortality could not be replicated is based on van der Ploeg, and reference to our answer to his critique will show that it is based on a very obvious fallacy.
>
> Spiegel mentions van der Ploeg's erroneous "findings" that numerous data-substitutions were discovered. Thus he states quite uncritically that different causes of death were attributed to the same person. In reality this refers to a single case where a similarity of names caused a mistake which could later be corrected. Every epidemiologist who works with extensive data concerning mortality knows that such minimal errors occur in every study and are not used, as by van der Ploeg, to discredit the whole study. Such errors, even if uncorrected, make no difference to the conclusions, and are usually discovered on going over the data a second and third time.

I think it would be informative for critical readers to compare the Spiegel et al. (1989) study with the original Grossarth-Maticek (1980) study of the effects of psychotherapy for one half of 24 matched pairs of terminal cancer patients, with the other matched patient randomly assigned to the control group. On issues such as matching for type of cancer, age, sex, social background, educated, and extent of the disease, the Grossarth-Maticek study is clearly superior; in other respects it is comparable. Yet it is not referred to by Spiegel et al., and for unknown reasons the Grossarth-Maticek study has been declared to give results "too good to be true," whereas the Spiegel et al. study, giving even better results, has been universally accepted. This is an odd divergence, and I gave a number of colleagues the two studies to compare on the grounds of scientific excellence, methodology, statistical analysis, and appropriateness of conclusions. They found little difference, with a slight advantage for the Grossarth-Maticek study. For me, the combination suggests two conclusions.

First, psychotherapy can prolong life in terminal cancer patients by a significant amount (i.e., close to 100%). Second, this effect can be produced by quite different approaches. We should now aim at replication with the express purpose of discovering the way in which the result is achieved—which means detailed investigation of changes in cortisol level, adrenocorticotrophic hormone, endogenous opiates, natural-killer-cell activity, and other hormonal, biochemical and also psychosocial variables. Grossarth-Maticek's method is based on sound theoretical principles, but a major part of his success may still be due to charismatic personality factors which may be difficult to reproduce. The same may be true of the entirely different Spiegel therapy. If the effects of diazepam are no greater than the effects of placebo (Shapiro, Struening, Shapiro, & Milcarek, 1983), perhaps a good many of the prophylactic and disease-related effects of psychotherapy are also due to placebo action? Even if there could be agreement on effects, causality still presents an insoluble problem at the moment.

Frentzel-Beyme

The comments by Dr. Frentzel-Beyme, who is a well-known member of the German Cancer Research Center, deserve special attention because he worked on Grossarth-Maticek's data, is intimately conversant with them, and would have been in the best position to spot any irregularities. His testimony is therefore especially welcome. Perhaps the language is rather stronger than called for, but he has suffered a good deal of provocation in his career because of his defense of Grossarth-Maticek. He answers several criticisms that I have dealt with only cursorily, and to repeat his points would have been redundant.

He brings out very clearly the emotional reaction that Grossarth-Maticek's work produced in many German epidemiologists, psychoanalysts, and physicians—a reaction so extreme that one may wonder, as he does too, what lies at the base of it. I have documented this reaction in my opening paragraph; I will not endeavor to explain it. Our opponents have often used arguments *ad hominem* in their dealings with Grossarth-Maticek (they have usually been kinder in their dealings with me); it would be foolish to respond in kind. Arguments have to be dealt with on their merits, and the motivation of critics is not part of the scientific process. If Frentzel-Beyme has at times transgressed that rule, this may be understood in terms of the unreasonableness of our opponents; here we are only concerned with his definite statement that after many years he has found no criticism to make of the method of data collection used by Grossarth-Maticek, other than small and fundamentally unimportant problems which every investigator encounters with such large numbers of subjects. Especially important is his point that minor deviations from perfect sampling (exclusions, refusals, etc.) do not violate the conclusions to be drawn from a prospective study as they cannot create the practical correlation between personality and disease.

Fox, and Shüler and Fox

Dr. Fox, with his accustomed sagacity, reanalyzes mainly results from earlier studies and points to statistical results that seem to him unlikely. His correspondence with Dr. Vetter, who carried out the analysis, is illuminating concerning the complexity of the resulting arguments. Both data collection and analysis occurred well before the time that I became involved in the Grossarth-Maticek saga, and I can but agree with his conclusion "that the prudent reader should be cautious about drawing conclusions from the reported results and about accepting unquestioningly the conclusions drawn by the authors from those results." I discuss what conclusions I think can be drawn from the data in the final part of my

reply, rather than here, so that readers are referred to these for a considerate answer.

Much the same must be said concerning Fox's other critique, with Dr. Schüler. They list a long catalogue of queries that Grossarth-Maticek should have addressed, but did not, in his early articles. Again, the criticisms are valid, and it is to be hoped that when his findings are published in book form, after the completion of the 1973 study, Grossarth-Maticek will make use of the great help given by Fox and Schüler in making clear exactly where information is lacking and what issues need to be addressed.

Schüler and Fox represent the more scientific part of the group of critics of Grossarth-Maticek's work who have remained unconvinced, and without doubt their concerns carry the most weight. The fact remains that certain data from Grossarth-Maticek's work have stood up to all criticism, and do suggest that there is some reality to his claims. This point, as already mentioned, is addressed in my final paragraphs.

It is unfortunate that Dr. Fox, in neither of his commentaries, consents to deal with the main import of my target article, but rather with some earlier and admittedly weaker work. As I have made clear, and as I thought Dr. Fox agreed when the group to supervise reanalysis and extension of the follow-up was set up, retrospective considerations are of much less importance than checking prior results against developments after the crucial deposition of all relevant data. It was agreed that if this extension of the follow-up gave data similar to, or identical with, those obtained previously, then the essential correctness of the major findings would be established. These data have now been collected, and an analysis is given in Appendix A; it is clear that the psychosocial predictor variables work as well post-1982 as they did pre-1982. This, I suggest, is the crucial finding on which we should center our attention, and I find it difficult to see how we can avoid a positive evaluation.

These general remarks may be supplemented by Grossarth-Maticek's detailed answers:

> Schüler and Fox concentrate on a few details which have either already been published in articles which they do not refer to, or which could not be published in journal-article form because of lack of space. A monograph publication is planned, covering the whole set of investigations. This point relates to questions concerning the exact value of certain measures, the number of measures taken in certain subgroups, and the number of probands in the stressed group who showed only stress and no additional risk factors. Similarly, data are available concerning the precise nature of the cancers causing death, the year of death, and so on. Again, the choice of therapy and control groups has been described in great detail. But it may be interesting to demonstrate how misunderstandings arise between the actual aims and methods of investigations, and their reception by Fox and Schüler.
>
> One typical misunderstanding relates to the interview experiments, which have not been extensively documented. This experiment had as its aim to research the conditions more closely which would enable proper prediction as well as meaningful intervention. We found that, as expected, prospective research results relying on the filling in of questionaires depended very much on the kind of relation established between interviewer and subject. Interviewers with a high degree of empathy, who took seriously individual differences in behavior, as relevant to the origin of diseases, and who managed to choose a proper moment for the beginning of the questionnaire-related interview, after a friendly preliminary discussion, achieved a more reliable and valid relation between personality variables and mortality. Less empathic interviewers who denied a synergistic relation between organic and psychological factors, and only believed in physical causation of disease produced low retest-reliabilities and poor validities.
>
> Concerning our intervention studies, Fox and Schüler seem to imagine that this constituted a human experiment analogous to the Metzler rat experiment. We were interested in the use of psychopharmaca with a function which would stimulate or depress central nervous system (CNS) functioning. Our hypothesis was that such pharmaca could interact with personality to increase excitement or inhibition. The empirical data were collected entirely by means of interview question, not by means of special medication. In our psychotherapeutic intervention we did not study systematically under experimental conditions the effects of psychoactive drugs on cancer or CHD. The use of drugs was decided on by the individual himself or herself, or the physician involved, not by us.
>
> Our intervention was always concerned with self-activation in people under stress. Several areas were touched on in connection with our repeated question: What is it you think you need?
>
> 1. Changes in relation to social interaction and cognitive emotional activity.
> 2. Changes in relation to sport and exercise, and physical movement generally.
> 3. Change of diet, including vitamin supplementation.
> 4. Stimulation or inhibition of CNS activity, through increase or decrease of coffee, for instance, or the use of stimulant or depressant drugs, in conjunction with the physician responsible for the proband.
>
> Careful records were kept of which probands changed in which respects. As the therapeutic groups were small, no significant effects were observed except for the cognitive-emotional reactions. As very few of the probands were given imipramine by their physicians, the drug was not found to have any cancer-preventing effect. (Proof of such an effect had to wait for a much larger study; Grossarth-Maticek & Eysenck, 1990a.) Possible interaction effects may exist but cannot be demonstrated on such a small sample. Our fundamental assumption of synergistic effect in the biological system constituted by humans suggests that single factors should never be interpreted in the sense of single causal factors, independent of everything else. Thus in the case of someone giving up smoking, we would assume that many additional factors, both physical and psychological, would change, so that there would be no proof of direct causality as regards smoking and lung cancer.
>
> The important feature of autonomy training is that we stimulate activity in the client which corresponds with his or her wishes. Thus the wish to use psychopharmaca with either stimulant or depressant effects was uttered by the proband; it is clearly not an example of the usual double-blind drug experiment. Consequently, the comparison between our studies and the rat experiment of Metzler are quite inadmissible because of the very different methodologies involved.

Fox and Schüler assume that we base our approach on a psychogenetic origin of cancer and formulate the criticisms that tumors develop over several years before clinical manifestations become apparent. In reality our fundamental assumption is that psychological factors interact in a synergistic fashion with organic factors—such as genetic predisposition, and/or physical, chemical, and microbiological or viral influences—to retard or accelerate the clinical manifestation of cancer. It is possible that physical factors are mainly concerned with the inhibition of cancer, whereas psychological factors may be of greater importance in the promotion of the disease. To imagine a purely psychological inhibition of cancer can only be based on a profound ignorance of the literature. As an example, consider our demonstration that when smoking, genetic predisposition, and chronic bronchitis are not present in a selected sample, lung cancer is equally present in stressed and unstressed probands, whereas incidence and mortality of lung cancer is much higher in the stressed than in the unstressed group when the physical factors are also present in equal measure (Eysenck, 1991).

Our critics suggest that the high mortality of the stressed group cannot be explained unambiguously in terms of psychological factors, but is rather due to organic risk factors. But we have shown that multivariate statistical methods demonstrate significant synergistic effects between psychological and organic risk factors—that is, mortality is high precisely where both physical and psychological risk factors are present, whereas mortality is significantly less where only one or the other risk factor is present.

It is rather puzzling to read in the Schüler and Fox critique that we are wrong in dealing with cancer as such and that different types of cancer correlate with different social factors. This is true, but hardly likely to constitute a criticism of our program, seeing that in many publications we have dealt with carcinoma of the cervix, carcinoma of the stomach, bronchial carcinoma, carcinoma of the breast, pharynx, and so on, taking into account the respective known physical risk factors. The suggestion of ignorance on our part can hardly be maintained. It does of course require very large prospective studies to enable the investigator to study relatively rare types of cancer (Grossarth-Maticek & Eysenck, 1990b), and often it is only possible to deal with aggregates (i.e., "cancer" as opposed to "CHD" and "other causes"). The success of such aggregation studies speaks well for the meaningfulness of the method, although when possible greater specificity undoubtedly pays.

Amelang

Dr. Amelang's critique is in some ways curious because he concentrates on some early publications and does not deal with the material of the target article. He also fails to deal with the unquestionable proof that the list of participants in the Heidelberg studies was given to the Oberburgermeister of Heidelberg in 1977, so that there could be no doubt about the participants; he was himself present when a comparison was made between that list and those Dr. Heller and others had been using in checking mortality and other data.

However, it seems best to leave the reply to Professor Grossarth-Maticek:

> In considering Amelang's critique, it is curious to note that it concerns itself mainly with an internal report dated 1977 that had a purely temporary purpose because the full set of data was not yet complete and the statistical treatment had to be done by hand. This explains minor deviations from later computer analyses. The seven criticisms regarding the Yugoslav study can all be answered rather simply:
>
> 1. The demand to publish simple predictions is unrealistic in view of well-known difficulties in publishing such articles. The separation of data-collection and mortality ascertainment is important to avoid accusations of data manipulation, and has been followed in the Heidelberg studies.
> 2. The physician who investigated the causes of death on the basis of death certificates could hardly have carried in his mind the prediction scores of 1,353 persons, so that this knowledge would have been used to influence his allocation of "cause of death" diagnoses.
> 3. The principles of selection have been published several times and are quite clear. The oldest inhabitant in every second house in Crvenka was asked to take part, but when this person was a female, she often preferred the next-in-age male to take over, obeying the patriarchal social norms then obtaining. This explains the distribution of age and sex variables. In addition, some "stressed" younger individuals were included. Professor Spielberger has analyzed the data for the aged and the stressed separately, and he found that the data for the aged gave even better results when the stressed sample was excluded.
> 4. Distortion effects, which can occur in repeated measurements at different points in time, are well known in psychological methodology where prospective studies are concerned; they are not peculiar to our study.
> 5. Interrogation of relatives was never the basis of investigation in any of our recent studies; only the preliminary study of 1977 is concerned with this form of data elicitation.
> 6. The method of evaluation of the RGM questionnaire used for the 1977 report, which is very largely the target of Amelang's criticism, was not used in later statistical treatments; it was used then as a preliminary way of looking at the data.
> 7. The refusal rate in the Yugoslav sample was very low because the investigator was personally known in the town and received great support from the local physicians and because of high motivation of the collaborators.
>
> All these points are irrelevant to our major publications, are easily explained, and cannot be used to furnish a negative impression of the work done.
>
> Regarding Amelang's methodological objections to the Heidelberg studies, the following are relevant. First, the names and scores of all participants of the 1972 study were deposited prior to 1982 at the independent university departments of Karlsruhe and Zurich, clearly in advance of the first mortality investigation. Second, contrary to Amelang's assertion, mortality was ascertained exclusively through neutral investigators (Institute of Statistics, University of Karlsruhe, and under the supervision of a worker in the Institute of Psychology at the University of Zurich). The cause of death so ascertained was transmitted to the Statistisches Landesamt Bad Ems for coding into the ICD system.
>
> Amelang's complaint concerning differences in dif-

ferent tables is explained by different rules of allocation. Thus in the original Yugoslav data there was no column for "failure to ascertain," so that all persons without known cause of death were counted as living; later those who were known to have died, but for whom cause of death was unknown, were counted separately as "other cause of death." (This information is from Dr. Vetter, who put the data into the computer.)

In regard to the instruments used for measurement, there is no problem. In both the Yugoslav and the Heidelberg prospective studies identical questionnaires were used to determine type. For the 1973 study, a new questionnaire was created that could be answered yes or no without additional explanations. In addition, two new personality types were introduced (rational/antiemotional, and antisocial) because previous studies had suggested such differentiation as meaningful. The 1973 prospective study was methodologically, theoretically, and by way of measurement a completely novel investigation which clearly marked out an advance compared with the investigations of 1972.

Amelang raises some questions concerning the extent, execution, and financing of the studies which are, to say the least, unusual in considering the value of a given study. However, the prospective studies were supported by four German and one Swiss foundation. In addition, money was contributed from wealthy relatives impressed by the social value of the research. In addition, Reynolds supported the further investigation and mortality ascertainment of the Heidelberg samples from 1982 to 1986.

Amelang's critique concentrates on peripheral issues, debates early trial runs with incomplete data, and does not address the most recent findings, most critical because they were derived from data bases which had been made known to independent judges prior to the beginning of the follow-up period. Thus they are not germane to the major issues.

Van der Ploeg

The critique by van der Ploeg is potentially the most serious of all, suggesting that the original data may have been manipulated in order to produce positive results. It is clearly the task of Dr. Grossarth-Maticek to answer these criticisms, and accordingly I simply quote his reply. In case readers should be confused by the apparent contradiction, I also quote later some comments by Dr. Heller, who was chosen as an independent assessor from the beginning, who had access to all the data, which had been lodged with the Karlsruhe Statistical Institute, where he was then working full time (and is still working part time), and who checked the death certificates in the Grossarth-Maticek studies. I have found him a completely honest, impartial, and critical person who gave much time and energy to his task because he considered it his duty to do so. His assessment of the differences between Grossarth-Maticek and van der Ploeg is to my mind the most trustworthy that can be carried out and agrees in every detail with what I know of the circumstances.

As far as van der Ploeg is concerned, I cannot say the same. He leaves out of his account several important facts, the most relevant being his insistence on receiving combinations of data that legally Grossarth-Maticek and Dr. Heller were unable to provide because of rules laid down by the *Datenschutz,* a German law of Draconian severity that protects individuals against disclosure of data concerning them. Violation of this law would result in the persons concerned being forbidden to carry out any further research activities in Germany, and possibly being sent to prison. Complaints about not being allowed access to certain data, or being given partial data, cannot be evaluated without knowing that furnishing such data would have laid Grossarth-Maticek and Dr. Heller open to legal prosecution and possible imprisonment. To fail to mention this point is to make it difficult to take seriously the accusations leveled against Grossarth-Maticek.

There is one other point. Originally, the second follow-up of the 1972 sample was financed by Reynolds Tobacco Company on the understanding that the final result would be discussed and published by the group appointed to supervise the study, consisting of Professor Spielberger, Dr. van der Ploeg, and myself. It was understood that such a report would be based on our joint efforts, would be carried out after ample consultation, and would be our unanimous opinion after bringing together our various investigations, reanalyses, and new follow-up results. Any problems and difficulties would then be resolved by reference to Grossarth-Maticek, Vetter, and Heller. Van der Ploeg has prematurely and without such consultation given his opinion, thus breaking our implicit agreement not to do so. This agreement would not prevent any of us from carrying out and publishing further work based on Grossarth-Maticek's data, or on new data produced using his questionnaires; both Dr. van der Ploeg and Professor Spielberger have carried out such new studies, which have been published or are in the process of being published.

This is the text of Grossarth-Maticek's reply to Dr. van der Ploeg:

> Our 1972 prospective study was followed up 10 years later with respect to mortality. To assure objectivity, a complex control system was introduced. The list of names of all subjects was given to the Institute for Statistics and Mathematical Business Studies in Karlsruhe, to facilitate control. In 1982, before the beginning of the mortality study, this list was signed and stamped by Dr. Heller. The same list had already in 1977 been given to the then Mayor of Heidelberg for safekeeping. The two lists were compared in 1990, in the presence of Professor Amelang, Dr. Heller, Dr. S. B. G. Eysenck, and Professor H. J. Eysenck, and found to be identical. Copies of the list signed by Dr. Heller were furnished to several other scientists in 1983, and the data concerning the subjects, together with code numbers corresponding to the original list of names, were entered into a computer system in Karlsruhe, Zurich, and Heidelberg. The total lists had originally been offered to the German Institute for Cancer Research, which did not agree to the proposal, and equally a law firm that had been approached for the same purpose refused to keep the list—hence, the rather later deposition of the list with the Mayor of Heidelberg.
>
> Mortality was ascertained, following the procedures for data protection under the control of the Institute of Statistics in Karlsruhe. Before this was done, interviewers were questioned by members of the same Institute to make sure that the collection of data had been carried out in exact conformity with the intentions, leaving no doubt regarding the correctness of the interviewing methods.
>
> In 1982 a possible argument arose to the effect that

it might have been possible theoretically that we could have known who in the time between 1972 and 1982 fell victim to cancer (i.e., before mortality was ascertained). In this way it might be possible to manipulate the data before sending them to Karlsruhe and Zurich. To make this argument inoperative I proposed to wait until 1986 and then look at the incidence and mortality data between 1982 and 1986, and compare these with the earlier data on mortality. It was partly for this purpose that names and data were deposited in Karlsruhe.

Mortality was indeed ascertained at the end of 1986, and the Institute for Statistics of the University of Karlsruhe can guarantee that the mortality data are really valid for the list of names given them in 1982. The first statistical evaluation by Dr. Vetter demonstrated that the results of subjects who fell victim to disease before and after the data had been deposited in Karlsruhe are indeed identical. This shows that our system to control results in Germany has proved reliable, and there was no further need for added controls. In 1985 Reynolds Tobacco Company offered to pay for a statistical reanalysis and new follow-up of our statistics. This would decide whether previous publications had been based on correct statistical methodology, and could be verified by the reanalysis and new follow-up. Professor Charles Spielberger and Dr. Henk van der Ploeg were asked to carry out these analyses and met with Dr. Heller to discuss the conditions of data protection. It was agreed that Dr. van der Ploeg would only obtain data concerning psychosocial and medical data, and data concerning mortality; the list of names of the subjects was to remain with Dr. Heller. Dr. van der Ploeg undertook to abide fully by the German conditions of the *Datenschutz*.

I was very astonished when a few days later Dr. van der Ploeg insisted he should be given the list of names as well as the list of the other data. He argued that otherwise he would not be able to carry out the reanalysis, and threatened that in the case of refusal he would publicly assert that we would not give him these data for fear of the results of such an analysis. At the same time I received a manuscript by Lydia Temoshok in which van der Ploeg gave a very negative assessment of our work, and considers the results as "not believable," even prior to a detailed examination.

It was difficult to know how to respond to van der Ploeg; it was decided to give him the possibility of a statistical reanalysis, but to obey the rules of the *Datenschutz* by inserting fictitious names, taken from another list, which has no relation to the code numbers. These names were crossed out with a pencil, but would still be legible. This list of names was made up by some of my scientific assistants. When on the basis of the original list it had been ascertained which probands had died, a group of fellow workers created a separate list of names of those who had died, whose names had been included in the list given to van der Ploeg. When Dr. Heller and his colleagues ascertained the mortality at the end of 1986, by accident van der Ploeg and Dr. Vetter were sent the wrong list of causes of deaths. Dr. Vetter drew attention to the resulting anomalies, and immediately he and van der Ploeg were sent the correct list with the correct causes of death. The resulting statistical evaluation by Dr. Vetter showed very clearly that the results of the first follow-up study (1972–1982) were replicated with considerable exactitude. (The results are given in Appendix A.) This original confusion of the two lists was no real problem for Dr. Heller, Professor Eysenck, or myself because the names of those who had died on the list could be compared with the list which had already been given to Dr. Heller in 1982, the list that had been deposited with the mayor of Heidelberg in 1977, and the list which had been sent to various other scientists. This comparison was made by Dr. Heller who decided quite firmly that the second list corresponded with the original names and the causes of death. Dr. van der Ploeg was informed of these procedures, and acknowledged them in his letter of September 10, 1990.

In this letter he made the following proposal. He suggested that it was necessary to put the list of names of 1972 in a sealed envelope to prove that the names which had been handed over in 1985 to him were the same. The list of those who had died and their causes of death should then be compared with the original list which in 1982 had been stamped by Dr. Heller, and also with the list which had been given to the Mayor of Heidelberg in 1977. If this could be done, then the "terrible accusation, of data manipulation, would be put aside and the proof given 'that you are right and I am wrong.' " Dr. Frentzel-Beyme of the German Center for Cancer Studies put the paper from the 1972 study in an envelope, which was sealed, and all the other conditions demanded by van der Ploeg were met. (Several paragraphs from van der Ploeg's letter showed that he agreed that fulfillment of these conditions would satisfy him.)

Van der Ploeg told me on the occasion of a meeting in Rotterdam in October 1990 that he was convinced of the agreement of the list stamped by Heller and the list of causes of death, and was convinced that no data manipulation had taken place.

On December 3, 1990, I received from van der Ploeg his article in which he put forward arguments that are quite contrary to his statement of September 10, 1990. His own proposal of an objective comparison of lists is not even mentioned nor is his knowledge that the correct names had never been given to him because of the laws relating to the *Datenschutz*. He also broke his promise to Dr. Heller not to insist on the connection between names and causes of death. He refused systematically to enter into any discussion of the true facts—that is, the way the measures he himself had suggested were used to maintain his hypotheses and prejudices.

A concrete example may be helpful. On August 31, 1990, van der Ploeg came to Heidelberg and met with Dr. Heller in the Hotel Perkeo. We had agreed that the aim of the meeting would be, first, to compare the correct list of causes of death with the original list of names of 1982 which had been stamped and signed by Dr. Heller, and after this, both would go to the Gesundheitsamt to test whether this list of deaths and causes of death would agree with the diagnosis which Dr. Heller had ascertained in 1986. Dr. van der Ploeg stated: "If there is agreement there, I can begin with a calm conscience with the statistical reanalysis and be absolutely certain that we are dealing with the correct causes of death." When Dr. Heller met me in the afternoon, he stated that van der Ploeg had shown strong hostile reaction against myself, refused to carry out the agreed plan of work, and had talked Dr. Heller into going to the Gesundheitsamt where the first (wrong) list of the correctness of cause of death would be compared. Dr. Heller was very astonished about this senseless procedure. On September 1, 1990, van der

Ploeg came to see me, and I asked him immediately why he had not carried out the agreed steps (for which alone there had been permission to visit the Gesundheitsamt). His answer was: "If I had done that, that would have given the proof that you are right and I am wrong." This is precisely the step undertaken by Professor Eysenck and Dr. Heller.

If one compares the critique with what is said by Dr. van der Ploeg in his letter of September 10, 1990, one can see that the measures suggested in the letter as leading to a very clear-cut verification are not even mentioned in the critique. In the article, almost every assertion is untrue and usually the opposite of the truth. Thus van der Ploeg writes several times that the mortality was not predictable in line with the list of names given him. Exactly the opposite is the case: Mortality was as predictable in the time spent from 1982 to 1986 as in the time spent from 1973 to 1982. (See the detailed analysis of these data by Dr. Vetter in Appendix A.) In another example, van der Ploeg asserts that he had no interest in the list of names, and only years after he had received the photocopied list was he able through some miracle suddenly to identify names that previously had been impossible to read. The truth is as follows: Already on receiving the photocopies he asked that the names would be crossed out only so lightly that they remained legible. My students and I were able when sending the list to read the names without any difficulty. Van der Ploeg had difficulties with the identification of some names already in 1986 several months after receiving the list, so that he asked me several times on the telephone to send him a typed list. Now he tells of the wonderful sudden ("most unexpectedly") legibility of names on the photocopied list, which he says he was able to include in his analysis only in August 1990.

Altogether I want to stress that the reanalysis team undertook only to carry out statistical evaluation and had no right to make connections between names, cause of death and personal details, and to transfer these outside Germany. It is clear that there could be no justification for that, for various reasons.

1. Dr. Heller and the Institute for Statistics of the University of Karlsruhe, who ascertained mortality and controlled the data, had the absolute duty vis-à-vis the Gesundheitsamt, Heidelberg, to separate the cause of death and the names and personal data, and not to hand on such a connection to anyone whatsoever.
2. Van der Ploeg had clearly undertaken in his dealings with Dr. Heller to adhere strictly to the German rules concerning *Datenschutz,* which means to make no connections on his part between names, causes of death, and personal data. I had a similar duty to Dr. Heller, and he in turn had undertaken not to connect his knowledge of causes of death with personal data, or hand these on to a third party.

At this point, I would like to clarify a point that may appear a little mysterious to readers. How could anyone test the truth or falsity of the data published by Grossarth-Maticek if they could not compare the names and the causes of death, personal data, and so forth? The answer is that the procedure agreed to with the *Datenschutz* implies an intermediary step. The first step connects the names of the people concerned with a code number. The second step connects the names with causes of death and other data. The third step connects the names of the people concerned with a code number. Thus Person A would know and be able to swear to the connection between names and code numbers, Person B the connection between code numbers and causes of death, and so on. The connection between names and causes of death, and so on, would be exclusively in the hands of someone agreed on by the *Datenschutz,* who would not be allowed to pass this on to a third party. It may seem a clumsy way of doing things, but it is the law, and van der Ploeg's attempts to circumvent the law, and potentially cause grave problems for Grossarth-Maticek and Dr. Heller, are surely most unusual in ordinary scientific discourse. Again, here is Grossarth-Maticek's reply:

> It is only an accident that we sent van der Ploeg at first the erroneous list with causes of death, containing names that were not connected with the prospective 1972 study. The important question is: Why would Grossarth-Maticek, if he had the intention of manipulating the results, send the wrong cause of death, leading to worse results, to the statisticians and not the correct cause of death, related to the original list of names, and giving positive results, when the cause of death for both lists was in fact ascertained on the same day in the Gesundheitsamt (and not some time afterwards)? In the reanalysis of van der Ploeg, he gives the impression that he kept asking questions for 5 years of Grossarth-Maticek, which were all answered correctly and without problems, and that he now concentrates his whole energy on an accidental error and refuses in spite of now knowing the truth to take this into account.
>
> The procedure adopted consisted in furnishing van der Ploeg with a list of names which he was told repeatedly was not identical with the original list because of the rules of the *Datenschutz,* and hence of no use to him. Van der Ploeg knew that originally in 1986, but has constantly refused to acknowledge the fact because it contradicted his wishes to be in possession of the original list.
>
> Unfortunately, later on, my student co-workers, engaged in ascertaining causes of death in the Gesundheitsemt, looked at the cause of death for the people in the original list who had died, as well as those in the van der Ploeg list. The van der Ploeg list contained names and causes of death from the prospective study of 1973. When this error was detected, I sent Dr. van der Ploeg the list of the original causes of death. Van der Ploeg told me several times by word of mouth and in writing that the explanation of this error did not present a problem to him because it was always possible to compare the list of those who had died with the original lists deposited in Heidelberg and Karlsruhe.
>
> I cannot resist the impression that van der Ploeg makes use of this unimportant error to suggest that Grossarth-Maticek carried out a manipulation of his data.

The essential facts of Grossarth-Maticek's account were confirmed as true in a letter from Dr. W. D. Heller, dated March 27, 1991. It agrees in every respect with my own knowledge and participation, and would seem to clarify the situation completely. Copies of the letter are available to interested parties. I would conclude that van der Ploeg's testimony is quite inaccurate and that his conclusions are incorrect. The mixture of *suggestio falsi* and *suppressio veri* is inadmissible as scientific evidence. Grossarth-Maticek continues:

A reply may be needed for one or two other criticisms of van der Ploeg. Many of van der Ploeg's remarks are difficult to understand, other than as indices of misunderstanding or incomprehensions. Thus he expresses astonishment about the high mortality in the control group of our intervention studies, compared with figures for the general population. He fails to mention that such a comparison is meaningless because, as frequently stated, both therapy and control groups were chosen because of high frequency of risk factors. In addition, the choice of probands was based on somewhat different risk factors, depending on the size of the population from which the sample was taken. The relatively small group of probands in the individual therapy sample was chosen from a total group of roughly 500 people, where the criteria for inclusion were presence of stress, as well as physical risk factors (high rate of smoking, high blood pressure, blood sugar and cholesterol level). The bibliotherapy group, on the other hand, was chosen from a much larger sample (roughly 20,000 persons); here additional risk factors were taken into account, such as drinking and genetic predisposition, as determined by the number of close relatives who had died of cancer or CHD. These considerations make it clear why mortality differs from one intervention study to another and why all are higher than population figures would suggest.

There may also be a misunderstanding of the concept of "healthy probands." This term was used to denote probands in the therapy and control groups who showed a very high proportion of psychosocial and physical risk factors, but showed no signs of suffering from cancer, CHD, or other serious diseases likely to cause death. The therapeutic aim was to demonstrate that by stimulating autonomy and self-regulated activity, and then producing a higher state of well-being, we could lower both mortality and incidence of cancer and CHD.

Vetter

The critique by Vetter, who carried out all the statistical analyses published by Grossarth-Maticek and myself (except for several early studies whose data collection and analyses were by hand rather than by computer) has certain surprising features. Thus he gives much space to an unpublished result—one that is impossible to reanalyze, criticize, or justify. But most disturbing is his refusal to stick with decisions jointly agreed on by himself, Grossarth-Maticek, and myself. As he states, he became sceptical about some of the data, and communicated this scepticism to me and to Grossarth-Maticek. We met together in Heidelberg to discover what could be done to deal with this problem. We agreed on two points. If the 1982–1986 follow-up gave results similar to those of 1972–1982, then skepticism would hardly be justified any longer. Preliminary analysis of the data by Dr. Vetter has shown that the data are essentially the same; his analysis is given in detail in Appendix A following this reply.

Vetter then mentioned that there might still have been some contamination, as he states in his critique, and suggested an absolutely conclusive test (his words). If we restricted ourselves, within the 1982–1986 follow-up, to those probands who died after a sudden onset of illness, later than the disposition of all relevant data, there could not be any possible contamination or manipulation of data. As he states, "I can say that this limited test turned out very favorably for Dr. Grossarth-Maticek." (Details of this test are given in Appendix B.) Here, then, as with van der Ploeg, we have the curious case of a man agreeing that if certain data turned out positive, then he would be convinced of the genuineness of the Grossarth-Maticek data, only to return to his doubts once the test came out favorably for Grossarth-Maticek! This is not my idea of proper scientific procedure. I have discussed this particular proof in the first part of this article and do not wish to add anything here, other than to draw attention to this rather interesting contradiction. I would also add that Vetter is carrying out further analyses of Grossarth-Maticek's data on my behalf which he would hardly do if he had serious doubts concerning their validity.

To deal with the details of his criticisms, I again leave it to Grossarth-Maticek to defend his data.

Vetter argues that 8 subjects of the Heidelberg study—namely, those with the highest stress-scores—had died. This fact might seem unlikely, but it is not when we consider the necessary background. These 8 people constitute 0.3% of the total number of participants (2,563 in all), and they died over a period of 14 years! According to the tests, these people had an unusual combination of unfavorable life events in conjunction with an inability to cope with them. They showed no competence in any direction to deal with such stress. This mortality rate does not surprise one who can picture the synergistic effect of suffering heavy subjective and objective stress. It is equally easy to understand how 2.5% (34 persons) of the Yugoslav sample, with the highest stress scores, died within 10 years. Vetter fails to mention not only the age of the people concerned, but also the fact that we are dealing with marked physical risk factors (i.e., high blood pressure, high blood sugar, high cholesterol level, high cigarette consumption, etc.). Neither does he mention that these people showed very significant synergistic effects between physical and psychological risk factors.

Concerning the data delivered after 1982, which apparently make Vetter skeptical, the following may be said. It had been agreed that for the purpose of replication only those data would be included that had been included in the Yugoslav as well as the Heidelberg studies. Hence, only those data were included in the original list which was communicated both to Karlsruhe and to Zurich. Data looked at later and additionally were analyzed to create new hypotheses and to replicate previous work. Hence these data are irrelevant to the main business of our study, namely, to test to what extent the 1982–1986 follow-up would give results similar to those of the earlier 1972–1982 study.

Vetter expresses astonishment that the predictions for the data 1978 to 1982 were better than those for the data from 1973 to 1977 in the Heidelberg prospective studies. He does not seem to understand the logic of a prospective study according to which psychosocial variable as predictors would logically be expected to increase their efficacy with increasing age, on the assumption that we are dealing with relatively stable personality traits. If we accept Vetter's assumption that psychological variables are to be considered rather as consequences of illness than as partly causal factors, and that personality variables change over time, then indeed it would be logical that predictive

accuracy would decrease over time. The facts seem to support the former, and deny the latter hypothesis.

Vetter again expresses astonishment that in the Yugoslav study the cholesterol values decreased dramatically in those who died of cancer. These values were ascertained on 3 to 7 occasions, the last time 3½ years before death. The results suggest that regardless of the time of origin of the cancer, constant decline of cholesterol predicts death from cancer within the next 3 to 4 years. The Heidelberg study replicated the same result, but in a rather weaker form, because here the measures were taken nearer the beginning of the study than was the case in Yugoslavia. It might be a very important finding that cholesterol decreases with advancing cancer, particularly in persons showing extremely high values (>300 mg/100 g as a rule) at the beginning of the study. Whatever might be the truth concerning this hypothesis, it is irrelevant to the matter of the target article.

Vetter only mentions in passing certain statistical data which he has worked out in great detail, relating to the causes of death from 1982–1986. He himself regarded as of central importance the comparison of causes of deaths that occurred after the delivery of the list of names in 1982, with mortality and diagnosis already known before 1982. Making the requisite statistical evaluation, he found that the predictability of cancer was identical in both groups. In his report of April 19, 1990, he stated: "The result shows very clearly that cancer prediction, using the predictors published in 1982, is no worse for those falling ill after 1982 than for those whose illness was already known in 1982." In his critique, Vetter simply states, very briefly, that the test did indeed give results favorable to me. In person he declared to Professor Eysenck and myself that he would be completely convinced if follow-up of all those who died between 1982 and 1986 showed that predictions of those falling ill after the delivery of the original data gave rise to predictions as good as those falling ill before that date.

He also fails to mention the multiple interaction, concerning which he has written detailed analyses occupying more than 100 pages for Professor Eysenck, where he expressed his conviction that data dealing with such complex interactions could not be manipulated.

In other words, Vetter after voicing his doubts suggested several criteria which, if met, would lay his doubts to rest. His own analyses showed in every case that his criteria were met, in the most clear-cut fashion. Yet in his critique he hardly mentions these crucial calculations. This is difficult to explain, and makes his continuing doubts hard to understand. It is also difficult to understand why his comments fail to deal with the target article, but rather cover very early findings of minimal interest (e.g., the relation of cholesterol level to cancer development) which I have nowhere mentioned. Normally criticism is devoted to the target article, not to extraneous matters, nor to unpublished data.

Conclusions and Future Directions

In reading through the various criticisms and rebuttals, allegations and counterallegations, accusations and countereaccusations, I wonder what the impression might be on the innocent reader? Perplexity? Incomprehension? Bewilderment? Mystification? Confusion? Bafflement? Stupefaction? Puzzlement? All these would be justified. Let me set down here what I think we may conclude from all the pages of print.

1. In the first place, Grossarth-Maticek has initiated and carried through several very important and impressive prospective studies which contain much invaluable material. Theoretical formulations and methodological refinements date back to the 1960s and should be judged by standards prevailing then; it is the fate of all prospective studies that when the payoff date comes, we all know we could mount a much better study now, even though such a belief is inevitably based on knowledge acquired in the meantime. When conceived, the study was brilliant, breathtaking, unique. Let us not forget its uniqueness in criticizing minor faults. The studies showed clearly that personality and stress were predictive of cancer and CHD, in a selective fashion. This verified theories that had been around for ages, but never properly tested.

2. These studies were complemented by another series of investigations demonstrating that a person's reactions to stress could be changed by behavioral therapy, so that cancer-prone probands and CHD-prone probands were less likely, after receiving therapy, to fall prey to these diseases than were control subjects. Different methods of therapy intended to increase personal autonomy were tried (individual, group, bibliotherapy), all with results which suggested important practical applications. The advantages of preventive medicine are so obvious, and the cost so slender compared with the expense of trying to cure patients suffering from cancer or CHD, that the social importance of the findings is obvious; these are even more important from the purely humanitarian point of view, of preventing pain, suffering, and other consequences of disease and death.

3. The unreasoning opposition to Grossarth-Maticek and his work illustrated in the opening paragraphs of this reply may appear unintelligible to most readers; are not scientists supposed to be objective, factual, detached, dispassionate, impartial, judicial, open-minded, unemotional? Alas, such a view is a stereotype more honored in the breach than the observance. Scientists are as emotional, prejudiced, rigid, and subjective as the next person; Barber (1961) discussed at some length the "resistance by scientists to scientific discovery." This is as strong and virulent as religious or political opposition; Galileo and the victims of Lysenko link hands with Semmelweiss, Pasteur, Koch, and other victims of a scientific group protecting its own! Most of the critics taking part in this symposium fortunately have adopted a more objective and fair-minded attitude, but occasionally objectivity vanishes.

The problem is that however unjustified the early attacks may have been, *aliquid semper haeret;* people feel that "where there is smoke, there is fire," and their attitude is changed once and for all. It is useless to appeal to fact when emotional attitudes are involved; psychologists should be the first to realize this truism. Given this successful besmirching of Grossarth-Maticek's name by a determined and powerful group, however contrary to fact, it will always be a case of "give a dog a bad name." Even the successful carrying out of the 1982–1986 follow-up is unlikely to satisfy people who have listened to the propaganda blasts against Grossarth-

Maticek from apparently authoritative (and authoritarian!) groups. As many critics have commented, only a large-scale, independent replication is likely to convince his peers.

It seems, to judge by history, that persecution is the likely result of genuinely creative effort. On that basis, Grossarth-Maticek must have been very creative indeed—although logic of course does not allow us to invert the syllogism. Creative scientists have also often laid themselves open to criticisms of data manipulation. Newton unashamedly changed the variables in his equations to get a perfect fit (and was known to do so). Mendel was found to have obtained results too close to the magic 3:1 proportion to be statistically acceptable. Even Ptolemy has not escaped similar well-informed criticism. Original thinkers inevitably tinker with the variables to find the best fit; replication is needed to discover whether they succeeded. Grossarth-Maticek tried many different questionnaires, formats, and methods; that can hardly be a cause of criticism. It would be foolish to stick with a suboptimal method just because it was the first tried.

4. Many criticisms have been made of the studies in question, particularly the prospective ones, and it would be idle to deny that some of these criticisms are justified. It is no easy way out to argue that similar and worse faults are frequently found in epidemiological studies that have attracted favorable mention and whose conclusions are widely accepted; I am particularly singling out studies of smoking (Eysenck, 1991). I have argued in my replies that some of the criticisms are misplaced, irrelevant, and do not touch the major conclusions. Nevertheless, there are sufficient justified criticisms—particularly concerning lack of detail in describing sampling procedures, measurement factors, and temporal sequences—to raise serious questions. I share these doubts, but I believe that we should consider the degree to which these details can affect the major findings. I think it will be found that they do not have an important influence on the final conclusions.

5. One of the most effective criticisms of the whole literature on the relation between personality and disease has been the demonstration by Fox that there are frequent contradictions in the evidence, and little agreement. This is true, but misleading. Meehl (1990) and I (Eysenck, 1984) have criticized the way psychological research is usually summarized. Meehl (1990), in his cogent monograph, explained "why summaries of research on psychological theories are often uninterpretable" (p. 195). Meehl listed "ten obfuscating factors whose effects are usually (1) sizeable, (2) opposed, (3) variable, and (4) unknown" (p. 195); the net effect of these 10 factors is to make "the usual research literature review well-nigh uninterpretable" (p. 195). We must distinguish studies beginning with a properly elaborated theory, using measuring instruments fashioned specifically to test that theory, and applied to appropriate samples, from studies having no specific theory, using unfocused universal questionnaires such as the Minnesota Multiphasic Personality Inventory (MMPI), on any sample that comes to hand; this blunderbuss approach is not likely to lead to success.

An article by Schmale and Iker (1971) illustrates the point. Testing the theory that feelings of hopelessness would be linked with cancer, and using interview ratings, they found a strong relationship; they failed to find any relationship at all using MMPI or a projective technique. A meta-analysis would summarize all that as one up, two down! This is a meaningless way of point scoring which is uninformative at best, and misleading at worst. There are of course problems in theory testing, as I have already pointed out; suppression of emotion is not easily quantified. But that is no excuse for using all-purpose instruments, or projective tests; they are seldom relevant to the theories tested. Hence, such apparently contradictory summaries cannot be held against positive outcomes of properly designed studies.

6. In a very large-scale undertaking such as this, minor irregularities are bound to occur; when hundreds of students are trained to collect data on thousands of subjects, errors are unavoidable. These can be kept to a minimum, but not eliminated. Transcription of data, manual handling of data, transfer to computer—all these can give rise to minor inaccuracies. Where millions of data entries are concerned, complete accuracy is to be attempted but probably unattainable. It is always important to try to assess the degree to which such failures to reach perfection might affect the final results.

7. Again, where so many highly motivated people are involved in data collection, the possibility of data manipulation cannot be ruled out. Thus a student, considering a proband to look ill and weak, might lead the interview in a direction that would allocate the proband to Type 1 or Type 2, rather than Type 4. It seems unlikely that such allocation would have any prognostic value, but it might. We know that such errors occur in the prognosis of lung cancer on the basis of smoking (Eysenck, 1991), where the basis of misclassification is much more obvious; it has not usually been guarded against there, and it would be difficult to do so here.

The crucial answer to such criticism is of course the consideration of the study after all the data have been deposited, as Grossarth-Maticek has outlined in his reply. This is what we have done, and the results, as outlined in Appendix A, gives us the assurance that if such manipulation occurred it cannot have had any great influence on the final outcome. Few other prospective studies have been able to give such proof of the objectivity of the methods used.

8. The crucial method of avoiding such criticisms is surely, as Grossarth-Maticek suggests, the publication (handing over to independent observers) of all names, codes, and data collection at time t, with a comparison of mortality and cause-of-death prediction of probands dying at time $t - x$, and such predictions at time $t + x$. When similar results are obtained under both conditions, clearly criticisms affecting predictions of probands dying before t cannot be relevant because they cannot affect predictions of probands dying after t. The fact that Grossarth-Maticek has jumped that hurdle was for me the crucial factor in carrying on with the analysis and publication of his data; had the results been negative I think we would all have ceased to be interested in his work. (See Appendix A.)

9. Even more searching was the test suggested by Vetter—namely, to compare only probands diagnosed after t, rather than dying after t, and possibly diagnosed before t. It is difficult not to consider the successful accomplishment of this further test as conclusive; few studies in epidemiology have been subjected to any test of similar severity. Note, in evaluating the evidence, that many critics seem to disregard what they themselves had beforehand agreed to consider as a crucial test, and continue to worry over small and often irrelevant details rather than face the major test and its positive outcome. This failure to come to terms with the facts

of the situation suggests a curious departure from scientific objectivity.

10. Can and should we then agree to accept that Grossarth-Maticek has succeeded in proving beyond doubt that his theoretical concepts can be used to isolate those traits that predict successfully cancer and CHD, and coalesce into types carrying out the same function? Can we agree to accept that the methods of therapy developed by Grossarth-Maticek are optimal in preventing cancer and CHD, or in prolonging life for terminal sufferers from cancer? The answer must surely be in the negative. The conceptions advocated by Grossarth-Maticek are not dissimilar to ideas going back over centuries, and adapted to psychometric testing by Kissen and Eysenck (1962) or by Schmale and Iker (1971). The measuring instruments constructed by Grossarth-Maticek constitute one way of bringing these theoretical concepts down to earth, but they are not the only ones, and they certainly require proper translation and considerable psychometric attention before they can be regarded as proper scientific instruments.

Similarly, the fact that Spiegel obtained even better results in prolonging life in terminally ill cancer patients than did Grossarth-Maticek might seem to suggest successful replication, but as the methods used were very different, it seems that there may be many different methods of stress management that might be equally successful. I think we must accept the prophylactic and life-prolonging effects of psychological therapy, but to discover what therapy is best for whom will take a long time yet; after all, the field is still wide open in the area of neurosis, where at least 600 studies have been done in the past 20 years (Eysenck & Martin, 1987.

11. The work of Grossarth-Maticek has given rise to fierce partisanship, with some regarding it as an outstanding contribution without flaws, others dismissing it outright. How is such divergence possible? It illustrates the fact that Thouless's law of certainty (1935; see Eysenck, 1954) obtains in science too. The law runs as follows:

> When, in a group of persons, there are influences acting both in the direction of acceptance and of rejection of a belief, the result is not to make the majority adopt a lower degree of conviction, but to make some hold the belief with a high degree of conviction, while others reject it also with a high degree of conviction. (Thouless, 1935, p. 16)

This has clearly happened here too, and I can only suggest that we should retain an intermediate position, accepting what seems to have been proved, but remaining critical of many aspects of the work. The unreasoning hostility I mentioned at the beginning of my reply is clearly over the top; so would be an uncritical acceptance of all the results reported.

12. As many contributors in this symposium have concluded, the obvious next step is a proper replication of Grossarth-Maticek's studies, with perhaps an extension to include measures of immune competence, degree of sclerosis, and so on, as well as constant monitoring of possible intermediary factors, such as smoking, drinking, and exercise. Such a study, also including intervention in selected cases, would of course have to be prognostic and carried out on a large scale; partial studies on a smaller scale would also be welcome but could not be decisive. Such large-scale studies are of course expensive, but seeing how much money has been spent on smoking-related studies, often of poor quality, it can hardly be argued that money is not available. Multiple Risk Factor Intervention Trial (Multiple Risk Factor Intervention Trial Research Group, 1982) cost 115 million dollars, and showed, if anything, that giving up smoking, reducing blood pressure, and controlling cholesterol level failed to change mortality. The money could have been better spent in investigating the effects of psychosocial factors. Indeed, there seems to be unanimity on the need for a replication, under independent supervision, but hopefully retaining Grossarth-Maticek in an advisory capacity; if this symposium has done nothing else besides bringing such a replication nearer, it will have served its purpose. For myself, I have always argued strongly in favor of such a replication, as has Grossarth-Maticek; let us all work toward it.

It may be useful to think of the whole debate in terms of Type 1 and Type 2 errors. Trying to avoid Type 1 errors (i.e., accepting a result that in fact is untrue), we may easily commit Type 2 errors (i.e., accepting the null hypothesis when it is false; I use these concepts, well aware of the arguments against the whole notion of testing the null hypothesis, as in Gilgerenzer et al., 1989. They do not affect the point I am making). Some of the critics, in my view, have leaned over backward to avoid Type 1 errors, only to fall seriously into Type 2 errors. I think we should be aware of the dangers involved in committing either type of error, and attempt to keep a reasonable balance, however difficult that may be.

One final word to indicate my estimate of the position Grossarth-Maticek's work holds in the light of the present situation concerning the effectiveness of primary prevention of cancer. There have been many claims, and much backslapping, but the outcome of serious research has shown the dismal state of the art. Hakama (1990) and Eysenck (1991) concluded that there is very little evidence for effectiveness of the major measures investigated. Much the same is true of clinical treatment measures, such as chemotherapy (Abel, 1990); there is much doubt of any effectiveness whatever. Given this generalized gloom, one might think that the hope held out to sufferers and prospective victims by Grossarth-Maticek's experiments with autonomic training would lead to rapid replication and exhaustive trials, funded by the National Institutes of Health in the United States, the Medical Research Council in the United Kingdom, and appropriate bodies in Germany. Of course, Grossarth-Maticek may be mistaken; of course, his charismatic personality may not be capable of being copied; of course, there can be no guarantee of success. But the possibility is so dazzling that the chance ought to be taken, particularly when we consider the millions wasted by the "orthodox" leaders in the field. Their achievements hardly justify pride and refusal to recognize the importance of independent creative thought. What, then, shall be our verdict? I suggest, as the Good Book says: "Thou shalt not muzzle the ox that treadeth the corn."

Note

Hans J. Eysenck, Institute of Psychiatry, University of London, Denmark Hill, DeCrespigny Park, London, SE5 8AF, England.

References

Abel, U. (1990). *Die Zytostatische Chemotherapie Fortgeschrittener Epithelialer Tumoren* [Psychostatic chemotherapy of advanced epithelial tumors]. Stuttgart: Hippokrates Verlag.

Bachman, C. (1981). *Die Krebsmafia* [The Cancer Mafia]. Monaco: Tamek.

Barber, B. (1961). Resistance by scientists to scientific discovery. *Science, 134,* 596–602.

Brown, G. W., & Harris, T. (1978). Social origins of depression: A reply. *Psychological Medicine, 8,* 577–588.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24,* 385–396.

Cox, D. R. (1970). *Analysis of binary data.* London: Chapman & Hall.

Darrock, J. V. (1974). Multiplicative and additive interaction in contingency tables. *Biometrika, 61,* 207–214.

Dixon, J. P., & Dixon, J. K. (1991). Contradictory tendencies in the perception of life conflicts in persons with cardiovascular disease and persons with cancer. *Personality and Individual Differences, 12,* 791–799.

Eaves, L., Eysenck, H. J., & Martin, N. (1989). *Genes, culture, and personality: An empirical approach.* New York: Academic.

Esterling, B. A., Antoni, M. H., Kumar, M., & Schneiderman, N. (1990). Emotional repression, stress disclosure responses, and Epstein–Barr viral capsid antigen titres. *Psychosomatic Medicine, 52,* 397–410.

Everitt, B. G. (1977). *The analysis of contingency tables.* London: Chapman & Hall.

Everitt, B. G., & Smith, A. M. R. (1979). Interaction in contingency tables: A brisk discussion of alternative definitions. *Psychological Medicine, 9,* 581–583.

Eysenck, H. J. (1954). *The psychology of politics.* London: Routledge & Kegan Paul.

Eysenck, H. J. (1984). Meta-analysis: An abuse of research integration. *Journal of Special Education, 18,* 41–59.

Eysenck, H. J. (1985). Personality, cancer and cardiovascular disease: A causal analysis. *Personality and Individual Differences, 6,* 535–556.

Eysenck, H. J. (1988). The respective importance of personality, cigarette smoking and interaction effects for the genesis of cancer and coronary heart disease. *Personality and Individual Differences, 9,* 453–464.

Eysenck, H. J. (1990). *Rebel with a cause.* London: W. H. Allen.

Eysenck, H. J. (1991). *Smoking, personality and stress: Psychosocial factors in the prevention of cancer and coronary disease.* New York: Springer-Verlag.

Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences.* New York: Plenum.

Eysenck, H. J., & Grossarth-Maticek, R. (1991). Creative novation behaviour therapy as a prophylactic treatment for cancer and coronary heart disease: II. Effects of treatment. *Behaviour Research and Therapy, 29,* 17–31.

Eysenck, H. J., & Martin, I. (Eds.). (1987). *Theoretical foundations of behavior therapy.* New York: Plenum.

Eysenck, H. J., & Rachman, S. (1965). *Causes and cures of neurosis.* London: Routledge & Kegan Paul.

Galtung, J. (1967). *Theory and methods of social research.* London: Allen & Unwin.

Gigerenzer, G., Switjink, Z., Porter, T., Daston, L., Beatty, J., & Krueger, L. (1989). *The empire of chance: How probability changed science and everyday life.* Cambridge, England: Cambridge University Press.

Grizzle, J. E., Starmer, C. E., & Koch, G. O. (1969). Analysis of categorical data by linear models. *Biometrics, 25,* 489–504.

Grossarth-Maticek, R. (1980). Social psychotherapy and course of the disease. *Psychotherapy and Psychosomatics, 33,* 129–138.

Grossarth-Maticek, R., & Eysenck, H. J. (1990a). Coffee-drinking and personality as factors in the genesis of cancer and coronary heart disease. *Neuropsychobiology, 23,* 153–159.

Grossarth-Maticek, R., & Eysenck, H. J. (1990b). Personality, smoking, and alcohol as a synergistic risk factor for cancer of the mouth and pharynx. *Psychological Reports, 67,* 1024–1026.

Grossarth-Maticek, R., & Eysenck, H. J. (1990c). Personality, stress, and disease: Description and validation of a new inventory. *Psychological Reports, 66,* 355–379.

Grossarth-Maticek, R., & Eysenck, H. J. (1990d). Prospective effects of psychoanalysis on cancer and coronary heart disease-prone probands, as compared with control groups and behaviour therapy groups. *Journal of Behaviour Therapy and Experimental Psychiatry, 21,* 91–99.

Grossarth-Maticek, R., & Eysenck, H. J. (1991). Creative novation behaviour therapy as a prophylactic treatment for cancer and coronary heart disease: I. Description and treatment. *Behaviour Research and Therapy, 29,* 1–16.

Grossarth-Maticek, R., Eysenck, H. J., Gallasch, G., Vetter, H., & Frentzel-Beyme, R. (1991). Changes in degree of sclerosis as a function of prophylactic treatment in cancer-prone and CHD-prone probands. *Behaviour Research and Therapy, 29,* 343–352.

Grossarth-Maticek, R., Eysenck, H. J., & Vetter, H. (1988). Personality type, smoking habit and their interaction as predictors of cancer and coronary heart disease. *Personality and Individual Differences, 9,* 479–495.

Grossarth-Maticek, R., Vetter, H., & Heller, W. D. (1986). Kausale Prädiktoren für Krebserkrankung, Herzinfarkt und Hirnschlag: Theorie, Methode und Ergebnisse der Heidelberger prospektiven psychosomatischen Interventionsstudie (1972–1982). Ein Beitrag zur Ätiologieforschung in der interdisziplinären Epidemiologie [Causal predictors of cancer, cardiac infarct and stroke: Theory, methods and results of the Heidelberg prospective psychosomatic intervention study (1972–1982). A contribution to the study of etiology in interdisciplinary epidemiology]. In E. D. Hager (Ed.), *Biomodulation and Biotherape des Krebses* (pp. 87–117). Heidelberg: E. Fischer.

Gudjonsson, G. (1981). Self-reported emotional disturbance and its relation to electrodermal reactivity, defensiveness and trait anxiety. *Personality and Individual Differences, 2,* 47–52.

Hakama, M. (1990). *Evaluating effectiveness of primary prevention for cancer.* Oxford, England: Oxford University Press.

Kissen, D. (1964). Relationship between lung cancer, cigarette smoking, inhalation and personality. *British Journal of Medical Psychology, 37,* 203–216.

Kissen, D. M., & Eysenck, H. J. (1962). Personality and male lung cancer patients. *Journal of Psychosomatic Research, 6,* 123–137.

Kneier, A. W., & Temoshok, L. (1984). Repressive coping reactions in patients with malignant melanoma as compared to cardiovascular diseased patients. *Journal of Psychosomatic Research, 84,* 145–155.

Kreitler, S., & Kreitler, H. (1990). Repression and the anxiety–defensiveness factor: Psychological correlates and manifestations. *Personality and Individual Differences, 11,* 559–570.

Meehl, P. E. (1990). Why summaries of research on psychological theories are often misinterpretable. *Psychological Reports, 66*(Monograph Suppl. 1), 195–244.

Multiple Risk Factor Intervention Trial Research Group. (1982). Multiple Risk Factor Intervention Trial. *Journal of the American Medical Association, 248,* 1465–1477.

Osler, W. (1906). *Aequanimitas.* New York: McGraw-Hill.

Plackett, R. L. (1974). *The analysis of categorical data.* London: Griffin.

Quander-Blaznik, J. (1991). Personality as a predictor of lung cancer: A replication. *Personality and Individual Differences, 12,* 125–130.

Rachman, S., & Hodgson, R. (1980). *Obsessions and compulsions.* Englewood Cliffs, NJ: Prentice-Hall.

Rosch, P. J. (1979). Stress and cancer: A disease of adaptations. In J. Tache, H. Selye, & S. B. Day (Eds.), *Stress and cancer* (pp. 43–58). New York: Plenum.

Rosch, P. J. (1980a). Lifestyle and cancer. *New York State Medical Journal, 80,* 2034–2038.

Rosch, P. J. (1980b). Some thoughts on the epidemiology of cancer. In S. B. Day, E. V. Sugarbaker, & P. J. Rosch (Eds.), *Readings in oncology* (pp. 1–6). New York: International Foundation for Biosocial Development and Human Health.

Schmale, A. H., & Iker, H. (1971). Hopelessness as a predictor of cervical cancer. *Social Science and Medicine, 5,* 95–100.

Schmitz, P. G. (in press). Personality, stress-reaction and disease. *Personality and Individual Differences.*

Scott, E. M., & Thomson, A. M. (1956). Psychological investigation of primigravidae. *Journal of Obstetrics and Gynaecology of the British Empire, 63,* 502–508.

Seltzer, C. (1989). Framingham study data and "established wisdom" about cigarette smoking and coronary heart disease. *Journal of Clinical Epidemiology, 42,* 743–750.

Shapiro, A. K., Struening, E. L., Shapiro, E., & Milcarek, B. I. (1983). Diazepam: How much better than placebo? *Journal of Psychiatric Research, 17,* 51–73.

Spiegel, D., Bloom, J. D., Kraemer, H. C., & Gottheil, E. (1989). Effects of psychosocial treatment on survival of patients with metastatic breast cancer. *Lancet, ii,* 888–891.

Tenant, C., & Bedington, P. (1978). The social causation of depression: A

critique of the work of Brown and his colleagues. *Psychological Medicine, 8,* 565–575.
Thouless, R. H. (1935). The tendency to certainty in religious beliefs. *British Journal of Psychology, 26,* 16–31.
Vetter, H. (1988). Probleme bei der Analyse von Kategorialen Variablen [Problems in the analysis of categorical variables]. In F. Faulbaum & H.-M. Vehlinger (Eds.), *Fortschritte der Statistik-Software* (pp. 46–53). Stuttgart: G. Fisher.
Weinberger, D. A., Schwartz, G. E., & Davidson, R. J. (1979). Low-anxious, high-anxious, and repressive coping styles: Psychometric patterns and behavioral physiological responses to stress. *Journal of American Psychology, 88,* 369–380.

Appendix A: Analysis of Mortality Data in the 1972 Prospective Heidelberg Study by Grossarth-Maticek, Covering the Period 1982 to 1986

Hans J. Eysenck
Institute of Psychiatry
University of London

Appendix A was written by Hermann Vetter at my request. It provides further details about the second follow-up period of the Heidelberg study.—Hans J. Eysenck.

"This follow-up covers all the probands originally tested in Heidelberg in 1972, inclusive of the probands in the intervention studies. These data relate to the status of the sample at the middle of 1982. Altogether there were 2,563 probands, of whom 417 could not be followed up, leaving 2,146 persons in all, of whom 1,512 were still alive, while 634 had died. Of these, 244 had died of cancer, 172 of heart infarct or stroke, and 218 of other or unknown causes. Of the 1,512 persons still alive in 1982, 199 died by 1986, 77 from cancer, 70 from CHD, and 52 from other or unknown causes. Those still living amounted to 1,313.

"Table A-1 shows the mortality of the probands in the normal and the stressed samples, and the intervention groups. Table A-2 shows for the sake of comparison similar figures for the 2,146 probands investigated in 1982.

"The average yearly mortality from 1973 to the middle of 1982 was thus 29.54%/9.5 = 3.1% per year. From the middle of 1982 to 1986 it was 13.28%/4.5 = 3.0% per year. The difference in mortality between 'normal' and 'stressed' samples was 45.7% and 11.4% respectively for 1973 to 1982, and 25.0% and 4.5% for 1982 to 1986. The yearly rate of mortality from 1972 to 1982 compared to 1982 to 1986 is 4.81% and 5.56% for the stressed sample, and 1.20% and 1.00% for the normal sample. The differences are highly significant for the stressed and normal groups for both time intervals.

"If we look at the small group of probands receiving health instruction, all of whom came from the stressed sample, mortality was 8.5% for 1973 to 1982, compared with 45.7% of the rest of the stressed sample; comparable figures for 1982 to 1986 are 1.9% and 25.0%. Both differences are highly significant ($p < .0002$), and even when the major predictors for mortality and cause of death are controlled (see discussion to follow) they remain at that level. (Using this analysis we controlled for any initial differences between instructed and noninstructed groups.) Details of the method of computing significances are given by Vetter (1988).

"Concerning the therapy groups (cancer prone and CHD prone) compared with the control groups, there was a marked reduction in mortality for 1982 to 1986 as well as for 1973 to

Table A-1. *Status of Group at 1986 Follow-Up*

Group	Still Living	Other Causes of Death	CHD	Cancer	Total; % Mortality
Normal	738	10	12	13	773
	95.47%	1.29%	1.55%	1.68%	4.53%
Stressed	425	40	50	52	567
	74.96%	7.05%	8.82%	9.17%	25.04%
Cancer Control	16	1	0	8	25
	64.00%	4.00%	0.00%	32.00%	36.00%
CHD Control	11	1	6	2	20
	55.00%	5.00%	30.00%	10.00%	45.00%
Cancer Treated	39	0	0	2	41
	95.12%	0.00%	0.00%	4.88%	4.88%
CHD Treated	31	0	3	0	34
	91.18%	0.00%	8.82%	0.00%	8.72%
Health Instruction	53	0	0	1	54
	98.15%	0.00%	0.00%	1.85%	1.85%
Total	1,313	52	71	78	1,514
	86.72%	3.43%	4.69%	5.15%	13.28%

Table A-2. *Status of Group at 1982 Follow-Up*

Group	Still Living	Other Causes of Death	CHD	Cancer	Total; % Mortality
Normal	773	43	27	29	872
	88.65%	4.93%	3.10%	3.33%	11.35%
Stressed	566	157	120	199	1,042
	54.32%	15.07%	11.52%	19.10%	45.68%
Cancer Control	25	4	5	12	46
	54.35%	8.70%	10.87%	26.09%	45.65%
CHD Control	20	5	14	0	39
	51.28%	12.82%	35.90%	0.00%	48.72%
Cancer Treated	40	3	2	0	45
	88.89%	6.67%	4.44%	0.00%	11.11%
CHD Treated	34	5	3	1	43
	79.07%	11.63%	6.98%	2.33%	20.93%
Health Instruction	54	1	1	3	59
	91.53%	1.69%	1.69%	5.08%	8.47%
Total	1,512	218	172	244	2,146
	70.46%	10.16%	8.01%	11.37%	29.54%

1982 (by chi-square test, all causes of death, always $p <$ .025).

"We must now turn to evaluating the predictive accuracy of various scales used in previous research on the 1982–1986 data. The following scales or type-scores were analyzed:

1. Hopelessness questionnaire.
2. Annoyances.
3. Hopelessness rating, 10-point scale.
4. Annoyances rating, 10-point scale.
5. Rationality/antiemotionality (only for stressed sample).
6. Type classification (four types).
7. Three-step autonomy rating.

"The statistical evaluation was as follows: For the quantitative variables (1 to 5) I calculated eta against (living – cancer mortality – CHD mortality – other causes of death). For 6 and 7, I calculated Cramer's coefficient of association *V*, assessing significance using chi square. Analysis of variance was calculated for contrasts between living and dead, and cancer versus CHD. Results are given in Table A-3.

"The particular relationships involved for the four-type classification are treated in more detail in Table A-4, which deals with the normal and the stressed Heidelberg sample excluding all intervention groups.

"Of particular relevance is the discrimination of Types 1 and 2 for cancer and CHD mortality, with 96 correct as opposed to 12 wrong predictions. Also noticeable is the very low mortality of Type 4, with 2% as compared with 19% for the other 3 types. Unlike the mortality data for 1973 to 1982 (Grossarth-Maticek, Vetter, & Heller, 1986), these data did not show a significant interaction between the hopelessness and annoyance scores.

"To investigate the interactions between physical and psychosocial risk factors, multiple-regression analysis was carried out, and significance levels calculated according to Vetter's (1988) formula. The first analysis concerned smoking (number of cigarettes smoked per day) and as the chosen psychosocial factor a linear combination of hopelessness (1 and 3), and rationality/antiemotionality (5). The interaction (product term) was significant at the $p = .018$ level. The picture is as follows: For probands whose psychosocial risk scores lie 1 *SD* below the average, the risk for smoking lies hardly at all above 0; it increases with increasing scores on the psychosocial variable.

"Concerning CHD and smoking, the psychosocial variable was a linear combination of the two anger–annoyance indicators (2 and 4). The interaction (product term) was significant at the $p < .0001$ level, with similar regression slopes as in the case of cancer.

"Looking finally at the interaction between blood pressure and psychosocial factors for CHD mortality, there was no interaction either for systolic or for diastolic measures. This is unlike the results of the analysis of 1973–1982 mortality data (Grossarth-Maticek et al., 1986)."

Table A-3. *Effectiveness of Various Scales in Predicting Mortality and Cause of Death, 1982 to 1986*

Scale[a]	Eta or Cramer's Coefficient[b]	$p <$	Living Versus Deceased ($p <$)	Cancer Versus CHD ($p <$)
1	.32	.001	.001	.001
2	.31	.001	.001	.001
3	.28	.001	.001	ns
4	.27	.001	.001	ns
5	.20	.001	ns	.001
6	.29	.001	not calculated	
7	.22	.001	not calculated	

[a]1 = hopelessness questionnaire; 2 = annoyances; 3 = hopelessness rating, 10-point scale; 4 = annoyances rating, 10-point scale; 5 = rationality/antiemotionality (stressed only); 6 = four-type classification; 7 = three-step autonomy rating.
[b]Eta for Variables 1 to 5, Cramer's coefficient for Variables 6 and 7.

Table A-4. *Mortality in Different Personality Types, 1982 to 1986*

Type	Living	Other Causes of Death	CHD	Cancer	Total
Type 1	179	20	11	56	266
	67.29%	7.52%	4.14%	21.05%	
Type 2	195	22	40	1	258
	75.58%	8.53%	15.50%	0.39%	
Type 3	321	5	7	5	338
	94.97%	1.48%	2.07%	1.48%	
Type 4	448	3	3	3	457
	98.03%	0.66%	0.66%	0.66%	
Total	1,143	50	61	65	1,319
	86.66%	3.79%	4.62%	4.93%	100.00%

Appendix B: Some Genuine Predictions of Grossarth-Maticek's Established

Hermann Vetter
Schönbrunn, Germany

In my commentary, I discussed the hypothesis that Grossarth-Maticek's psychosomatic predictions of deaths and causes of death in Heidelberg after 1982 might have been due to artificial assignment of the predictors on the basis of existing cancer or a history of cardiovascular disturbances in the subjects. It was also mentioned that a limited test of this hypothesis had led to favorable results for him. Here I wish to describe this test in more detail.

Among the cancer deaths that occurred in the Heidelberg cohort from mid-1982 to 1986 there were a few for which the duration or first diagnosis of cancer was recorded on the death certificate. The idea of the test was to compare the predictive success for cases in which this first knowledge was available prior to 1982 with cases in which it was available only later. In the Heidelberg stressed sample (including control groups, but excluding intervention groups), first knowledge prior to 1982 was recorded for 10 cases and later for 9 cases; in the normal sample, prior knowledge was recorded for 1 case and later for 4 cases. Because a comparison of the prediction for 4 cases and a single case did not seem advisable, the 4 later cases were compared to all the rest, which consisted of 13 cancer deaths.

A Methodological Caution

The unequal size of the two groups makes necessary a caution with respect to the possible size of the (point-biserial) correlations. We have generally: $R^2 = V_s/V_t = V_s/(V_s + V_e)$, where V_s is the systematic variance, V_e is the error variance, and $V_t = V_s + V_e$ is the total variance of the dependent variable. In the point-biserial case, the error variance is the variance of the dependent variable within the two groups. The systematic variance is $p(1 - p)d^2$, where d is the difference of the means of the two groups, and p, $1 - p$ are the relative frequencies of the two groups. We consider the influence of p on R^2 when d and V_e are constant. We have

$$R^2 = p(1 - p)d^2/[p(1 - p)d^2 + V_e].$$

When R^2 is small, it is approximately proportional to $p(1 - p)d^2$; and when p is small, $p(1 - p)$ is approximately proportional to p. Hence, when R^2 and p are small, R^2 is approximately proportional to p. Under these conditions, then, R^2 is approximately proportional to the size of the smaller group when the group means and the within-group variance are constant. So in the group of 4 subjects we have to expect a R^2 less than one third of that in the group of 13 subjects when other things (i.e., the goodness of the prediction in terms of d and V_e) are equal.

Predictors Used

It was decided to predict the cancer deaths with a multiple-regression model. The predictors had to be selected from the variables handed over to me in 1982 and deposited by me at two independent university institutes. Selection was made in terms of the bivariate correlations with cancer deaths after mid-1982. Prediction was made separately in the normal and in the stressed sample, and the sets of available predictors were not quite identical. Two sets of predictors were used in each case, a smaller one including only variables related to cancer, and a larger one containing also variables related to other causes of death. Furthermore, in the stressed sample, where a considerable number of cancer deaths beside those in the two prediction groups was available, two prediction strategies were used. The first was the same as the only one available in the normal sample, namely, using the given set of predictors to discriminate the "prior" group and the "later" group of cancer deaths, respectively, from all subjects that had not died of cancer, and to compare the two multiple correlations. The other strategy, available in the stressed sample only, consisted in deriving the regression coefficients of the predictors from those cancer deaths after mid-1982 for which neither "prior" nor "later" information was available; to use this regression function for discriminating the "prior" and the "later" group, respectively, from the noncancer subjects; and to compare the two correlations.

The smaller and the larger sets of predictors in the two subsamples are shown in Table B-1.

Results

The squared multiple correlations achieved by these sets of predictors are given in Table B-2. All these squared multiple correlations (or squared bivariate correlations in the case of the regression coefficients derived from different subjects) are significantly different from zero with $p < .05$, with the exception of those in the stressed sample, prior group, where $p = .07$ and $p = .17$, respectively.

An inspection of the correlations shows that in the stressed sample they are always higher for the "later" group (although it is slightly smaller than the "prior" group). In the

Table B-1. *Cancer-Related and Other Predictors of Cancer Deaths by Subsample*

Subsample	Smaller Set (Cancer Predictors Only)	Additional in Larger Set
Normal Only	Interviewer rating of rationality with respect to withdrawing objects	Interviewer rating of rationality with respect to disturbing objects
Stressed Only	Rationality/Anti-emotionality	
Both Samples	Number of life events leading to hopelessness	Number of life events leading to anger
	Intensity of hopelessness	Intensity of anger
	Interviewer rating of hopelessness	Interviewer rating of anger
		Need for harmony and closeness
		Nonhypochondria

Table B-2. *Predictor Squared Multiple Correlations*

	Set of Predictors	
	Smaller	**Larger**
Normal Sample		
"Not Later" Group (n = 13)	.0263	.0415
"Later" group (n = 4)	.0142	.0357
Stressed Sample		
"Prior" Group (n = 10)	.0157[a]	.0234[b]
"Later" Group (n = 9)	.0260	.0354
Regression Coefficients Derived From Rest of Cancer Deaths		
"Prior" Group (n = 10)	.0135	.0169
"Later Group (n = 9)	.0191	.0218

Note: Values differ significantly from zero ($p < .05$) except for the stressed sample, prior group.
[a]$p = .07$. [b]$p = .17$.

normal sample, too, where less than one third of the values of the "not later" group is to be expected in the "later" group on account of its size, the values are greater than that. So it emerges that on all counts, the predictions stand the test of genuineness very well.

Discussion

I cannot think of any possibility of manipulating the predictors supplied in 1982 with respect to deaths afterward once the use of morbidity information is ruled out. Hence, even given all the doubts put forward in the commentary, it should be admitted that Grossarth-Maticek has succeeded in making at least some genuine psychosomatic predictions. It is deemed highly desirable, however, to extend the test to all (or most) of the deaths that occurred from 1982 to 1986 by collecting information from doctors and/or relatives when the disease that led to death first became known.

Addendum: The "Morbidity Hypothesis" Retracted

Hermann Vetter
Schönbrunn, Germany

In my commentary, "Some Observations on Grossarth-Maticek's Data Base" (this issue), I discussed the hypothesis that "predictions" of deaths and causes of death from 1982 to 1986 might have been due to knowledge of existing relevant diseases in subjects; in Appendix B ("Some Genuine Predictions of Grossarth-Maticek's Established") of Dr. Eysenck's response (this issue), I presented in more detail a limited test of this hypothesis that quite clearly spoke against it. The hypothesis had been suggested by the fact that virtually all those predictors that had not broken down after 1982 worked better for deaths that occurred from 1973 to 1977 than from 1978 to 1982. This finding had been obtained by means of the following "obvious" method: Take the whole cohort that entered the study in 1972 and predict deaths and causes of death in 1973–1977; then take the survivors in 1977 and predict deaths and causes of death in 1978–1982; finally, take the survivors in 1982 and predict deaths and causes of death in 1982–1986.

Recently it occurred to me that this method is not as obvious as it may seem. The prediction for 1973–1977 requires a discrimination of the deceased subjects from a rather heterogeneous group composed of those who died soon after 1977 as well as those who survived at least till 1986. Therefore, I carried out the following alternative analysis: Divide all deaths that occurred from 1973 to 1986 into those from the periods 1973–1978 (instead of 1973–1977, to make groups as nearly equal in size as possible), 1979–mid-1982, and mid-1982–1986 and discriminate each of these groups from the survivors. The result was that all correlations were higher for the first period than for the second period and that they dropped little from the second period to the third period. This is exactly the expectation I uttered in my commentary. Hence, the motivation for the "morbidity hypothesis," which had been disproved in the limited test anyway, evaporates, and I herewith wish to retract it.