The Dilution Effect: Nondiagnostic Information Weakens the Implications of Diagnostic Information

RICHARD E. NISBETT

The University of Michigan

HENRY ZUKIER

The New School for Social Research

AND

RONALD E. LEMLEY

The University of Michigan

In a series of studies, subjects were asked to make predictions about target individuals. Some subjects were given information about the target which pretest subjects had judged to be "diagnostic"—that is, had judged to be usefully predictive of the outcome. Other subjects were given a mix of information judged to be diagnostic and information judged to be "nondiagnostic" by pretest subjects— that is, judged to be of little value for predicting the outcome. Subjects given mixed information made much less extreme predictions than did subjects given only diagnostic information. It was argued that this "dilution effect" occurs because people make predictions by making simple similarity judgments. That is, they compare the information they have about the target with their conception of outcome categories. The presence of individuating but nondiagnostic information about the target reduces the similarity between the target and those outcomes that are suggested by the diagnostic information. One of the major implications is that stereotypes and other "social knowledge structures" may be applied primarily to abstract, undifferentiated individuals and groups and may be largely set aside when judgments are made about concrete, individuated people.

The accumulating evidence on the accuracy of people's predictions has led to some striking observations. Even expert judges make predictions that are substantially less accurate than those obtained from optimal actuarial formulas employing the same information (Meehl, 1954; Goldberg, 1968). Expert judges are less accurate, in fact, than a simple linear equation modeling the judge's own weighting system and they are less accurate

248

even than equations in which beta weights are randomly assigned (Dawes & Corrigan, 1974). Research by Kahneman and Tversky (1973) suggests that predictions are suboptimal in part because people rely on the "representativeness heuristic." Features of the target are simply compared to features of possible outcomes, and the outcome that is most representative of the target is chosen. Since this strategy ignores the important consideration of relative outcome frequency, predictions are often too extreme, that is, they depart too much from the central tendency of the outcome distribution.

Most prediction research employs the following model: Subjects are given information about a target case which they believe to be *diagnostic* of some outcome, for example, of a trait or of membership in some group or category, and subjects then are asked to make predictions about the degree to which the target possesses the trait or whether the target belongs to the category. For example, clinical psychologists are presented with MMPI profiles and are asked to predict the severity of mental illness of the patient who generated the profile. Or subjects are told that a target person is a German and are asked to predict the target's efficiency. Or subjects are told that a target "shows no interest in political and social issues and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles" (Kahneman & Tversky, 1973, p. 241) and are asked to predict whether the target is an engineer or lawyer.

Though this research has been quite successful in elucidating people's prediction strategies, it does not capture an important element of many real-world prediction tasks. People normally possess not only information that they believe to be diagnostic, but also information that they do not believe to be diagnostic. Thus clinicians in real clinical settings typically possess much more information about patients than just their MMPI scores, and people who meet a real German know much more about the individual than his nationality. Some of this additional information may be regarded as diagnostic of degree of mental illness, or efficiency, but much of it, if not most of it, probably would be regarded by the individual as having little or no diagnostic value.

Previous research therefore has not focused on an important question: How do people combine items of information that they believe to be diagnostic—that is, to be useful for predicting some outcome—with items of information that they believe to be nondiagnostic—that is, to have little or no value for predicting the outcome? Under the simplest conditions, where it can be assumed that the implications of nondiagnostic items do not interact either with each other or with the implications of diagnostic items, it might be expected that predictions about a target would rest only on the diagnostic information, and that the addition of nondiagnostic in-

formation would not affect predictions. There are good theoretical grounds, however, for expecting that people might not behave in that way. One set of theoretical considerations leads to the expectation that nondiagnostic information would *enhance* the extremity of predictions, and another set of considerations leads to the opposite expectation that nondiagnostic information would *dilute* the extremity of predictions.

Enhancement of the extremity of predictions due to the presence of nondiagnostic information might be expected on the basis of prevailing assumptions about the way people process social information. Just as ambiguous information is sometimes interpreted in terms of stereotypes and other knowledge structures (Cantor & Mischel, 1979; Hamilton, 1979; Nisbett & Ross, 1980), nondiagnostic information might be assimilated to, and even made to seem supportive of, the diagnostic information. A clinician might interpret any number of neutral behaviors as being supportive of his diagnosis of a patient as paranoid, and the layperson might render any number of stray facts about a given German consistent with his stereotype about the efficiency of Germans. In addition, if the diagnostic information is obtained *prior* to the nondiagnostic information, then set and primacy effects could play a further role in enhancing the extremity of predictions. The diagnostic information might function as an initial "hypothesis" about the target which could then be "confirmed" by the subsequent nondiagnostic information.

On the other hand, the work of Kahneman and Tversky (1972, 1973) suggests that the presence of nondiagnostic information might result in the very opposite of enhancement. If the information that subjects possess about the target is highly representative of, or similar to, the features of a given outcome, subjects predict with confidence that the target possesses the outcome value or is a member of the outcome category. Judgments of similarity are a positive function of the number of features common to both target and outcome and a negative function of the number of non-common features (Tversky, 1977). Thus the addition of common features increases the similarity and the addition of noncommon features decreases the similarity. In this terminology, diagnostic information about the target is information common to both the target and to the subject's conception of the outcome. For example, the target has little political interest and likes mathematical puzzles; most engineers are believed to have little political interest and to enjoy asocial, scientific hobbies like mathematical puzzles. By contrast, nondiagnostic information is information that characterizes the target but which neither characterizes nor contradicts the individual's conception of the outcome.

It seems likely that most people would regard the great bulk of facts about any given individual as nondiagnostic of almost all outcomes. Such facts as that a target is 5 ft 10 in. tall, has two siblings, is ambitious, or gets

along well with his colleagues, would all be regarded by most people as nondiagnostic for the prediction "engineer" vs "lawyer." For most people, neither the stereotype of an engineer nor that of a lawyer either includes such features or suggests that they should be absent. It might be expected that such individuating and noncommon information (that is, information which characterizes the target but not the outcome) would be ignored in prediction. But if people make predictions on the basis of the similarity between target and outcome, and if noncommon features serve to reduce the perceived similarity, then predictions should become less extreme when nondiagnostic information is available. The effect of nondiagnostic information, then, might be to "dilute" the implications of diagnostic information by reducing the perceived similarity between the target and the outcome.

In all of the studies described below, pretest subjects judged the diagnosticity for some outcome of individual items of information that later were presented to experimental subjects. The predictions of experimental subjects given only information judged by pretest subjects to be diagnostic were compared to those of subjects given a mixture of the same "diagnostic" information and other information judged by pretest subjects to be nondiagnostic. It was anticipated that subjects given the mixed information would make less extreme predictions for the target than subjects given only diagnostic information.

STUDIES 1–3

One of the most important implications of the present analysis is that diagnostic information in the form of group membership labels, for which common cultural stereotypes exist, might be diluted by exposure to small amounts of individuating but nondiagnostic information about group members. In the first series of studies, subjects were asked to make predictions about the behavior of male participants in psychological investigations. Advantage was taken of the stereotypes that exist concerning people who are interested primarily in the humanities vs those who are interested primarily in the sciences. Subjects were asked to predict how much electric shock would be tolerated by a student majoring in the sciences vs a student majoring in the humanities, and also how many movies would be attended by each type of student. It was anticipated that subjects would predict that science majors, because of their greater presumed masculinity and experience with apparatus, electrical and otherwise, would take more shock than humanities majors, and it was anticipated that subjects would predict that humanities majors, because of their greater cultural interests, would attend more movies. Other subjects were given, in addition to group membership labels, information about the students that was selected for its manifest irrelevance to the behaviors to

be predicted. It was anticipated that this additional information would dilute the extremity of predictions based on information about group membership alone.

Study 1

Overview

The experimental conditions were intended to capture three ecologically common prediction tasks: (1) Predictions about a group of people sharing a stereotyped category label; (2) predictions about a single individual described only by the stereotyped category label; and (3) predictions about a single individual described both by the stereotyped category label and by additional information. Two sets of subjects thus received only category label information about the target individuals before making their predictions. In the group version of this diagnostic-information-only condition, subjects were asked to predict shock tolerance for a group of engineering majors and a group of music majors, and to predict movie attendance for a group of premedical students and a group of English majors. In the individual version of this condition, subjects made predictions for an individual described only by first name, last initial, college major, and career plans, for example, "John V., an engineering major planning a career in aerospace engineering." The two versions of the diagnostic-information-only condition make it possible to determine whether subjects make the same extreme predictions for specific individuals who are described with a particular category label as they do for abstract groups of people who are described with that label.

The chief concern was with a comparison of the two diagnostic-information-only groups with a group exposed to brief videotaped interviews with four different individuals. The interviews provided "background information" judged by pretest subjects to be nondiagnostic. A comparison between the predictions in the interview condition and those in the diagnostic-information-only conditions would show whether subjects make more or less extreme predictions for group members when they also possess nondiagnostic information about them.

Procedure

One hundred and eight University of Michigan students of both sexes[1] who were enrolled in introductory psychology participated in the study in groups of 6–12. Subjects were told that the investigators were concerned with the ability of people to predict the behavior of subjects in psychological studies. They were asked to read about two studies which, they were told, had recently been conducted using male students at The University of Michigan. The first study was an experiment on the pain-suppressant properties of the drug Tiborium. Subjects read the protocol of the study provided in Nisbett and Borgida (1975). The protocol

[1] There were no sex differences in any of the studies reported.

stated that participants in the shock study were requested to accept as much electrical shock, applied through finger electrodes, as they could bear. It was emphasized that the average subject had tolerated 529 μA, enough to cause a strong jolt extending through the hand on which the electrodes were placed. The average for all subjects was provided so as to reduce the variance of estimates. The value 529 μA was not chosen arbitrarily, however. It was the mean value that pretest subjects estimated would be tolerated by a random sample of University of Michigan males.

The second study was a survey of attendance at cultural events at the university and in the community. One of the questions concerned the number of movies attended over the course of the semester. It was emphasized that the average number of movies attended was 12 (the mean value estimated by pretest subjects).

Upon arrival at the laboratory, subjects were randomly assigned to one of the diagnostic-information-only conditions or to the videotaped interview condition.

Diagnostic-Information-Only Conditions

Some subjects were asked to estimate the amount of shock taken and the number of movies attended after being given information only about the majors and career plans of participants in the studies. The cover sheet for these subjects stated that the investigators were interested in knowing "what kinds of information are useful to people" in making judgments of the type requested. "We will ask you to make judgments about the behavior of subjects" after being given "only a small amount of information" about them. "You may or may not find the information you are given to be useful in making your guesses."

There were two versions of the condition. In the *group version,* 24 subjects were told that 9 of the participants in the shock experiment had been engineering majors and 4 had been music majors. They were asked to estimate how much shock each group had tolerated on the average. After reading about the cultural-events survey, they were told that 16 of the respondents had been premedical students and 11 had been English majors. They were asked to estimate how many movies had been attended by each group.

In the *individual version,* 24 subjects were asked to estimate how much shock was tolerated by particular participants: "John V., an engineering major planning a career in aerospace engineering," and "Allen H., a music major planning to be either a professional musician in an orchestra or band or a teacher of music." After reading about the survey, they were asked to estimate how many movies would have been attended by "Tom W., a premedical student majoring in biology," and by "Bill R., an English major planning to go to law school."

Diagnostic plus Nondiagnostic Information: Videotaped
Interview Condition

Sixty subjects were asked to estimate shock tolerance and movie attendance for people whom they saw in videotaped interviews of about 2-min duration. The interviews allegedly had been conducted with participants

in the studies. In fact, the interviewees were undergraduate acquaintances of the investigators who happened to be majoring in appropriate fields for the purposes of the study: engineering, music, biology, and English.

Each interviewee provided two interviews. In one, the interviewee answered every question truthfully. In the other, every question except the ones about major and career plans were answered truthfully: Science majors posed as humanities majors and humanities majors posed as science majors. Thus the engineering major provided one interview in which he identified himself correctly as an engineer planning a career in aerospace engineering and one interview in which he identified himself as a music major planning a career as a member of an orchestra or a teacher of music. Half the subjects saw interviews in which all interviewees correctly identified their majors and half saw interviews in which science majors posed as humanities majors and vice versa.

The questions asked in the interview were intended to provide biographical information that would have little or no diagnostic value for predicting either shock tolerance or movie attendance. (This attempt was successful, as will be seen below.) The question asked, and the answers provided by the engineering major, were as follows:

Q. What is your name?
A. John Varner.
Q. What is your year in school?
A. I'm a sophomore.
Q. What is your major?
A. Engineering.
Q. What do you plan to do for a career?
A. I'll go into some facet of aerospace engineering.
Q. What is your grade point average?
A. 3.1.
Q. Where are you from originally?
A. Detroit.
Q. What is your father's occupation?
A. He's a sales manager for a steel company.
Q. What is your mother's occupation?
A. She's a housewife.
Q. What is your religious background?
A. Catholic.
Q. Did you consider going to any other schools besides the University of Michigan?
A. I actually went to Western Michigan for a semester and then transferred here.
Q. Can you tell me something about your spare time activities . . . sports, hobbies?
A. I like to snowski, and play chess and read and listen to music.
Q. What kind of reading do you do for pleasure?
A. Mostly science fiction and philosophical literature.
Q. Do you happen to remember the name of the last book you read?
A. Uh . . . "Cat's Cradle," by Kurt Vonnegut.

During most of the interview, the interviewee was shown in a head-and-shoulders close-up. Subjects were seated 6 ft (1.82 m) from a 19-in. (0.48-m) television monitor.

Dependent Measures

Subjects gave their estimates of the shock-taking behavior of the engineering major(s) and music major(s) on the scale below.

Microamperes of Shock Tolerated

0	50	100	250	500	1000	2000	3000
Took no shock at all	Tickling sensation in fingers	Tingling sensation in fingers	Small jolt felt through hand	Strong jolt to hand	Jolt causing hand and wrist to jerk	Jolt causing forearm to jerk	Jolt causing entire arm to jerk

Subjects gave their estimates of the number of movies attended by the premedical student(s) and English major(s) on the scale below.

Number of Movies Seen

0   5   10   15   20   25   30   35   40   50   60   70   80   90   100

Pretest for Diagnosticity of Information on Videotapes

Prior to conducting Study 1, all items of interview information on the videotapes were presented singly to pretest subjects. Pretest subjects read the same protocols presented to subjects in Study 1 and then were asked to indicate for each item "how helpful the information would be for a prediction." For the shock study, subjects were asked, "If you think the information indicates that the person took *more shock* than average, place a plus (+) in the blank provided." "If you think the information indicates that the person took *less shock* than average, place a minus (−) in the blank." "If you think the information *would not be very helpful* in making a prediction about whether the subject took more or less shock than the average, place a zero (0) in the blank."

Major and career plan information were presented as a single item, e.g., "major is engineering and he plans a career in aerospace engineering." All other items were presented as narrative versions of a single question and answer, e.g., "father is a sales manager for a steel company." Items for the science major and the humanities major were combined into a pool with filler items and presented to 16 of the subjects in one random order and to 15 of the subjects in another random order.

For the shock study, 81% of pretest subjects thought that an engineering major was indicative of high shock tolerance and 74% thought a music major was indicative of low shock tolerance. For the movie study, 81% of

the subjects thought an English major was indicative of going to a large number of movies and 77% thought a premedical career was indicative of going to a small number of movies.

As anticipated, none of the remaining items was viewed as diagnostic by a majority of the subjects, and few were regarded as diagnostic by even a substantial fraction of the subjects. All items that were viewed as diagnostic by more than 4 of the 31 pretest subjects (13%) were deleted from the videotaped interview. On the average, the items retained on the videotape (except for major and career plans) were regarded as nondiagnostic by 94% of pretest subjects.

Results

If people make predictions by judging the similarity between target information and outcome features, then subjects in both the group and individual versions of the diagnostic-information-only conditions would be expected to make more extreme predictions than subjects in the videotaped-interview condition. The "high-shock-tolerance" outcome, for example, would be expected to include the features "masculine," and "acquainted with apparatus." Both engineers as a group, and "John V., an engineering major planning a career in aerospace engineering," would be expected to share these features with the high-shock-tolerance outcome. "John V.," the participant seen on the videotape, also shares these features, but in addition has features, e.g., "Catholic," "3.1 grade point average," not included in the outcome. These additional features should reduce the similarity between the target and the high-shock outcome.

Figure 1 presents the differential estimates of subjects in all three conditions. The Y axis of Fig. 1A graphs mean estimated shock tolerance of engineers minus mean shock tolerance of music majors, corrected by dividing by the sum of both scores in order to reduce variance due to individual differences in mere magnitude of estimates. Figure 1B presents the similarly corrected mean differential estimates for movie attendance by English majors and premedical students.

It may be seen that differential estimates of shock tolerance were substantially greater for subjects with knowledge only about the major and career plans of targets than for subjects with access also to the information provided in the videotaped interviews. Both the group version of the diagnostic-information-only condition and the individual version of that condition produced sharply different estimates for the two types of major, and the two versions differed only trivially from each other ($t < 1$). The t value contrasting the combined diagnostic-information-only conditions with the videotaped interview conditions was 2.61 ($df = 106, p < .01$).[2]

---

[2] All p values reported are based on two-tailed tests.

A. Shock Study

Engineer Shock - Music Major Shock / Total Shock

40 — .30 — .20 — 10 — 0

Diagnostic Information Only-Group Version | Diagnostic Information Only-Indiv- idual Version | Video- Taped Interview

B. Movie Study

English Major Movies - Premedical Student Movies / Total Movies

.40 — .30 — .20 — .10 — 0

Diagnostic Information Only-Group Version | Diagnostic Information Only-Indiv- idual Version | Video- Taped Interview
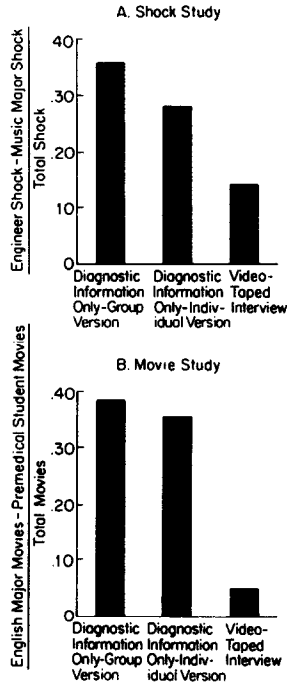
FIG. 1. Differential predictions of shock tolerance and of movie attendance as a function of nature of information about the target.

The pattern is the same for estimates of movie attendance. Both the group version of the diagnostic-information-only condition and the individual version produced sharply different estimates of movie attendance for English majors vs premedical students, and the two versions differed only trivially from each other ($t < 1$). The combined diagnostic-information-only conditions differed from the videotape condition at the .001 level of significance; $t(106) = 5.30$.

This first, demonstrational study shows that subjects make very different predictions about science majors and humanities majors than they do about particular, individuated science majors and humanities majors. "Engineers" take a great deal of shock and "John V., an engineer planning a career in aerospace engineering," takes a great deal of shock, but a living, breathing engineer named John Varner, a Catholic sophomore from Detroit with a GPA of 3.1, etc., takes only a little more shock than average. This finding is consistent with the hypothesis that information without diagnostic value serves to dilute predictions based on highly diagnostic information. Before such an interpretation can be accepted, however, several alternative explanations must be dealt with.

(1) The physical appearance and mannerisms of the interviewee may have contradicted the stereotype for his major. This might have happened, for example, if the investigators had chosen a particularly effeminate engineer and a particularly masculine-appearing music major. It will be recalled, however, that the design of the study makes it possible to rule out such a trivializing explanation. Half of the subjects saw a real engineer and a real music major, while the other half of the subjects saw the music major posing as an engineer and the engineer posing as a music major. If the real music major had seemed physically the sort who would have tolerated a great deal of shock, then the real music major should have been expected to tolerate more shock than the real engineer, regardless of the major claimed by either. In fact, however, the "music major" was guessed to have tolerated an almost identical amount of shock whether it was really the music major whom subjects saw or the engineer claiming to be a music major ($t < 1$), and the "engineer" was similarly guessed to have tolerated an almost identical amount whether he was or was not the real engineer ($t < 1$). The same lack of any effect for real vs pretended major was found for the movie study.[3]

There is a nontrivial explanation, however, related to the above alternative, which the results of Study 1 cannot rule out. It is possible that subjects are reluctant to make extreme, stereotypic predictions for *any* real, living-and-breathing individual. This intriguing possibility was explored in Study 2.

(2) Subjects in the videotape condition might have been swamped with information and might have forgotten the critical diagnostic information about major and career plans by the time they made their estimates. This possibility can be ruled out. At the end of the study, subjects were asked to recall the majors of each of the interviewees they saw. Though this occurred some 15 min after exposure to the first interview, 88% of subjects correctly recalled the majors of *all four* interviewees.

There is also a nontrivial version of this latter explanation, however.

---

[3] It should be noted that this design feature of counterbalancing major and career (diagnostic) information across the additional package of (nondiagnostic) information also rules out the possibility that the additional package of information *as a whole* suggested either more or less shock tolerance or moviegoing. If so, there would have been an effect of who played the role of the particular major, which there was not, for either study. It should also be noted that individual differences in beliefs about the diagnosticity of the items of additional information could not account for the dilution effect. If some subjects had believed that one or another item in the additional package was strongly diagnostic of increased shock tolerance or moviegoing, while other subjects had believed that some item was strongly diagnostic of decreased shock or moviegoing, this would have produced higher variance in the videotape condition than in the diagnostic-information-only conditions. In fact, however, variance was only trivially higher in videotape conditions than in the other conditions ($p > .25$ for both studies).

Though subjects might have been able to recall the diagnostic information, its salience might have been substantially reduced by the nondiagnostic information which followed it and which immediately preceded the prediction task. Such an explanation would be telling commentary on people's inability to combine information for purposes of prediction, but the explanation is quite different from the "dilution-of-similarity" hypothesis. The "salience" explanation cannot be ruled out by Study 1, but was explored in Study 2.

Study 2

Two conditions were added to the diagnostic-information-only and videotaped interview conditions of Study 1. In one, the videotaped diagnostic-information-only condition, the major and career plan information were given by the target himself, in a brief videotape clip from the interview. In the other, all the verbally presented information from the videotape was presented in written, paragraph vignette form. If the dilution effect in Study 1 was due simply to the reluctance of subjects to apply stereotyped predictions to a flesh-and-blood person, then the brief videotape exposure should be sufficient to produce a substantial dilution effect and the full, written information condition should produce little dilution effect. If, on the other hand, the dilution effect is due to reduced similarity between target and outcome, then the brief videotape clip, because of the quantitatively small amount of nondiagnostic information it contains should produce little dilution, while the written interview information should produce a large dilution effect comparable to that obtained in the full videotaped interview condition.

The "salience" alternative was also examined in Study 2. If the diagnostic information in the videotape condition of Study 1 was diluted because its salience was reduced by the nondiagnostic information presented subsequently, then there should be less of a dilution effect if the diagnostic information were presented *after* the nondiagnostic information and just before the prediction task, when it should be relatively salient. In order to examine this possibility, half of the subjects in the written interview information condition were presented with paragraphs that gave the diagnostic information early, immediately after the target's name and year in school, while the other half of the subjects in this condition were given the information late in the paragraph.

Method

The diagnostic-information-only condition, individual version, and the videotaped interview condition of Study 1 were duplicated exactly. A videotaped diagnostic information condition was added. In this condition, subjects were shown a 10- to 15-sec clip from the videotaped interview in which the target announced his major and career plans. The target

was seen in a head-and-shoulders close-up throughout the segment. A written interview information condition was also added. In this condition subjects read paragraph vignettes giving all the verbally presented information from the videotaped interviews. For half the subjects in this condition, the diagnostic, major, and career plan information appeared as the third item, immediately after name and school class; and for the other half of the subjects the diagnostic information appears as the third from the last item, with 6–8 items of nondiagnostic information preceding it.

There were 80 male and female introductory psychology students in the written version of the interview information condition and 36 in each of the other conditions. Procedures and dependent measures were identical to those of Study 1.

Results

The conditions of Study 2 correspond to a 2 (diagnostic-information-only vs diagnostic-plus-nondiagnostic information) × 2 (videotaped vs written presentation) ANOVA design. If the dilution of predictions in the videotaped interview condition of Study 1 was due simply to subjects' reluctance to apply a stereotype to a real person, then the effect of seeing the target on videotape should be large relative to the effect of the nondiagnostic information. If the dilution effect was obtained because the nondiagnostic information reduced the similarity between target and outcome, then the effect of the nondiagnostic information should be large relative to the effect of the videotape.

It may be seen in Fig. 2 that the dilution effect is largely due to the presence of nondiagnostic information. The effect of the nondiagnostic information was substantial for both the shock study and the movie study; $F(1,184) = 13.53$ and 7.04, respectively, $p < .001$ and $< .01$, respectively. The effect of the videotape was slight in both cases ($F < 1$ for shock study, $F = 1.70$, n.s., for movie study). For the shock study, the effect of the videotape was clearly nil both for subjects who received diagnostic information only and for those who also received nondiagnostic information. For the movie study, the effect of the videotape was clearly nil for subjects who received the nondiagnostic information, but there is a strong suggestion that the videotape produced some dilution for subjects who received only diagnostic information. The contrast between the written and the videotaped diagnostic-information-only conditions is marginally significant. (The uncorrected and hence overly liberal t is actually significant at $p = .05$.) This single comparison aside, it is clear that the dilution effect is largely due to the presence of nondiagnostic information about the target and is not much influenced by exposure to the target on videotape.

The salience interpretation of the dilution effect was tested by examining the effect of presenting the diagnostic information early in the written vignettes vs late. If the diagnostic information had a reduced impact in the videotaped interview condition in Study 1 because its salience was lowered by the subsequently appearing nondiagnostic information, then the
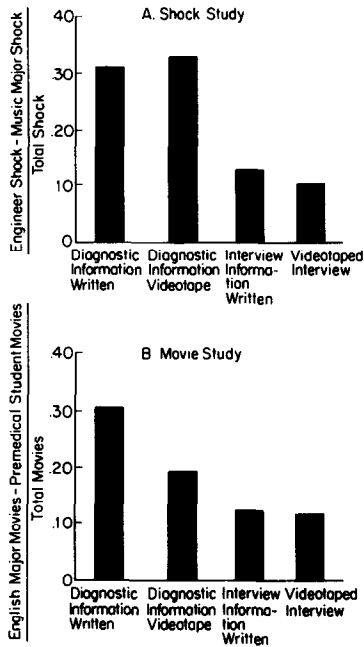
FIG. 2. Differential predictions of shock tolerance and of movie attendance as a function of videotaped vs written presentation of target information and as a function of presence vs absence of nondiagnostic information.

dilution effect should have been greater when diagnostic information was presented early than when it was presented late, immediately before predictions were made. In fact, however, the effect of order of presentation of diagnostic information was very slight. For both the shock study and the movie study, the dilution effect was only trivially greater when diagnostic information was presented early than when it was presented late (t = 1.12 for shock study, $t < 1$ for movie study).

Study 3

Studies 1 and 2 provide support for the notion that individuating but nondiagnostic information serves to dilute the effects of diagnostic information. In both studies, however, the diagnostic information continued to exert at least some sway over predictions. In both videotaped interview conditions of both Study 1 and Study 2, and in both of the written interview information conditions of Study 2, the predictions based on diagnostic information significantly survived the introduction of nondiagnostic information. It would be of interest to know whether, with somewhat less strongly diagnostic information, the effect of nondiagnostic information

might be sufficiently powerful to completely overcome the effects of the diagnostic information. This possibility was tested in the final study of this series.

The college majors in Studies 1 and 2 were chosen because the investigators suspected that they would produce maximally different predictions. For example, engineering seems intuitively to represent the college major with the strongest implications for high shock tolerance and music seemed one of the majors with the strongest implications for low shock tolerance. It seemed likely that a simple switch of the major pairs in the shock study with the major pairs in the movie study would produce weaker differential predictions in both studies. If so, it would be possible to examine the effects of nondiagnostic information on information that was less extremely diagnostic than that presented to subjects in Studies 1 and 2.

Study 3 also presented an opportunity for reexamining the effect of diagnostic information presented via videotape vs diagnostic information presented in written form. Study 2 was less than completely conclusive on this point, because, for the movie study, there was a trend toward dilution of the diagnostic information due to the brief exposure to the target on videotape.

Method

The procedure and dependent measures were identical in every respect to Study 2, except that the major pairs were reversed, and the written interview information condition was dropped. Subjects were 96 male and female University of Michigan introductory psychology students.

Results

It may be seen in Fig. 3 that, as anticipated, differential predictions were somewhat less extreme than in Study 1 or Study 2 for the diagnostic-information-only conditions. Differences between predictions for science majors and humanities majors in the written condition were still highly significant, however [$t(31)$ for shock study $= 3.23, p < .01$; $t(31)$ for movie study $= 5.16, p < .001$]. These differences largely sustained themselves when subjects briefly saw the interviewees on videotape [$t(31)$ for shock study $= 2.02, p = .05; t(31)$ for movie study $= 2.86, p < .01$]. For neither the shock study nor the movie study were the differential predictions significantly reduced in strength by the brief videotape exposure to interviewees (both t's < 1). Thus it may be concluded that merely seeing and hearing the target is not sufficient to reduce appreciably the predictions that would be made on the basis of knowledge of his major alone.

In contrast, it may be seen that exposure to the full 2-min interview was
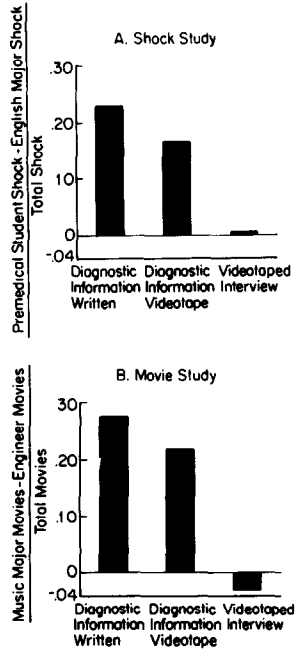
FIG. 3. Differential predictions of shock tolerance and of movie attendance as a function of means of presentation of diagnostic information and presence of nondiagnostic information.

sufficient literally to obliterate inferences made on the basis of expsoure to diagnostic information only. The relatively weaker diagnostic information in Study 3 thus had no effect whatever on predictions when nondiagnostic information was added. The differential predictions of subjects exposed to the full videotaped interview for the shock study were significantly less extreme than predictions of subjects exposed only to written information about major and career plans [$t(63) = 1.95, p = .06$], though not significantly less extreme than predictions made by subjects exposed to videotaped information about major and career plans; $t(63) = 1.31$, n.s. For the movie study, differential predictions for subjects in the full videotape condition were less extreme than those both of subjects in the written diagnostic information condition [$t(63) = 3.12, p < .01$] and the videotape diagnostic information condition; $t(63) = 2.21, p < .05$.

STUDIES 4 AND 5

The results of Studies 1–3 are, collectively, subject to a variety of interpretations which seemed to be best addressed by moving to new materials and designs. First, it is possible that the dilution effect is obtainable *only* when the diagnostic information is of the stereotypic sort

employed in Studies 1–3. It is possible that any real person, as seen in the videotaped interview or as portrayed in the vignettes, would fail to sustain the stereotyped image of an "engineer" or a "premed." The label "engineer" may call up an entire, culturally shared "prototype" (Cantor & Mischel, 1979), but when individuating details are provided, these might serve to cast doubt on the "prototypicality" of the person. The dilution effect might then result from a failure of the individuated target to match the prototype and consequent reluctance on the part of subjects to make stereotypical predictions. This account is a variant of the similarity explanation of the dilution effect: The nondiagnostic individuating information might serve to render the person *less similar to the prototype* than would be the case if there were no individuating information at all, and the individuated person is then seen as being less similar to the outcome than is the prototype.

We suspect that such prototype weakening is at least partially operative in Studies 1–3, but we also believe the dilution effect will be found even when the diagnostic information does not serve to call up a powerful, culturally shared stereotype. In Studies 4 and 5, the diagnostic information used was not a social category label. Instead, the diagnostic information served to suggest an *outcome* that was a social category label that might be expected to have some of the properties of a stereotype.

There is another explanation of the initial findings that is not so readily interpretable in terms of our similarity analysis. Taken singly, the items used in Studies 1–3 might have been nondiagnostic, but it is conceivable that in combination they might nevertheless have been seen as diagnostic. Thus a given subject might not have thought that the information that the target was "Catholic" was diagnostic of shock tolerance or that "being from Detroit" was diagnostic of shock tolerance, but still believed that the information that the target was "a Catholic from Detroit" *was* diagnostic of (average) shock tolerance. Alternatively, some of the items might have been seen as diagnostic in conjunction with the information about the target's major, even though the same items were not seen as diagnostic in isolation. Thus a given subject might not have believed that "Catholic" was by itself diagnostic of shock tolerance but still believed that Catholic engineers take less shock than engineers in general. These alternatives were hard to rule out for Studies 1–3 because so many items of nondiagnostic information were presented for each target person. One cannot be confident that such interaction effects did not occur with at least some combinations of items.

In the studies that follow, a sensitive, within-subject design was used, in which each subject was exposed both to a "diagnostic-information-only" condition and to conditions providing varying amounts of nondiagnostic information. It was hoped that a small number of items would be

sufficient to dilute significantly the implications of the diagnostic infor-
mation. In that event, it should be possible to assess the plausibility of the
hypothesis that the nondiagnostic items took on diagnosticity in conjunc-
tion with each other or in conjunction with the diagnostic items.

Finally, the studies employed a quite different subject population from
that in Studies 1–3. Subjects were graduate students in social work,
making judgments of a sort that are very common and very important in
their profession—predictions about child abuse.

Overview

Subjects made predictions about the likelihood that several "middle-
class, male social-work clients," were child abusers. Though the child-
abuse outcome might be expected to be a social category with some of the
properties of a stereotype, at least for social-work students, the diagnostic
information presented to subjects was not of the category label type.
Instead, the diagnostic information consisted of personal facts, such as
"has severe financial pressures and unpaid debts," which were expected
to suggest a predisposition toward child abuse.

The "clients" were described in brief vignettes. In Study 4, each vi-
gnette contained either two items of information judged to be diagnostic of
child abuse by pretest subjects or two items of information judged to be
counterdiagnostic of child abuse. Vignettes contained 0, 2, 4, or 8 items of
information judged to be nondiagnostic by pretest subjects. The design
was thus a 2 (diagnostic vs counterdiagnostic information) × 4 (levels of
nondiagnostic information), with each subject receiving every treatment.
The design makes it possible to examine the effect of increasing amounts
of nondiagnostic information on both diagnostic and counterdiagnostic
information.

Study 5 was identical to Study 4 except that only one item of either
diagnostic or counterdiagnostic information was included in each vi-
gnette. In addition, a separate group of control subjects was exposed to
vignettes containing only the various amounts of nondiagnostic informa-
tion, and no diagnostic or counterdiagnostic information. This control
group made it possible to determine whether increasing amounts of non-
diagnostic information are increasingly suggestive of child abuse (or of
nonabuse).

Subjects and Setting

All of the subjects were enrolled in a 2-year Master's program at the University of Michi-
gan's Graduate School of Social Work. On the average, subjects had more than 2½ years of
practical experience in social-work settings. All questionnaires were completed during reg-
ular class sessions. No subject participated in more than one of the three phases of data
gathering.

Pretest

Fifty-nine graduate students in social work were given questionnaires dealing with child abuse. The cover story—employed in the two experiments which follow with only minor modifications—asserted that recent evidence had indicated higher rates of child abuse among the middle class than formerly had been recognized: "In light of the seriousness of child abuse, and the helplessness of its victims, it is important to develop diagnostic procedures which will aid helping professionals in identifying the middle-class abusing parent. To design effective diagnostic training procedures which will enable helping professionals to better predict instances of child abuse," the cover story continued, "it would be useful to know how accurately people can predict, or diagnose, child abuse before receiving any special training." Subjects were then given 150 items of information about *adult, male, middle-class* clients allegedly drawn from existing records at a community mental health agency. "Some of these items are drawn from the records of clients who *are known to be* child abusers. Others *are not* child abusers, although of course they do have other social and personal problems" for which they would have sought help. The subjects' task was "to evaluate each separate piece of information and rate whether you think that information is helpful with respect to diagnosing child abuse."

The 150 information items were arranged in four random orders and subjects were randomly assigned to one of the four orders. In a blank next to each item subjects rated the item in one of the following ways:

Diagnostic: "+ (plus) means that the piece of information, taken by itself, suggests that the person *is* a child abuser."

Counterdiagnostic: "− (minus) means that the piece of information, taken by itself, suggests that the person *is not* a child abuser."

Nondiagnostic: "0 (zero) means that the piece of information, taken by itself, is of no value in deciding whether or not the person is a child abuser."

Diagnostic items. The eight items that most strongly suggested to subjects that the client was an abuser were selected. At least 63% of subjects believed that each item was diagnostic of child abuse. The items included the following: "He is aroused by sadomasochistic sexual fantasies," "He has a drinking problem," "He was sexually assaulted by his stepfather," and "He has no friends."

Counterdiagnostic items. The eight items that most strongly suggested to subjects that the client was not an abuser were selected. At least 44% of subjects believed that each item was counterdiagnostic. The items included the following: "He would like to adopt a second child," "He has many close friends," "He works with a local group to ease school integration," "He does volunteer work for a hot line."

Nondiagnostic items. Twenty-eight nondiagnostic items were selected such that at least 85% of the pretest subjects had judged the items to be "of no value" in diagnosis and no more than 10% thought the item to be diagnostic of either abuse or nonabuse. On the average, the items were believed to be nondiagnostic by 90% of the subjects. These items were the most diagnostically useless, in the subjects' estimation, of the 150-item pool. Sample items included: "He manages a hardware store," "He lost two fingers on his left hand," "He was arrested for joyriding at age 17," and "He has an IQ of 110."

Study 4

Procedure

Forty-eight subjects rated the likelihood that each of 12 "clients," described in vignettes of varying lengths, was one of the "known abusers" in a sample of clients. Ratings were made on 11-point scales, ranging from 1 ("A very good chance that he *is not*") [one of the abusers] to 11 ("A very good chance that he *is*"). The cover story was essentially the same as that provided for pretest subjects, with modifications to make it clear that each vignette, composed of 2 to 10 items, described one particular client. Each subject rated (a) four vignettes consisting of 2 diagnostic items plus 0, 2, 4, or 8 nondiagnostic items, (b) four vignettes consisting of 2 counterdiagnostic items plus 0, 2, 4, or 8 nondiagnostic items, and (c) four filler vignettes of 2, 4, 6, and 10 items constituting a mix of weakly diagnostic and weakly counterdiagnostic items.

Questionnaire Design

Nondiagnostic item sets. Of the 28 nondiagnostic items, 6 specified an occupation, 2 an IQ score, and 2 a place of birth. The remaining 18 items covered a range of topics including habits, hobbies, youthful indiscretions, and physical characteristics. Four different sets of nondiagnostic items, consisting of one cluster each of 2, 4, and 8 items, were constructed from the pool of 28 items. Items were first assigned randomly to one of two sets consisting of three clusters, with the constraints that there be one occupation item in each cluster but no more than one occupation, IQ, or birthplace item could appear in a given cluster. These two sets were then doubled into four total sets by creating two different random orders of each of the 14 items in the set. Each subject received each of the 28 items exactly once.

Pairs of diagnostic and counterdiagnostic items. The eight diagnostic items were rank-ordered in terms of degree of diagnosticity as judged by pretest subjects. These items were paired—the strongest (1) with the weakest (8), 2 with 7, 3 with 6, and 4 with 5—to create four pairs of roughly equal diagnostic strength. The eight counterdiagnostic items were also rank-ordered and paired (1 and 8, 2 and 7, 3 and 6, and 4 and 5).

Construction of vignettes. A Latin square design was used to rotate diagnostic and nondiagnostic pairs through each of the nondiagnostic item sets. Each pair appeared equally often in each nondiagnostic set. The two

items in each pair were inserted in the second and second-to-last positions in each cluster of nondiagnostic items. The order in which each vignette appeared in the questionnaire was random. The design resulting from this method of constructing vignettes made it possible to assess three different variables: (1) the effect of diagnostic vs counterdiagnostic items, (2) the effect of number of nondiagnostic items, and (3) the effect of the particular set of nondiagnostic items in which diagnostic and counterdiagnostic pairs were embedded.

Study 5

Study 5 provided a replication of Study 4, with the exception that only one item of diagnostic or counterdiagnostic information was used in the vignettes instead of two. Study 5 also examined the effect of increasing amounts of purely nondiagnostic information, in the absence of any diagnostic or counterdiagnostic information.

Single diagnostic item replication of Study 4. Forty-eight subjects drawn from the same population as Study 4 rated 12 vignettes. The four items judged by pretest subjects to be most diagnostic of abuse and the four items judged to be most counterdiagnostic were selected. These items appeared as the sole piece of information and also in conjunction with two, four, and eight nondiagnostic items. When inserted in the nondiagnostic clusters the middle position was used, so that the diagnostic or counterdiagnostic item appeared as the second of three items, the third of five items, or the fifth of nine items. All diagnostic and counterdiagnostic items were rotated through the Latin square design as in Study 4.

Increasing amounts of purely nondiagnostic information. Forty-eight additional subjects rated six vignettes consisting of purely nondiagnostic information. The nondiagnostic clusters by themselves were now presented alone as two-, four- and eight-item vignettes. The vignettes were randomly interspersed with six filler vignettes moderately suggestive of an abuser and six moderately suggestive of a nonabuser.

Results

The results of the two studies were entirely congruent, and are therefore presented together. Figure 4 presents the ratings of the likelihood that the client was a child abuser as a function of the number of nondiagnostic items present in the vignettes.

Effect of nondiagnostic information on vignettes containing diagnostic information. The results were analyzed separately for diagnostic and counterdiagnostic vignettes, because the effects of nondiagnostic information were quite different in the two cases. Analyses were 4 (levels of nondiagnostic information present) × 4 (set of nondiagnostic items) repeated-measures ANOVAs. When two diagnostic items were present,
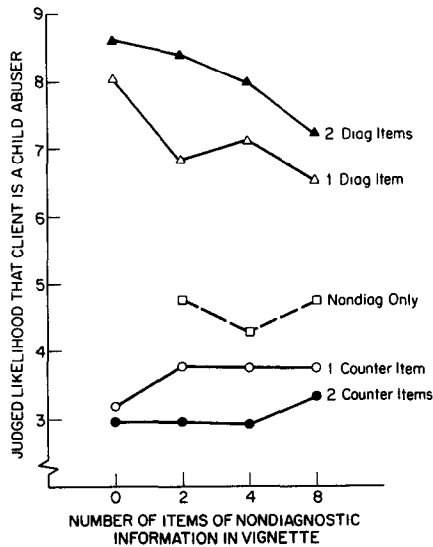
FIG. 4. Judged likelihood that the client is a child abuser as a function of diagnostic vs counterdiagnostic information and as a function of number of nondiagnostic items. Diag = items diagnostic of abuse; Counter = items counterdiagnostic of abuse; Nondiag = nondiagnostic items.

the effect of increasing amounts of nondiagnostic information was pronounced; $F(3,132) = 6.49, p < .001$. Neuman–Keuls tests comparing ratings for vignettes containing the diagnostic items alone with ratings for vignettes containing two, four, or eight items of nondiagnostic information showed that the presence of as few as four items of nondiagnostic information was sufficient to "dilute" significantly the extremity of ratings based on diagnostic information alone; $p < .01$. Thus, for example, a client who "is quite short-tempered" and "has severe financial pressures and unpaid debts" was rated as significantly less likely to be a child abuser when it was also known that "his mother is a housewife," "he has tatoos on both forearms," "he manages a hardware store" and "he fixes up old cars in his spare time." The effect of the particular set of nondiagnostic items present in the vignette was not significant, for this or any of the other analyses.

The effect of increasing amounts of nondiagnostic information was also pronounced when only one item diagnostic of abuse was present in the vignette; $F(3,132) = 8.87, p < .001$. Neuman–Keuls tests showed that as few as two items of nondiagnostic information were sufficient to dilute the extremity of ratings based on one item of diagnostic information; $p < .01$. Thus, for example, a client who "wanted his wife to have their child aborted" was judged as significantly less likely to be a child abuser if it

was also known that he "is currently employed as a shoe salesman" and "has a serious hearing defect."

Effects of nondiagnostic information on vignettes containing counter-diagnostic information. In contrast to the effect of nondiagnostic information when it appeared in vignettes containing diagnostic information, the effect of nondiagnostic information was very slight when it appeared in vignettes containing counterdiagnostic information. Whether there were two counterdiagnostic items or only one, nondiagnostic information had no significant effect on ratings; $F < 1$, $F(3,132) = 1.63$, n.s., respectively. Thus a client who "has many close friends" was judged unlikely to be a child abuser, and nondiagnostic information such as "he has tatoos on both forearms" or "his mother is a housewife" did not serve to dilute significantly such a judgment, even when there were as many as eight such items.

Effect of increasing amounts of nondiagnostic information in the absence of diagnostic or counterdiagnostic information. There was no detectable effect of increasing the amount of purely nondiagnostic information; $F(2,92) = 2.21$, n.s. Increasing amounts of nondiagnostic information apparently suggest neither an increasing likelihood of abuse nor a decreasing likelihood of abuse. The finding that increasing amounts of nondiagnostic information produced decreasing judgments of the likelihood of child abuse, for subjects presented with diagnostic information, is therefore not artifactual: Increasing amounts of nondiagnostic information do not by themselves suggest a lower likelihood of abuse.

The data for purely nondiagnostic vignettes speak also against an "interactive diagnosticity" explanation of the dilution effect. Across all 12 of the combinations of nondiagnostic information (three levels of amount of information—two, four, or eight—times four different clusters of items at each level), the range of mean judged likelihood of abuse is very narrow—from 4.08 to 5.21. Thus, *for subjects as a group,* it is quite unlikely that any of the combinations became powerfully diagnostic or powerfully counterdiagnostic. It of course remains possible that there were equal and opposite interaction effects for some (or even all) combinations such that the singly nondiagnostic items became powerfully diagnostic for some subjects and these same items became powerfully counterdiagnostic for others. This is implausible and unparsimonious, however, and, in addition, the variation around the mean at the various combinations differed too little to give any support to this possibility: The standard deviations across the 12 combinations ranged only from 1.59 to 2.20. If some of the combinations had seemed very diagnostic to some subjects and very counterdiagnostic to others while other combinations seemed uniformly nondiagnostic to most subjects, differences in the variance across combinations should have been substantial.

Finally, there is the possibility of "interactive diagnosticity" between items judged as diagnostic by pretest subjects and items judged as nondiagnostic, such that the "nondiagnostic" item acquired diagnosticity (or counterdiagnosticity) in conjunction with the diagnostic item. An examination of the simplest combinations—one diagnostic item with two nondiagnostic items—indicates that such a possibility does appear plausible for one of the diagnostic items in combination with each of the four clusters of two nondiagnostic items. This was the diagnostic item "he has a drinking problem." Since all of the clusters of nondiagnostic items, including the two-item clusters, presented an occupational item, this means a distinct possibility of interactive counterdiagnosticity: the fact of employment in *any* job might be expected to mitigate the presumed severity of the drinking problem and thus to reduce the likelihood that the "drinker" also abuses his children. For none of the other diagnostic items employed in Study 5 ("sexually assaulted by his stepfather," "wanted his wife to have their child aborted," "quite short-tempered") does such a pattern of genuine interactive diagnosticity seem plausible. (The two-item nondiagnostic pairs were: "he is currently employed as a shoe salesman, he has a serious hearing defect;" "he is currently employed as a bartender, he was born in Muskegon;" "he manages a hardware store, he once ran away from home as a boy;" and "he works as an insurance salesman, he has a very handsome and athletic appearance.") In fact, however, the "drinking problem" diagnostic item showed slightly and nonsignificantly *less* dilution than the other diagnostic items as a group. Thus "interactive diagnosticity" does not seem a highly credible explanation for the dilution effect. In the only combination of items where such a pattern seems at all plausible, the dilution effect was not greater than for combinations where it seems quite implausible.

Dilution of diagnostic information vs dilution of counterdiagnostic information. Why was there such a marked dilution effect for the diagnostic items while there was an undetectable dilution effect for the counterdiagnostic items? One possibility is that there was simply a ceiling effect for the counterdiagnostic items. The mean for the counterdiagnostic items was closer to the mean for purely nondiagnostic items than was the mean for the diagnostic items. If the average of the purely nondiagnostic vignettes is taken as the central value which both diagnostic and counterdiagnostic vignettes could approach as a limit, then the amount of dilution possible was much greater for diagnostic than for counterdiagnostic items.

The above interpretation is both plausible and parsimonious. However, we believe it is possible that a much more interesting process, consistent with the similarity interpretation of the dilution effect, might have been at work. Routine, neutral, unexceptional facts about a person such as our

nondiagnostic items may be seen as more dissimilar to the unpleasant, undesirable diagnostic items than to the admirable counterdiagnostic items. People may be more prepared to find the good and the mundane mixed together in the same person than to believe that the evil and the mundane can coexist. A storm of protest greeted Hannah Arendt's 1965 (1968) description of Adolph Eichmann, the man who oversaw the destruction of European Jewry, as a bland, ordinary bureaucrat, a man indistinguishable in his outward features from millions of people with purely innocent occupations and concerns. Thus neutral, nondiagnostic attributes may reduce the similarity between the sinister, diagnostic target features and the child-abuse outcome more than they reduce the similarity between the admirable, counterdiagnostic target features and the nonabuse outcome.

DISCUSSION

Person Perception and the Dilution Effect

Perhaps the most interesting aspect of the present work is that which speaks to the role of stereotypes and knowledge structures in the processing of social information. The results of Studies 1–3 indicate that nondiagnostic information weakens the implications of social stereotypes. The results of Studies 4 and 5 indicate that nondiagnostic information weakens the implications of diagnostic information, which, though not of a stereotypic nature in itself, suggests a social category outcome that may have some of the characteristics of a stereotype. One interpretation of these results is that stereotypes may play a somewhat less powerful and pervasive role in person perception than has been assumed by many social psychologists. Stereotypes may operate most powerfully in the abstract, applying primarily to undifferentiated groups or individuals, and they may exert relatively little impact on judgments about concrete, individuated targets. This proposition has also been suggested by Locksley, Borgida, Brekke, and Hepburn (1980; see also Borgida, Locksley, and Brekke, 1980) in their work on sex stereotypes. These investigators found that knowledge of a target's sex did not influence judgments about the target's assertiveness or other stereotypically sex-related traits when the target was individuated by presenting additional *diagnostic* information about the target's assertiveness. This information consisted of descriptions either of highly assertive or highly nonassertive behavior on the part of the target. We may speculate that, whether individuating information is diagnostic or nondiagnostic, judgments about concrete targets may show little effect of the same stereotypic labels that exert a pronounced effect on judgments about abstract, undifferentiated groups or individuals.

Thus, although stereotypes exert a demonstrable effect on the interpretation of ambiguous information and on recall of information about

targets (Hamilton, 1979), stereotypes nevertheless may turn out to be less powerfully determinative of many kinds of social judgments than might be presumed. This supposition has the advantage of accounting in a satisfactory way for the "some of my best friends are . . ." phenomenon. People may believe that "blacks are lazy" without believing that any of the blacks whom they know personally are lazy. And one may speculate that the pernicious effects of social stereotypes may consist largely in the fact that they prevent the stereotype holder from exposure to the individuating information that, were it known, would serve to weaken or undermine the applicability of the stereotype.

Judgmental Strategies Underlying the Dilution Effect

Counternormative Similarity Judgment or Normative Prediction Strategy?

We have suggested that the dilution effect is counternormative, that is, that it represents an inappropriate use of information for making predictions. The presumption of counternormativeness rests on the view that the "nondiagnostic" information remains nondiagnostic in the full judgmental context and that its utilization in prediction is therefore inappropriate. Unfortunately, this argument is intrinsically very difficult to establish. By some standards, the information is diagnostic by definition, since it is actually used in judgments (cf. Anderson, 1968). For the present results we can only argue that it is implausible that subjects literally believe that the individual items of nondiagnostic information are diagnostic when they are combined with each other or with the diagnostic information. It seems very unlikely to us, for example, that subjects believe that the information that a client "is currently employed as a bartender" suggests a decreased likelihood of child abuse when combined with the information that "he was born in Muskegon," or believe that either item somehow logically mitigates the diagnostic implications of the knowledge that "he wanted his wife to have their child aborted."

To rule out the possibility that subjects are using a normatively appropriate strategy, however, will require several converging lines of evidence. A start toward such convergence has been made by Zukier (1981). Normative models of prediction require that subjects be responsive both to the target's scores on predictor variables and to the presumed association between the predictor variables and the outcome variable. Zukier showed that, when the target's scores on predictor variables are *average*, subjects are quite unresponsive to the degree of association which they believe to exist between the predictors and the outcome. This finding strongly suggests that subjects do not in fact respond to the diagnosticity of information in the way that normatively appropriate strategies require.

In the present work, the strongest *prima facie* evidence for the counter-normativeness of the dilution effect is found in Study 3. Subjects exposed only to the diagnostic information about the target's major were heavily influenced by it. They predicted that premedical students would tolerate over 40% more shock than English majors and predicted that music majors would attend over 50% more movies than engineers. Yet this same specification of major exerted no detectable influence on predictions about targets seen in the full videotaped interview. One cannot have it both ways. Either premedical students tolerate a great deal more shock than English majors or they do not. If premedical students do tolerate a great deal more shock than English majors, then any particular, individuated premedical student tolerates more shock than any particular, individuated English major, unless the individuating information strongly contradicts the implications of the major by itself. It seems quite unlikely that subjects actually believed that the individuating information contradicted the shock-tolerance implications of the major. If not, then subjects are inconsistent to a degree that, on the face of it, certainly seems to be counternormative.

"Cognitive Algebra" or "Similarity Judgment"?

We should also emphasize that we have no direct evidence that it is similarity judgments that underlie the dilution effect rather than some other mechanism. We believe we have ruled out a number of alternative explanations for the effect, including all of the trivializing ones that have been suggested to us, but such a process cannot be completely satisfying. Again, converging lines of evidence will be necessary to establish similarity judgments as the primary mechanism underlying the dilution effect. We should speak directly, however, about an alternative view stemming from Anderson's work (e.g., 1968). This is the possibility that subjects simply "average" the "scale value" of each of the items of information while ignoring the appropriate "weights" of the nondiagnostic items (which by definition should be zero).[4] This view seems to us to be less an "alternative explanation" of the results than a description of the results, which is compatible with our own, and which is silent as to the specific mechanism that might produce such averaging. Put somewhat differently, it is our view that similarity judgments produce the apparent "averaging" of scale values.

Practical Implications

"Regressive" Predictions from Nonregressive Prediction Strategies

One of the chief implications of the present research is that predictions made in complex, information-rich judgmental contexts may sometimes

[4] See Note Added in Proof.

be less extreme than previous findings would suggest. Kahneman and Tversky (1973) showed that, when subjects are presented with exclusively diagnostic information, predictions are highly nonregressive. Subjects' predictions about outcome variables tend to be as extreme as scores on predictor variables, even when subjects do not believe that there is a very strong association between predictor variables and outcome variables. For example, subjects predict that a target with an extremely good sense of humor would have an extremely high grade point average even though they do not believe that there is a very strong association between sense of humor and grade point average. The present work suggests that if subjects also knew a great many individuating but nondiagnostic facts about the individual with a good sense of humor they would make less extreme predictions because such information would lower the similarity between the target and the high-grade-point-average outcome.

It should be emphasized, however, that such "regressiveness" would not necessarily be expected in all real-world judgmental contexts. In the present work, subjects' attention was by design directed toward the non-diagnostic information and this may have heightened greatly the salience of such information. In Tversky's (1977) formalization of similarity judgments, he stressed that it is the similarity of *salient* features of targets and outcomes that are compared. In many naturalistic contexts, the salience of nondiagnostic information might be substantially less than in the present research. Indeed, when subjects cannot rely on "external" sources of information but must search their memories for the relevant information, they sometimes may examine *only* facts believed to be diagnostic and thus may make predictions that are fully as nonregressive as those demonstrated by Kahneman and Tversky.

Implications for Professional Diagnostic Settings

Finally, it should be noted that the present research provides one more demonstration in a long line of studies indicating that expert judges show the same deficiencies in their professional judgmental strategies as untutored subjects. Subjects in Studies 4 and 5 were graduate students with professional experience making judgments similar to those made in the conduct of their professional work.

It is important to note some of the implications of the present work for decisions such as those made by social workers and other institutional "gatekeepers." Many years ago, Goffman (1961) pointed to the dangerous potential of case records, for mental patients and other "deviants," which contain only "diagnostic" information about the individual. Goffman's concern was that predictions based on such highly selected, exclusively diagnostic information might be too extreme and yield too confident a judgment of deviance. Research by Kahneman and Tversky (1972; 1973) provides strong evidence favoring Goffman's presumption.

People do indeed make predictions that are much too confident and extreme when they are provided only with diagnostic information.

The present research, however, suggests that the complementary danger may also arise. If the case record, or other information available to the gatekeeper, contains large amounts of information which, taken singly, would be believed by *the judge himself* to be valueless for purposes of prediction, extreme predictions, *even when correct,* may be set aside in the context of the complete file of information. As a consequence, individuals who are a danger to themselves or others might be likely to be untreated, released, or paroled, simply as a function of the amount of worthless information about the individual which the gatekeeper happens to possess. The present results thus provide yet another reason for expert judges to abandon intuitive strategies of judgment whenever more formal methods of assessment are possible.

Note added in proof. Since this article was accepted for publication, our attention was drawn to an article by C. M. Troutman and J. Shanteau ("Inference based on nondiagnostic information," *Organizational Behavior and Human Performance,* 1977, 19, 43–55) who work in the Anderson tradition. These investigators showed in a bookbag and poker chip study that subjects were influenced by information that the investigators believed to be nondiagnostic. This demonstration is conceptually similar to the present one, although it is not the same, since no attempt was made to determine whether the subjects themselves believed, outside the full judgmental context, that the information was nondiagnostic.

REFERENCES

Anderson, N. A simple model for information integration. In R. P. Abelson, et al. (Eds.), *Theories of cognitive consistency: A source book.* Chicago: Rand McNally, 1968.

Arendt, H. *Eichmann in Jerusalem: A report on the banality of evil.* Chicago: Rand McNally, 1968.

Borgida, E., Locksley, A., & Brekke, N. Social stereotypes and social judgment. In N. Cantor & J. Kihlstrom (Eds.), *Cognition, social interaction and personality.* Hillsdale, NJ: Lawrence Erlbaum Associates, 1980.

Cantor, N., & Mischel, W. Categorization processes in the perception of people. In L. Berkowitz (Ed.), *Advances in experimental social psychology,* New York: Academic Press, 1979. Vol. 12.

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

Goffman, E. *Asylums: Essays on the social situation of mental patients and their inmates.* Garden City, NY: Doubleday, 1961.

Goldberg, L. R. Simple models or simple processes? Some research on clinical judgments. *American Psychologist,* 1968, 23, 483–496.

Hamilton, D. A cognitive-attributional analysis of stereotyping. In L. Berkowitz (Ed.), *Advances in experimental social psychology.* New York: Academic Press, 1979. Vol. 12.

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

Kahneman, D., & Tversky, A. On the psychology of prediction. *Psychological Review,* 1973, 80, 237–251.

Locksley, A., Borgida, E., Brekke, N., & Hepburn C. Sex stereotypes and social judg-
    ments. *Journal of Personality and Social Psychology,* 1980, **39,** 821–831.
Meehl, P. E. *Clinical versus statistical prediction: A theoretical analysis and a review of the
    literature.* Minneapolis: Univ. of Minnesota Press, 1954.
Nisbett, R. E. & Borgida, E. Attribution and the psychology of prediction. *Journal of Per-
    sonality and Social Psychology,* 1975, **32,** 932–943.
Nisbett, R. E., & Ross, L. *Human inference: Strategies and shortcomings of social judg-
    ment.* Englewood Cliffs, NJ: Prentice–Hall, 1980.
Tversky, A. Features of similarity. *Psychological Review,* 1977, **84,** 327–352.
Zukier, H. *The dilution effect: Normative prediction strategy or counternormative similarity
    judgment?* Unpublished manuscript, The New School for Social Research, 1981.