ELSEVIER

Acta Psychologica 104 (2000) 87–101

acta psychologica

www.elsevier.com/locate/actpsy

The weighing of pathological and non-pathological information in clinical judgment

Yoav Ganzach *

Faculty of Management, Tel Aviv University, 69102 Tel Aviv, Israel

Received 17 May 1999; received in revised form 18 October 1999; accepted 19 October 1999

Abstract

On the basis of the classic data of Meehl (1959), I examine how clinical psychologists use the MMPI scales to judge the degree of pathology of psychiatric patients by comparing linear models of the judgment to a linear model of the criterion (the actual diagnosis of the patients). This comparison reveals that excessively heavy weight is assigned to pathological information in comparison to non-pathological information. Additional analyses reveal that this biased weighing also influences the actual diagnosis and that it is a major determinant of the accuracy of clinical judgment. It is suggested that these effects arise from a confirmation bias associated with the hypothesis that a patient has severe, rather than mild, pathology.

PsycINFO classification: 2340; 3040; 3310

Keywords: Clinical-judgment; Decision-making; Psycho-diagnonsis

1. Introduction

During their career people go through a professional socialization process in which they learn the theories of their field and adopt its metaphysical assumptions. This process defines for the members of the profession the phenomena that are

* Tel.: +11-972-3640-9959; fax: +11-972-3640-9560.
E-mail address: yoavgn@post.tau.ac.il (Y. Ganzach).

0001-6918/00/$ - see front matter
PII: S0001-6918(99)00055-4

worthy of observation, and the type of information that is relevant to their professional judgments. While this process is an integral part of the development of professional skills, it may also lead professionals to base their judgments on perceptions derived from the dominant theories in their field, and ignore, to some extent, important relationships among variables in the environment (DiMaggio & Powell, 1983). This phenomenon has been demonstrated in domains such as managerial decision making (Dearborn & Simon, 1958; Walsh, 1988; Staw, 1981) and scientific research (Kuhn, 1962).

For many years, an important characteristic of the professional socialization process of clinical psychologists has been an emphasis on the psycho-pathological aspects of the mind. The training of clinical psychologists concentrated on issues such as the origins of psycho-pathology, its development, and its remedy; it focused on the diagnosis of the pathological rather than on the identification of the benign; and it centered on the study of the deviant rather than the understanding of the normal. Furthermore, to a certain extent, even normal behavior were often understood by clinical psychologists to be the result of unconscious pathological aspects of the mind, such as murder impulses, incestuous fantasies, and death wishes.

Did a socialization process that emphasize pathology influence the professional judgment of clinicians? A number of studies which have dealt with this question concluded that the answer is positive (e.g., Renaud & Estes, 1961; Rosenham 1972; Langer & Abelson, 1974). Rosenham's (1972) paper entitled 'On being sane in insane places' is the most well-known. In this paper, Rosenham reports that behaviors, which would otherwise appear normal, were judged as pathological by the staff of psychiatric hospitals. However, Rosenham's results could be attributed to initial false information which was supplied to the judges in his study (e.g., Wishner, 1974), or to the high base-rate probability of pathology for inmates of psychiatric hospitals (e.g., Davis, 1979).

The current paper examines the hypothesis that in making clinical judgments, psychologists assigne excessively heavy weight to information regarding the presence or absence of severe pathology (which will be labelled pathological information), in comparison to information regarding the presence or absence of mild pathology (which, within the context of the current study, could be labelled non-pathological information). The paper presents evidence suggesting that such biased weighing influence clinical judgment, and explores the outcomes of this bias. Finally, the cognitive processes underlying this bias are discussed in terms of a confirmation bias in hypothesis testing – the tendency to overemphasize information confirming, rather than disconfirming, expectations; and it is suggested that clinicians' confirmatory hypothesis are associated with the existence of severe, rather than mild, pathology.

The data used in the paper were collected in the mid-1950s by Meehl. The analysis of these data played a major role in the study of the validity of clinical judgments (e.g., Meehl, 1959; Goldberg, 1965, 1970; Dawes & Corrigan, 1974). In these studies, researchers were primarily interested in the *actuarial* validity of the judgment, and in particular, in whether clinical judgments have a higher correlation with the criterion

than the predictions of a linear model. The approach of the current paper to the study of the validity of clinical judgment is different. The paper focuses on analyzing and explaining biases, or systematic deviations, from optimal actuarial validity. As a result, the method used in the paper is also different from the method used in the previous studies. While in previous studies the validity of clinical judgment was studied by correlating judgment with the criterion, in the current study it is studied by comparing models of the judgment to models of the criterion. In particular, the study compares the weight of information associated with severe pathology to the weight of information associated with mild pathology in these models.

2. Method

Meehl's data include 861 MMPI profiles of psychiatric patients – the patients' scores on the 11 most commonly used scales of the MMPI. They also include the criterion – the diagnosis given to the patient in the clinic in which he/she received treatment. 47% of the patients were diagnosed as psychotics and 53% diagnosed as neurotics. These diagnoses were based primarily on information about the patient's past and present behavior, and, to a certain extent, on the results of various psychological tests. For some of the patients, the information on which the diagnosis was based did not include the MMPI profiles, but for some, the MMPI profiles were available when the clinic's diagnosis was made (Meehl, 1959).

The data also include evaluations of the 861 profiles that were made by 29 clinicians, whose schooling represented a wide variety of approaches to Clinical Psychology at the time Meehl's experiment took place (Meehl, personal communication). Each clinician judged the 861 MMPI profiles on an 11-step forced normal distribution scale from least psychotic (1) to most psychotic (11). The clinicians were instructed that the patients could be either psychotics or neurotics. (see Meehl (1959) for a detailed description of the data).

One aspect of the data which is particularly important to the current study is that the MMPI scales of the 861 profiles have a clear dimensional organization. (See Ganzach (1995), for the results of a factor analysis of the scales.) One dimension is associated with the neurotic scales of the MMPI, another with the psychotic scales, and a third with scales that identify defensiveness in test taking. These dimensions, and in particular the neurotic and the psychotic dimensions, were likely to have played an important role in the process by which the clinicians used the MMPI profiles in their diagnostic judgments in Meehl's experiment (Ganzach, 1995).

Because of the high multicollinearity in the data, the independent variables of the models were the dimensions rather than the individual scales:

$$Y = \alpha + \beta_1 F_1 + \beta_2 F_2 + \beta_3 F_3,$$

where F_1 is the unit weight mean of the scales of the neurotic dimension (hypochondriasis, depression, hysteria, and psychoasthenia), F_2 the unit weight mean of the scales of the psychotic dimension (paranoia, schizophrenia, hypomania,

psychopathic deviation, and eccentricity), [1] F_3 the unit weight mean of the scales of the defensiveness dimension (the lie and defensiveness scales), and Y is the dependent variable (e.g., judgment, criterion). For the purpose of comparability between their weights, the three dimensions were standardized prior to the analyses.

The criterion was modeled using a logistic regression, where Y is the log of the odds of having been diagnosed as psychotic by the clinic. The judgment was modeled in three ways. First, by modeling the mean judgment, which was created by averaging the judgments of all 29 clinicians to each profile. Second, by modeling the judgments of each of the 29 judges. And third, by modeling a binary variable, called the *MMPI diagnosis*, created by rank ordering the profiles according to their mean judgment, and labeling the top 47% as having diagnosis of psychosis and the bottom 53% as having a diagnosis of neurosis. [2] The MMPI diagnosis represents the diagnosis that would have been assigned to the patients based on a consensus judgment of the MMPI profiles, and it could be compared to the criterion, or the *clinic's diagnosis*, which was made primarily on the basis of actual behavior. Although some loss of information is involved in using the MMPI diagnosis rather than the untransformed judgment, I discuss the results primarily in terms of this binary variable, because it makes the judgment directly comparable to the criterion. However, the results of the models of the untransformed judgments are reported as well.

Two points regarding the comparison of the judgment models to the criterion model are relevant here. First, the environmentally appropriate *relative* weights of the dimensions in the judgment model should be equal to their relative weights in the criterion model. The reason for this is that relative weights are not affected by the error in the independent variable (Dawes & Corrigan, 1974). Second, when the judgment and the criterion are measured on a binary scale (e.g., when the MMPI diagnosis is compared to the clinic's diagnosis), and when the error variance is larger in the criterion than in the judgment, the optimal weights of the dimensions in the judgment model should be larger than their weights in the criterion model. [3]

Finally, there are two important methodological questions that could be asked. One concerns the justification for using a linear model, given that the true model of the judgment and the true model of the criterion are not known. The justification for using linear models is that they give a good fit to both the criterion and the judgment

[1] Whereas the psychopathic deviation scale and the scale associated with eccentric thoughts and behaviors (the F scale) are usually not regarded as psychotic scales, in the current data they loaded heavily on the psychotic dimension, most likely because the sample consisted only of psychotic and neurotic patients.

[2] From a Bayesian perspective, this is the optimal classification, given the distribution of diagnosis in the sample.

[3] Note that the second requirement is also true if both the criterion and the judgment represent rank ordering in general (and in our case, even the row judgments represent rank ordering). It is not true when the criterion and the judgment represent actual outcome values. In this case, the weights of the predictors in the optimal judgment model should equal their weights in the criterion model.

even if the true model is nonlinear (Dawes & Corrigan, 1974; Goldberg, 1971). [4] Furthermore, linear models are appropriate for psychological inquiry even if they do not directly mimic the process by which the independent variable (e.g., judgment, criterion) is determined (Hoffman, 1960; Goldberg, 1968).

The second question concerns the justification of using the clinic's diagnosis – by itself an erroneous measure of the patient's true state–as a criterion for evaluating clinical judgment. However, an error in the clinic's diagnosis is not a problem for the current analysis. First, a random error in the criterion would result in equal decrease in all three regression weights of the criterion model (Dawes & Corrigan, 1974), but our case for biased weighing is not built on the weights of the criterion model being smaller than the weights of the judgment model. Second, a systematic error in the criterion causing the weights of the criterion model to be *unduly similar* to the weights of the judgment model would suggest that our estimates of the differences between the weights of the two models are conservative in relation to the true differences. Third, a systematic error in the criterion causing the weights of the criterion model to be *unduly different* from the weights of the judgment model is unlikely, since it implies that the clinic's diagnosis, which is based on ample information about the patient's past and present behavior, is less accurate than a diagnosis made solely on the basis of the MMPI. However, this problem cannot be completely ruled out. I will return to it in the discussion section, where it will be evaluated against the results of the analyses.

3. Results

3.1. The model of the judgment vs the model of the criterion

In the analyses reported in this section, the regression weights of the neurotic and the psychotic dimensions in the judgment model are compared to their weights in the criterion model. The hypothesis that excessively heavier weight is assigned to pathological information as compared to non-pathological information implies that, in comparison to the weight of the neurotic dimension, the weight of the psychotic dimension is higher in the judgment model than in the criterion model.

The regression weights of the judgment models and the criterion model are presented in Table 1. Columns 2 and 3 present the weights of the models of the judgment. Column 4 (labelled 'all profiles') presents the weights of a criterion-model which was build using all the 861 profiles (the last two columns will be discussed later). Note that the *dimensions were standardized* before the analyses, so their weights could be compared within each of the models. Note also that the dimensions' weights of the criterion model and the MMPI diagnosis are of logistic regression,

[4] Although it was recently shown that nonlinear models give a better fit to the judgment in Meehl's data than the linear model (Ganzach, 1995), the additional fit supplied by these models was rather minimal, and thus should not decrease the descriptive usefulness of the linear model.

Table 1
Regression coefficients of the judgement and criterion models[a]

Dimension	Judgement		Criterion (Clinc's diagnosis)		
	Mean judgment	MIMPI diagnosis	All profiles	Contaminated profiles	Uncontaminated profiles
Neurotic dimenstion	−0.38[b]	−0.71[b]	−0.85[b]	−0.96[b]	−0.83[b]
	(0.04)	(0.13)	(0.10)	(0.14)	(0.17)
Psychotic dimension	1.62[b]	3.81[b]	1.07[b]	1.22[b]	0.64[b]
	(0.04)	(0.27)	(0.10)	(0.14)	(0.16)
Defensiveness dimension	−0.04	0.03	0.38[b]	0.40[c]	0.32[d]
	(0.03)	(0.11)	(0.08)	(0.11)	(0.13)

[a] Entries are regression coefficients after the three-dimensions were standardized. For the criterion and the MMPI diagnosis entries are coefficients of logistic regression. Numbers in parenthesis are standard errors.
[b] $P < 0.0001$.
[c] $P < 0.001$.
[d] $P < 0.05$.

and therefore cannot be interpreted as ordinary standardized weights. However, being that the two models are logistic models with standardized independent variables, their weights are comparable. Therefore, the judgment weights are compared to the criterion weights primarily by contrasting the regression weights of the MMPI diagnosis (column 3) to the regression weights of the clinic's diagnosis (column 4).

It is clear from Table 1 that, whereas in the criterion model the weight of the psychotic dimension is only slightly larger than the weight of the neurotic dimension, in the judgment models the weight of the psychotic dimension is much larger than the weight of the neurotic dimension. The coefficients of the neurotic and psychotic dimensions are, respectively, −0.71 and 3.81 in the model of the MMPI diagnosis and −0.85 and 1.07 in the model of the clinic's diagnosis. That is, the weight of the psychotic dimension is 5.4 times larger than the weight of the neurotic dimension in the judgment model, but only 1.3 times larger in the criterion model. [5] These results are consistent with the biased weighing hypothesis. Note also that there is a large discrepancy between the coefficient of the defensiveness dimension in the criterion model, where it is highly positive, and its coefficient in the judgment models, where it

[5] Significance tests for the difference between the weights in the judgment were: $F(1.857) = 1207, P < 0.0001$ with regard to the mean judgment, and $\chi^2(1) = 183.7, P < 0.0001$ with regard to the MMPI diagnosis. Although the difference between the weights in the criterion model was also significant, $\chi^2(1) = 6.0, P < 0.05$, as will be discussed later, when the problem of criterion contamination is controlled, the difference between the two weights is non-significant (and the weight of F_1 is even larger than the weight of F_2). Note that in these tests, the independent variables were rescaled to have the same sign (i.e., by multiplying the F_1 by −1).

is very close to zero. Since the defensiveness dimension is related to deception in test taking, it appears that deception, being positively related to pathology in our sample, is effectively ignored in the judgment (see footnote 7).

Perhaps the most interesting comparison between the judgment and the criterion is the one in which the weights of the neurotic and psychotic dimensions in the MMPI diagnosis are directly compared to their weights in the clinic's diagnosis. This comparison reveals that whereas the weight of the psychotic dimension is much higher in the judgment than in the criterion, the weight of the neurotic dimension is higher in the criterion than in the judgment. These data are more consistent with underweighing of non-pathological information than with over-weighing of pathological information, since the optimal dimensions' weights in the judgment model should be larger than their weights of the criterion model (the R^2 of the models are 0.70 and 0.23, respectively; see Nagelkerke (1991), for a discussion of R^2 in logistic regression models). Thus, it is likely that the locus of the excess weight assigned to pathological information over non-pathological information is in the underweighing of non-pathological information.

3.2. Individual judgment strategies and accuracy

The focus of the previous section was on aggregate measures of judges' strategy. However, there are variations in individual strategies which are relevant to the issues under consideration. Table 2 presents the OLS regression weights of a model in which each clinician's judgments of the degree of psychosis were regressed on the three (standardized) dimensions. It is clear from this table that, whereas there is a consensus among the 29 clinicians that the effect of the psychotic dimension on the likelihood of psychosis is positive, there is no consensus that (keeping the other dimensions constant), the effect of the neurotic dimension is negative. The regression slope of the neurotic dimension is significantly negative for 20 of the judges, and significantly positive for six of them, whereas the regression slope of the psychotic dimension is significantly positive for all the 29 judges (Table 2 also presents the significance levels for testing the null hypothesis that the individual regression coefficients are equal to zero). Thus, quite a large percentage of the judges do not recognize the appropriate direction of the relationship between the neurotic information and the criterion. These findings suggest that one reason for the underweighing of non-pathological information observed in the *aggregate* measures is confusion about the appropriate sign of the environmental relationship. [6] Note, however, that underweighing of the neurotic dimension also occurs for the judges who correctly perceived the effect of the neurotic dimension; for all of them the

[6] One way by which this confusion can occur is that the perceived relationship between the neurotic dimension and the likelihood of severe pathology is determined not by the environmental relationship, but by the relationship between the neurotic dimension and the pathological information (i.e., the psychotic dimension) in the MMPI. Since this last relationship is positive (the correlation between the neurotic and psychotic dimensions is 0.51), the perceived relationship between the neurotic dimension and the likelihood of psychosis may be positive.

Table 2
Individual weights and accuracy[a]

Judge	Neurotic dimension	Paychotic dimension	Defensiveness dimension	Accuracy
1	−0.81[b]	1.60[b]	−0.19[c]	0.33
2	−0.75[b]	1.97[b]	0.06	0.30
3	−0.40[b]	1.60[b]	0.16[d]	0.30
4	−0.84[b]	1.74[b]	0.19[c]	0.31
5	−0.40[b]	1.70[b]	−0.03	0.33
6	−0.61[b]	1.72[b]	−0.24[b]	0.33
7	−0.61[b]	1.72[b]	0.12	0.35
8	−0.26[b]	1.70[b]	0.09	0.26
9	−0.05	1.62[b]	−0.25[b]	0.23
10	0.52[b]	1.16[b]	−0.05	0.17
11	0.24[b]	1.51[b]	−0.20[b]	0.22
12	−0.59[b]	1.63[b]	−0.11[d]	0.29
13	−0.30[b]	1.59[b]	−0.07	0.26
14	0.15[d]	0.86[b]	−0.03	0.15
15	−1.05[b]	1.91[b]	0.18[c]	0.39
16	−0.53[b]	1.82[b]	−0.24[b]	0.31
17	−1.35[b]	1.66[b]	0.10[d]	0.36
18	−0.88[b]	1.91[b]	0.13[d]	0.38
19	−0.24[b]	1.68[b]	−0.07	0.28
20	0.28[b]	1.44[b]	0.52[b]	0.24
21	0.50[b]	1.35[b]	−0.25[b]	0.14
22	−0.09	1.65[b]	−0.22[b]	0.25
23	−0.59[b]	1.58[b]	−0.04	0.30
24	−0.56[b]	1.58[b]	0.09	0.31
25	0.19[c]	1.53[b]	−0.12[d]	0.21
26	−0.59[b]	1.73[b]	0.08	0.32
27	0.06	1.40[b]	0.27[b]	0.22
28	−0.81[b]	1.89[b]	−0.04	0.36
29	−0.56[b]	1.70[b]	−0.30[b]	0.29

[a] The weights are of OLS regression in which each clinician's judgment of the degree of pathology was regressed on the three standardized dimensions.
[b] $P < 0.0001$.
[c] $P < 0.001$.
[d] $P < 0.05$.

weight of the neurotic dimension is still substantially smaller than the weight of the psychotic dimension. [7]

The difficulties associated with the appropriate utilization of the non-pathological information is a major determinant of inter-judge differences in accuracy. The accuracy of each of the judges – the correlation between the 861 individual judgments

[7] Such lack of consensus exists also with regard to the defensiveness dimension. The regression slope of F_3 is significantly positive for 8 of the judges an significantly negative for 10 of them. Note that this lack of consensus is the reason that the aggregate weight of F_3 in the judgment model is close to zero.

and the criterion – is presented in the fifth column of Table 2. Across the 29 judges, the correlation between the dimensions' weights and accuracy was $-0.92, +0.81$, and $+0.31$ for the neurotic, psychotic and defensiveness dimensions, respectively. Whereas these high correlations are of interest in themselves – they demonstrate the efficiency of judgment strategies that resemble linear weighing [8] – our interest here is on the fact that the neurotic dimension has a larger impact on accuracy than the psychotic dimension. The difference between their (absolute) correlations with accuracy is significant, $t(26) = 2.8,\ P < 0.01$ in testing for a difference between dependent correlations. Thus, although the impact of the psychotic dimension on judgment is substantially larger than the impact of the neurotic dimension, the appropriate use of the neurotic dimension is a more important determinant of accuracy. Note, however, that the psychotic dimension still predict differential accuracy quite well, despite the overweighing associated with it.

3.3. A comparison between contaminated and uncontaminated profiles

The above analyses represent a conservative estimation of biased weighing because some of the profiles were contaminated – the test scores of the MMPI were available to the clinic's staff when they made the diagnosis. Such contamination is likely to make any observed biases smaller than they really are. To study the effect of contamination, I divided the profiles into contaminated and uncontaminated profiles based on Goldberg's (1965) classification of the profiles in Meehl's data (the 92 profiles for which Goldberg could not determine the degree of contamination were excluded from the analysis). Columns 6 and 7 of Table 1 present, respectively, the dimension weights of the *criterion* model in the contaminated and uncontaminated profiles. It is evident from the data that the weight of pathological information is higher in the contaminated than in the uncontaminated profiles. In the contaminated profiles, the weight of the psychotic dimension is significantly larger than the weight of the neurotic dimension ($\chi^2(1) = 6.7,\ P < 0.01$), whereas in the uncontaminated profiles their weights is about equal (and the weight of the neurotic dimension is even larger than the weight of the psychotic dimension).

The differences between the dimensions' weights in the contaminated and uncontaminated profiles suggest that biased weighing occurred not only in the experiment, but also in the real world. When the test results of the MMPI were available to the clinic's staff, they ignored the non-pathological information (the weight of the neurotic dimension in the two types of profiles is quite similar), but took into account the pathological information (the weight of the psychotic dimension in the contaminated profiles is higher than its weight in the uncontaminated profiles). To test this effect, I examined a criterion model of all the 769 profiles whose contamination status was known, which included the three dimensions, the contamination

[8] Interestingly enough, neither Goldberg (1970) nor Einhorn (1974) found a relationship between the multiple correlation of the cues and the judgment – which could be viewed as a measure of judge's reliance on linear strategy – and accuracy.

status (a binary variable), and the interaction between the dimensions and the contamination status. The null hypothesis that the weights are equal in the two types of profiles (i.e., the null hypothesis that there is no interaction between weight and contamination status) was rejected with regard to the psychotic dimension, $\chi^2(1) = 7.6,\ P < 0.006$; but it was not rejected with regard to the neurotic dimension, $\chi^2(1) = 0.4,\ P > 0.5$ (nor was it rejected with regard to the defensiveness dimension, $\chi^2(1) = 0.3,\ P > 0.6$).

4. Discussion

The current study showed that, in clinical judgment, excessively heavier weight has been assigned to pathological information in comparison to non-pathological information, and suggested that this biased weighing is associated with underweighing of non-pathological information. In diagnosing psychotics from neurotics, clinicians relied primarily on information associated with psychosis, and ignored information associated with neurosis. This diagnostic strategy is not justified, since, with regard to the criterion, the importance of these two types of information is about the same. This underweighing of non-pathological information influenced not only the judgments in a controlled experimental situation (i.e., the judgments elicited in Meehl's experiment). It also influenced real-life diagnoses, those given in the clinics. When the MMPI results were available, the staff of the clinics incorporated into their diagnoses the test results which were associated with severe pathology, and ignored the results associated with mild pathology.

The underweighing of non-pathological information could be viewed as an instance of *confirmatory judgment strategy*. It could be argued that in processing clinical information, clinicians attempted to confirm the hypothesis that the patient is highly pathological; and that the underweighing of non-pathological information is associated with the tendency to disregard disconfirming information (e.g., Fischhoff & Beyth Marom, 1983; Klayman & Ha, 1987). Indeed, in a review of the relevant literature, Turk and Salovey (1985) concluded that "in the clinical context, what the clinicians expect to observe is pathology. This expectancy may ... introduce systematic biases in the direction of overestimating psychopathology and underestimating more positive features" (p. 24).

While confirmatory strategies are usually demonstrated in the context of information search (e.g., Wason, 1960; Snyder & Campbell, 1980), a number of studies have demonstrated confirmatory *judgment* strategies. These strategies can be divided into two types. One type is *configural confirmatory strategies*, in which the confirmatory hypothesis leads to attribute weights which depend on attribute values. Thus, for example, when the confirmatory hypothesis is that a rental-apartment is suitable, the more positive attributes receive relatively higher weight in the evaluation of the apartment, whereas when the confirmatory hypothesis is that the apartment is unsuitable, the more negative attributes receive higher weight (Ganzach, 1993). Another example is Birnbaum and Stegners (1979) demonstration of the effect buyers' and sellers' 'point of view' in judgments of the fair price of a product:

with regard to sellers, the more favorable the product information, the higher its weight, whereas with regard to buyers, the more favorable the information, the lower its weight.

A second type of confirmatory judgment strategies is *non-configural confirmatory strategies*, in which the information is organized in dimensions or categories whose weight depends on the confirmatory hypothesis. Examples are Tversky's (1977) demonstration that the weight of common features is relatively high in similarity judgment whereas the weight of distinctive features is relatively high in dissimilarity judgment, and Shafir's (1993; see also Ganzach & Schul, 1995) demonstration that the weight of positive features is relatively high in accept decisions whereas the weight of negative information is relatively high in reject decisions. [9]

The underweighing of non-pathological information that emerges from the analyses presented in the current paper is a non-configural confirmatory strategy, since it does not involve value-dependent weighing. However, elsewhere I have demonstrated that the judgments in Meehl's experiment also involve configural confirmatory strategies. In integrating the neurotic scales to arrive at an overall evaluation of the neurotic dimension, the most neurotic information tends to receive the highest weight, whereas in integrating the psychotic scales to arrive at an overall evaluation of the psychotic dimension, the most psychotic information tends to receive the highest weight (Ganzach, 1995). Fig. 1 presents a schematic representation of a process by which three neurotic and three psychotic scales are integrated to arrive at an overall judgment of the likelihood of psychosis. In Fig. 1, P_3 and N_3 are respectively the *highest*-value psychotic and neurotic scales, and the thick arrows indicate relatively heavier weight in the integration process. Fig. 1 shows that there are two stages in the integration process. The first, the integration of the dimensions' scales, involves configural confirmatory strategy; the second, the integration of the dimensions, involves a non-configural confirmatory strategy. Note, however, that whereas the second stage involves bias in weighing, it is not clear if the first stage involves such a bias (Ganzach, 1998).

Finally, although the discussion of biased weighing relies primarily on the concept of confirmatory judgment strategy, it should be noted that this concept does not give a complete account of our data, since it does not explain why the confirmatory hypothesis is associated with severe pathology (psychosis) and not with mild pathology (neurosis). To explain this, a framework which identifies the confirmatory hypothesis is necessary. I propose that dominant professional theories, and professional socialization processes which lead to adopting these theories, provide such a framework. Note also that Clinical Psychology may not be the only domain in which the confirmatory hypothesis is associated with pathology. For example, in describing

[9] Note that these examples of non-configural confirmatory strategies, as well as other examples discussed in the literature (e.g., Lehman Krosnick, West & Li, 1992), involve features, rather than dimensions, as input for the judgment (in fact, configural strategies cannot occur when the input information is in the form of features). Unlike these examples, the current study involves non-configural confirmatory strategy when the input information is dimensional.

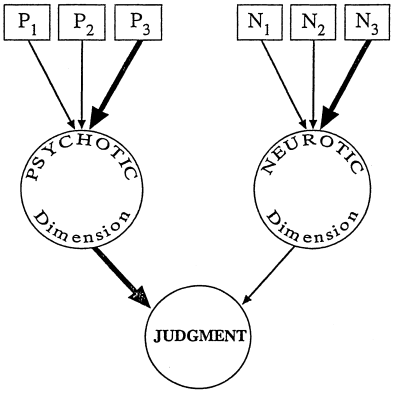

Fig. 1. A schematic representation of a process by which three neurotic and three psychotic scales are integrated to arrive at an overall judgment of the likelihood of psychosis.

physicians frame of mind, John Donne writes ("The First Anniversarie – An Anatomy of the World", lines 91–92):

There is no health, Physicians say that we at least, enjoy, but a neutrality.

A somewhat different explanation for the findings is based on the compatibility principle, which suggests that "the weight of any input component is enhanced with its compatibility with the output" (Tversky, Sattath & Slovic, 1988, p.376). According to this explanation, the psychotic dimension received excessively higher weight in comparison to the neurotic dimension because it was compatible with the response scale, which was phrased in terms of the likelihood of psychosis. For example Schul and Ganzach (1995) showed that positive features weigh more when the judgment scale is an attractiveness scale (e.g., how good is the stimulus) than when it is an unattractiveness scale (e.g., how bad is the stimulus). Note, however, that the response-scale explanation is not independent of the confirmatory bias explanation. It is likely that Meehl used a scale in which responses were given in terms of the degree of psychosis, because this scale is more natural than a scale in which responses are given in terms of the degree of neurosis. But the former scale is more natural because in many diagnostic situations, the presence of psychosis is of more interest to clinicians than the presence of neurosis. [10] Note also that a response-scale explanation cannot account for some of the findings reported in this section, such as the difference between the contaminated and uncontaminated profiles in the criterion model.

[10] It is worthwhile to distinguish here between experimenter-induced frame and theory induced frame. In Meehl's experiment, the two are confounded. However,it is possible to unconfound them by comparing the model of the criterion to models of judgments which were generated in different framing conditions. For example, by comparing a condition in which judgment was framed in terms of the extent of severe pathology (e.g., the degree of psychosis), to a condition in which judgment was framed in terms of the extent of mild pathology (e.g., the degree of neurosis).

It was noted in the introduction that an alternative explanation suggesting that biased weighing is due to an error causing the criterion to be unduly similar to the judgment cannot be ruled out completely. However, in view of the results, this explanation is highly unlikely. First, such an explanation would suggest that the locus of the bias is that, in the clinic, pathological information is underweighed relative to non-pathological information. Given the literature about clinicians tendency to emphasize pathological information, this is theoretically unappealing. Second, the fact that the MMPI was developed as a criterion keyed inventory, and that the discriminative ability of its scales is about the same, is consistent with the similar weights of the neurotic and psychotic dimensions in the criterion model, but not with their dissimilar weights in the judgment models.

The area of decision making has seen a proliferation of research on biases in judgment and choice (see for example, Kahneman, Slovic & Tversky, 1982). Almost all work concerning these biases used one of two paradigms. In one paradigm, decisions in one condition are compared to decisions in another, normatively equivalent, condition, and biases are identified as the gaps between decision output in the two conditions (e.g., Lichtenstein & Slovic, 1971). In the other paradigm, subjects are presented with problems for which normative solutions exist, and biases are identified as the systematic portions of the gaps between subjects' responses and normative responses (e.g., Kahneman & Tversky, 1972). From this perspective, the current study belongs to the second paradigm. However, whereas in this paradigm the normative responses are derived from some normative theory about rational behavior, in the current study the normative responses are derived from real-life criterion.

Finally, it is interesting to note how the current work stands with regard to the two main traditions in judgment and decision making research, the heuristic and biases tradition, and the functionalist (Brunswikian) tradition (see Hammond, 1990, for a discussion of this distinction). On the one hand, the current study is performed within the functionalist approach of studying expert judgment which emphasize non-orthogonal design, representative stimulus, redundancy of information, and measures of achievement (see, Stewart, 1997 for a recent example). On the other hand, this study focus on internal cognitive processes which typify the heuristics and biases approach. Whereas a number of studies conducted in the functionalist tradition did attempt to examine such cognitive processes (e.g., Adelman, 1981; Gaeth & Shanteau, 1984; Brehmer, 1973, 1974), to achieve this purpose they often tended to sacrifice the representativeness of the design (e.g., most of them did not rely on real-life cues or criteria). The current study use Meehl's unusual database to overcome this difficulty. It attempts to uncover cognitive processes underlying expert judgment as they occur in the natural environment.

Acknowledgements

I would like to thank Dr. Lewis R. Goldberg for his help, both in supplying the data and in answering many of my questions. I would like to thank Robyn Dawes,

Nimrod Kleinman, Avi Kluger, David Krantz, Dany Mingelgrin Yaacov Schul and Robert Griphin for their comments on earlier versions of this paper. Financial support was provided by the Kurt Lion foundation and by the Israel Institute for Business Research.

References

Adelman, L. (1981). Decisions and design. *Organizational Behavior and Human Decision Processes*, *27*, 423–472.

Birnbaum, N. H., & Stegner, S. E. (1979). Source credibility in social judgment: bias, expertise, and the judges point of view. *Journal of Personality and Social Psychology*, *37*, 48–74.

Brehmer, B. (1973). Single cue probability learning as a function of the sign and magnitude of the correlation between cue and criterion. *Organizational Behavior and Human Performance*, *9*, 377–395.

Brehmer, B. (1974). Hypothesis about the relations between scaled variables in the learning of probabilistic inference. *Organizational Behavior and Human Performance*, *11*, 1–27.

Davis, D. A. (1979). What's in a name? A Bayesian rethinking of attributional biases in clinical judgment. *Journal of Consulting and Clinical Psychology*, *47*, 1109–1114.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin*, *81*, 95–106.

Dearborn, D. C., & Simon, H. A. (1958). Selective perception: a note on departmental identification of executives. *Sociometry*, *21*, 140–144.

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, *48*, 147–160.

Einhorn, H. (1974). Cue definition and residual Judgement. *Organizational Behavior and Human Decision Processes*, *12*, 30–49.

Fischhoff, B., & Beyth-Marom, R. (1983). Hypothesis evaluation from a Bayesian perspective. *Psychological Review*, *90*, 239–260.

Gaeth, G. J., & Shanteau, J. (1984). Reducing the influence of irrelevant information on experienced decision makers. *Organizational Behavior and Human Decision Processes*, *33*, 233–282.

Ganzach, Y. (1993). Goals as determinants of nonlinear, noncompensatory judgment strategies. *Organizational Behavior and Human Decision Processes*, *56*, 422–440.

Ganzach, Y. (1995). Nonlinear models of clinical judgment: Meehl's data revisited. *Psychological Bulletin*, *118*, 422–429.

Ganzach, Y. (1998). Nonlinear models in decision making: the diagnosis of psychosis versus neurosis from the MMPI. *Organizational Behavior and Human Decision processes*, *74*, 53–61.

Ganzach, Y., & Schul, Y. (1995). The influence of quantity of information and valence framing on decision. *Acta Psychologica*, *89*, 23–36.

Goldberg, L. R. (1965). Diagnosticians vs diagnostic signs: the diagnosis of psychosis vs neurosis from the MMPI. *Psychological Monographs*, *79*.

Goldberg, L. R. (1968). Simple models or simple processes: some research on clinical judgment. *American Psychologist*, *23*, 483–496.

Goldberg, L. R. (1970). Man versus model of man: a rationale plus some evidence for improving clinical inferences. *Psychological Bulletin*, *73*, 422–434.

Goldberg, L. R. (1971). Five models of clinical judgment: an empirical comparison between linear and nonlinear representation of the human inference process. *Organizational Behavior and Human Performance*, *6*, 458–479.

Hammond, K. R. (1990). Functionalism and illusionism: Can integration be usefully achieved? In R. B. Hogarth, Insights in decision making. Chicago: The University of Chicago Press.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin*, *57*, 116–131.

Kahneman, D., & Tversky, A. (1972). Subjective probability: a judgment of representativeness. *Cognitive Psychology*, *3*, 430–454.

Kahneman, D., Slovic, P., & Tversky, A. (1982). Judgment under uncertainty: Heuristics and biases. London: Cambridge University Press.

Klayman, J., & Ha, Y. W. (1987). Confirmation, disconfirmation, and information in hypothesis testing. *Psychological Review*, *94*, 211–228.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: The University of Chicago Press.

Langer, E. J., & Abelson, R. P. (1974). A patient by any other name...: clinical group differences in labelling bias. *Journal of Consulting and Clinical Psychology*, *42*, 4–9.

Lehman, R. L., Krosnick, J. A., West, R. L., & Fan, L. (1992). The focus of judgment effect: a question wording effect due to hypothesis confirmation bias. *Personalis and Social Psychology Bulletin*, *18*, 690–699.

Lichtenstein, S., & Slovic, P. (1971). Reversal of preference between bids and choices in gambling decisions. *Journal of Experimental Psychology*, *89*, 46–55.

Meehl, P. (1959). A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles. *Journal of Counseling Psychology*, *6*, 102–109.

Nagelkerke, N. J. D. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*, 691–692.

Renaud, H., & Estes, F. (1961). Life history interview with one hundred normal America males: pathogenicity of children. *American Journal of Orthopsychiartly*, *31*, 786–802.

Rosenham, D. L. (1972). On being sane in insane places. *Science*, *179*, 250–258.

Schul, Y., & Ganzach, Y. (1995). The effects of accessibility of standards and decision framing on product evaluation. *Journal of Consumer Psychology*, *4*, 61–83.

Shafir, E. (1993). Choosing versus rejecting: why some options are both better and worse than others. *Memory and Cognition*, *21*, 546–556.

Snyder, M., & Campbell, B. H. (1980). Testing hypotheses about other people: the role of the hypothesis. *Personality and Social Psychology Bulletin*, *6*, 421–426.

Staw, B. (1981). The escalation of commitment to a course of action. *Academy of Management Review*, *6*, 577–587.

Stewart, T. R. (1997). The importance of the task in analyzing expert judgment. *Organizational Behavior and Human Decision processes*, *69*, 205–219.

Turk, D. C., & Salovey, P. (1985). Cognitive structures cognitive processes and cognitive behavior modification: II. Judgments and inferences of clinicians. *Cognitive Therapy*, *9*, 19–33.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.

Tversky, A., Sattath, S., & Slovic, P. (1988). Contingent weighing in judgment and choice. *Psychological Review*, *95*, 371–384.

Walsh, J. P. (1988). Selectivity and selective perception: an investigation of managers' belief structures and information processing. *Academy of Management Journal*, *31*, 873–896.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*, 129–140.

Wishner, J. (1974). Psychopathology: defective concepts or defective practice? In Proceedings of the 18th International Congress of Applied Psychology, Montreal, Canada (Invited address).