Journal of Personality and Social Psychology
1982, Vol. 43, No. 1, 22-34

0022-3514/82/4301-0022$00.75

Confirmatory and Diagnosing Strategies in Social Information Gathering

Yaacov Trope and Miriam Bassok
Hebrew University of Jerusalem, Jerusalem, Israel

This article concerns the strategies people use to gather information about another's personality. In a series of three experiments, subjects were asked to test the hypothesis that another person possesses a certain personality trait by selecting sources of behavioral evidence. The two main informational properties on which these sources of evidence varied were (a) the probability of the evidence under the hypothesized trait, from highly probable evidence to highly improbable evidence and (b) the diagnosticity of the evidence (i.e., the extent to which the evidence was differentially probable under the hypothesized and alternative traits). The results showed diagnosticity to be the major determinant of the information-gathering preferences. There was very little evidence for a confirmatory strategy wherein evidence probable under the hypothesized trait is preferred to evidence improbable under this trait. In fact, improbable evidence was preferred when it was more diagnostic than probable evidence. Thus, the confirmatory strategy, whatever appeal it may have had to subjects, did not reduce the diagnostic power of the information assembled.

Confirmatory Information Gathering

Past social judgment research has mainly been concerned with inferences from given pieces of information (see Einhorn & Hogarth, 1981; Kahneman, Slovic, & Tversky, 1982; Nisbett & Ross, 1980). This research has neglected the preceding phase of information gathering: the choice among available information sources, the construction of new information sources, and the determination of the amount of information from each.

An important exception to the neglect of information gathering is Synder and Swann's research program on hypothesis testing (Snyder, 1981; Snyder & Campbell, 1980; Snyder & Swann, 1978). These psychologists sought to investigate how people gather information for the purpose of testing a hypothesis about another's personality. Their subjects were instructed to select interview questions to test whether a target person possessed a certain trait. Some subjects tested the hypothesis that the target was an extrovert whereas others tested the hypothesis that the target was an introvert. On the basis of their studies, Snyder and Swann concluded that people gather information by using a hypothesis-confirming strategy that was defined as "the preferential soliciting of behavioral evidence whose presence would tend to confirm the hypothesis under scrutiny" (Snyder, 1981, p. 280). Thus, subjects who were asked to test the hypothesis that the target was an extrovert preferred "precisely those questions that one typically asks of people already known to be extroverts" (p. 279) such as "What would you do if you wanted to liven things up at a party?" (p. 279). In contrast, when testing the hypothesis that their target was an introvert, subjects preferred questions that are typically asked of people already known to be introverts, such as "What factors make it hard for you to really open up to people?" (p. 280). The mere phrasing of the hypothesis in terms of one alternative rather than the other, regardless of its prior likelihood, was sufficient to produce the preferential soliciting of hypothesis-confirming evidence.

According to Snyder and Swann, this

Thanks are due to Eugene Burnstein, Ruma Falk, and Michael Ross for their comments on an earlier version of this article.

Requests for reprints should be sent to Yaacov Trope, Department of Psychology, Hebrew University of Jerusalem, Jerusalem, Israel.

strategy is likely to produce erroneous behavioral confirmation of any hypothesis that the information gatherer happens to be testing. In Snyder's (1981) words:

> After all, most (if not all) people have been to some parties and have spent some time alone in their lifetimes. Moreover, most (if not all) people will be particularly likely to talk about parties when asked to do so in conversation. Similarly, they will be particularly likely to talk about quieter times when conversations move in that direction. (p. 278)

Indeed, interviewees who were asked the questions selected by subjects actually behaved in a manner that confirmed the tested hypothesis.

Snyder, Swann, and their associates conducted several investigations to identify limiting conditions for confirmatory hypothesis testing. None was found, not even a procedure "that successfully diminishes the magnitude of the preferential soliciting of hypothesis-confirming evidence" (Snyder, 1981, p. 283). This research suggests that even if people were perfectly error-free information processors, they would still be trapped in perpetual confirmation of their hypotheses due to their biased information search strategies.

The Diagnosing Strategy in Information Gathering

As we previously noted, except for Snyder and Swann's work, there is little direct research on gathering of information about another's personality. However, a vast amount of indirectly related theorizing and research suggests that people can take into account factors affecting the diagnostic value of behavior (i.e., the extent to which its probability depends on possession of the personal disposition under consideration). Indeed, several psychologists have interpreted attribution formulations in terms of factors affecting behavior diagnosticity (Ajzen & Fishbein, 1975; Trope, 1974; Trope & Burnstein, 1975). Specifically, Kelley's (1967) well-known criteria for attributional validity specify that a behavior would not be attributed to an underlying cause (e.g., a person factor) unless the behavior is distinctively and consistently associated with the cause. Obviously a distinctive and consistent behavior is more probable when the cause is present than when it is absent and is therefore diagnostic of that cause.

Moreover, in the absence of such covariation data, people can assess the diagnosticity of behavior according to their cognitive representations of personality traits (Rosenberg & Sedlak, 1972; Schneider, 1973) and of the causal structure of their social environment (Heider, 1958; Tversky & Kahneman, 1980). These knowledge structures can guide such fine judgmental operations as discounting and augmenting the role of a given personal disposition in producing a behavior when other facilitative or inhibitory causes are detected (Kelley, 1972). In the presence of facilitative causes, the behavior is of little diagnostic value because its probability remains high regardless of whether the person possesses the focal disposition, whereas in the presence of an inhibitory cause the behavior is highly diagnostic because the disposition is necessary for the behavior to occur. Indeed, more than two decades of research on trait attribution has demonstrated at least some degree of sensitivity to various factors affecting diagnosticity, such as statuses (Thibaut & Riecken, 1955), roles (Jones, Davis, & Gergen, 1961; Trope & Burnstein, 1975), surveillance (Strickland, 1958), and amount of freedom of choice (Steiner, 1970; Trope, 1978; Trope & Burnstein, 1977). Finally, a large body of research has shown that when evaluating their skills, people know how to select tasks (Trope, 1975, 1979, 1980; Trope & Brickman, 1975) and others to compare themselves with (Ruble & Boggiano, 1980) that are maximally diagnostic.

With such judgmental skills, the information gatherer wishing to decide whether or not a target person possesses the hypothesized trait should be capable of selecting questions that elicit subjectively diagnostic answers. The selection of questions may depend on the person's knowledge about the behaviors, feelings, preferences, and other personal features associated with the hypothesized and alternative traits. Such features will be judged diagnostic to the extent that the conditional probability of the feature given the hypothesized trait is believed to be different from that given the alternative

trait. Thus, in the diagnosing strategy, the information gatherer will select questions asking about features that are maximally diagnostic with respect to the hypothesized and alternative traits. In this strategy, questions about diagnostic features should be preferred regardless of whether the features are probable or improbable under the hypothesized trait.

In contrast, in the confirming strategy, the information gatherer will ask about a feature to the extent that it is probable given the hypothesized trait. More specifically, this strategy can be interpreted to mean that the information gatherer concentrates on the hypothesized trait, disregarding the alternative(s), and asks questions involving features that are probable given the hypothesized trait while rejecting features that are improbable given this trait. The extent to which a feature is diagnostic or differentially probable under the hypothesis and alternative(s) is irrelevant to the confirming strategy and should not affect the choice of a feature for inclusion in a question. For example, in testing the hypothesis that the target is an introvert by the confirming strategy, the information gatherer is unlikely to ask "Do you like parties?" because the feature "liking parties" is improbable under the introvert hypothesis. This question will be rejected even if it is of subjective high diagnosticity.

Since Snyder and Swann did not study diagnosticity effects, these contrasting predictions remained untested. To provide such a test, the present series of experiments orthogonally varied the probability of features under the hypothesis and the diagnosticity of questions about the features. As in Snyder and Swann's studies, subjects were provided with the hypothesis that the target possessed one of two personality traits. Subjects were then presented with a list of behavioral features and were asked to state their preferences among them for testing their hypothesis.[1]

Experiment 1

Overview

The subjects' task was to select questions about handwriting features for testing the hypothesis that the target has a given trait. Half of the subjects tested the hypothesis that the person is predominantly intuitive (intuitive hypothesis) and the remaining subjects tested the hypothesis that the person is predominantly analytic (analytic hypothesis). There were eight handwriting features such as slanting, curvature, and printing. These trait categories and features were chosen on the basis of informal pilot studies to be somewhat familiar, yet sufficiently ambiguous to make any manipulated association between a category and a feature seem plausible. It was thus possible to vary the probability and diagnosticity of each feature by presenting fictitious rates of occurrence of each feature in the handwriting of intuitive and analytic persons. Four levels of probability under the hypothesis were crossed with two levels of diagnosticity, resulting in a total of eight features.

Method

Subjects

The subjects were 110 students from the University of Toronto introductory psychology classes. They served as subjects to satisfy a course requirement. The experiment was conducted in small groups varying in size from 3 to 5.

Procedure

Subjects received a booklet entitled *Handwriting and Personality*, which first indicated that "The study is concerned with the use of handwriting for determining personality traits." Subjects read that a large amount of psychological research has shown that people can be classified into those who are predominantly analytic and those who are predominantly intuitive. Subjects then read brief descriptions of these traits: "The predominantly analytic person is characterized by careful attention to details, conscious judgments and systematic identification of separable aspects of information" and "The predominantly intuitive person relies on direct apprehension, displays spontaneous insights, and responds to holistic configurations."

The booklet continued as follows:

> Consider now a group of male and female undergraduate students half of which are analytic persons and half intuitive. A panel of expert graphologists carefully analyzed the handwriting of each student,

[1] To simplify presentation, the term *diagnosticity* of (or preference for) a feature will be used to refer to the diagnosticity (or preference for) a question asking about the value (e.g., presence or absence) of a feature.

using a list of eight independent handwriting features such as slanting of the handwriting. The graphologists determined whether or not any given feature appears or does not appear in each student's handwriting. Your task is to select those handwriting features which will enable you to determine whether or not any given student is predominantly analytic.

In the intuitive-hypothesis condition, the term *intuitive* replaced *analytic* in the last sentence.

The eight traits were the following: (a) elimination of unessential elements, (b) emphasis on the lower zone, (c) deviations from standard form, (d) parallel lines, (e) printing, (f) threadlike connection, (g) variability in writing angle, and (h) variation in margin.

The only additional information that subjects received about each feature was its rate of occurrence in the handwriting of analytic and intuitive persons. This information was presented in two bar-diagrams of equal height, one for the analytic and one for the intuitive category, each of which was divided into a dark area (representing the percentage of people in the category in whose handwriting the feature was present) and a blank area (representing the complementary percentage of people in the category in whose handwriting the feature was absent).

Manipulation of diagnosticity and probability given the hypothesized trait. The percentage information was varied according to diagnosticity and probability given the hypothesized trait. Diagnosticity can be defined in terms of the likelihood ratio of the feature's presence, LR(F), and its absence, LR($\bar{F}$), where LR(F) = $P(F/T_1)/P(F/T_2)$ (i.e., the ratio of the conditional probability of the feature given one trait to its conditional probability given the other trait) and the likelihood ratio of the feature's absence, LR($\bar{F}$) = $P(\bar{F}/T_1)/P(\bar{F}/T_2)$ (i.e., the analogous ratio of conditional probabilities associated with the absence of the feature). Note that the indexes on the traits are chosen so that the numerator is equal to or greater than the denominator. Diagnosticity of a question about a feature can be defined, then, as the expected likelihood ratio, D, that is

$$D = P(F)LR(F) + P(\bar{F})LR(\bar{F}), \qquad (1)$$

where $P(F)$ and $P(\bar{F})$ are the respective total probabilities of the presence or the absence of the feature (for a more detailed discussion of Equation 1, see Trope, 1975). With fixed values of the two experimental variables (diagnosticity and the conditional probability of the feature given the hypothesis) and assuming equiprobable traits, Equation 1 was solved for the conditional probability of the feature under the alternative. Specifically, D was set at 1.25 for low diagnosticity and at 3.75 for high-diagnosticity features. The probability of the feature given the hypothesized trait was set at four levels: .2, .4, .6, and .8—from highly improbable to highly probable, given the hypothesized trait. Table 1 presents the resulting conditional probabilities of the features under the hypothesized trait, $P(F/H)$, and under the alternative trait, $P(F/A)$, for the eight kinds of features. These conditional probabilities comprised, of course, the percentages presented in the dark areas of the diagrams.

To avoid any confounding of the experimental variables by the a priori meanings of the handwriting features, the assignment of the handwriting features to levels of probability under the hypothesized trait and levels of diagnosticity was counterbalanced across subjects according to an 8 × 8 Latin square design.

Table 1
Reported Conditional Probabilities of Nine Features Given the Hypothesized-Trait Category and the Alternative-Trait Category

Diagnosticity of feature	Probability of feature under hypothesized trait			
	VL	ML	MH	VH
Low				
$P(F/H)$	.200	.400	.600	.800
$P(F/A)$	.278	.510	.490	.720
High				
$P(F/H)$	.200	.400	.600	.800
$P(F/A)$	.778	.890	.110	.220

Note. $P(F/H)$ = probability of feature under hypothesized trait; $P(F/A)$ = probability of feature under alternative trait; VL = very low; ML = moderately low; MH = moderately high; VH = very high.

Each subject thus considered eight handwriting features, each accompanied by a bar diagram and a rating scale. The features were presented in a random order that was varied across subjects. Subjects were instructed as follows:

Please indicate the extent to which you would like to use each of these features for determining whether or not any given student from the group is predominantly analytic (intuitive). That is, you are asked to indicate the extent to which you want to be informed whether or not each feature appears in a student's handwriting for the purpose of deciding whether or not the student is analytic (intuitive).

Subjects indicated their preferences for each feature on an 11-point scale ranging from "not at all interested" (0) to "very interested" (10).

Manipulation check. To avoid an effect of the manipulation check on the dependent measure or vice versa, a separate group of 35 subjects was used to obtain a check on the manipulation of diagnosticity. The procedure for this group was identical to that of the experimental group except that they were not provided with a hypothesis, but were asked instead to judge the extent to which each feature discriminated between predominantly analytic and predominantly intuitive students. They indicated their judgments on an 11-point scale ranging from "not at all" (0) to "extremely well" (10). After rating the features, all subjects received an explanation of the purpose of the study and the need for the use of fictitious data regarding the handwriting features.

Results

The data provided by the manipulation-check group showed that subjects were able

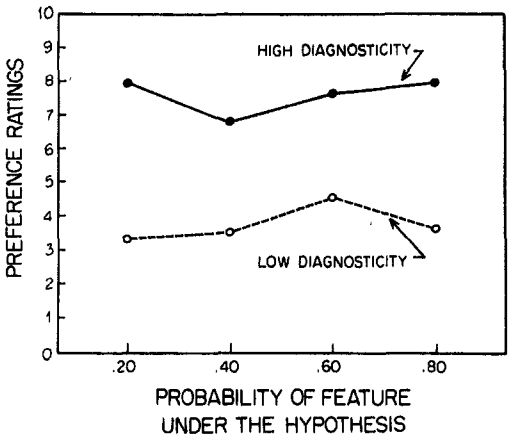

Figure 1. Mean preference ratings of eight features as a function of diagnosticity and probability of the feature given the hypothesis (Experiment 1).

to assess properly the diagnosticity of the handwriting features. Specifically, the mean ratings of the extent to which the features discriminated between analytic and intuitive persons were higher for high-diagnosticity features (7.01) than for low-diagnosticity features (2.95). Diagnosticity was the only significant source of variation in the analysis of these data, $F(1, 29) = 34.35$, $p < .001$.

Figure 1 presents mean preference ratings for the eight kinds of features. These data were subjected to an analysis of variance with hypothesis (analytic hypothesis vs. intuitive hypothesis) as a between-subjects factor and diagnosticity of the feature and its probability under the hypothesis as within-subjects factors. The analysis yielded a highly significant main effect of diagnosticity, $F(1, 73) = 91.36$, $p < .001$. As Figure 1 clearly shows, subjects preferred high-diagnosticity features ($M = 7.34$) to low-diagnosticity features ($M = 3.73$). Furthermore, this effect did not depend on whether the analytic or the intuitive hypothesis was being tested ($F < 1$ for the Diagnosticity × Hypothesis interaction).

Probability of the feature under the hypothesis also produced a statistically significant effect, $F(3, 219) = 3.76$, $p < .05$. However, this effect was inconsistent in that it was not a monotone function of the feature's probability under the hypothesis; instead, as this probability increased, the ratings initially decreased and then increased. Also, the effect of probability under the hypothesis was not uniform across diagnosticity levels, resulting in a significant Probability × Diagnosticity interaction $F(3, 219) = 3.36$, $p < .05$. It should be noted, however, that the magnitude of these effects was extremely small in comparison to that of diagnosticity. Specifically, diagnosticity accounted for .97 of the total variance (including error variance), whereas probability under the hypothesis and the Probability × Diagnosticity interaction accounted for only .008 and .005 of the variance, respectively.

Discussion

These findings are clearly at odds with the proposition that in testing their hypotheses, people prefer questions about evidence to the extent that the evidence is probable under the hypothesis. Such a strategy would suggest that if, for example, the information gatherer believes slanted handwriting is associated with analytic personality and curved handwriting with intuitive personality, then to test the analytic hypothesis he or she will gather information on slanting, regardless of its association with the alternative category of analytic personality; and to test the intuitive hypothesis, he or she will gather information on curvature, regardless of its association with analytic personality.

Subjects' preferences showed no trace of such a strategy. Features that were highly improbable under the hypothesized trait were included in questions if they discriminated between the hypothesis and the alternative. Thus, subjects preferred to ask about evidence with the weakest link to the hypothesized trait than to ask about evidence with the strongest link to the hypothesized trait if the former was more diagnostic. Guided almost exclusively by the diagnosing strategy, subjects' information-gathering preferences properly utilized their knowledge about behavioral manifestations of the hypothesized trait as well as those of the alternative trait.[2]

[2] It might be argued that subjects took into account the information about both trait categories because they were unclear as to which constituted the hypothesis.

This is not to say that Equation 1 is a totally accurate model of subjects' judgments. Various violations of this Bayesian model are conceivable. For the present purposes, the important point is that information gathering was sensitive to the differential association of the features with the hypothesized- and alternative-trait categories. Nevertheless, it is quite possible that subjects used some rule other than a weighted sum of likelihood ratios to combine the conditional probabilities with diagnosticity judgments. A weighted sum of differences between conditional probabilities is a plausible example of such a rule. Indeed, substituting differences for ratios in Equation 1 yields a set of diagnosticity values that quite accurately reproduce the order of mean preferences for the eight features. Furthermore, these predicted values are perfectly correlated with the order of mean ratings from the manipulation-check group regarding the extent to which the features discriminate between the hypothesis and the alternative. The fit between these ratings and the predicted values on the one hand, and the preferences on the other hand, clearly suggests that the sole determinant of subjects' information-gathering preferences was the perceived diagnostic value of the evidence. Although the minor effects of the feature's probability under the hypothesis are not predicted by diagnosticity as formulated in Equation 1, they can be accounted for by variation in subjective assessments of the features' diagnosticity.

Experiment 2

Like Experiment 1, this experiment examined preferences among questions about features varying in diagnosticity and conditional probability under the hypothesis. However, in Experiment 1 the total probability of the features (i.e., the average of their conditional probabilities under the two categories) was free to vary, being lowest (.239) for the feature with low diagnosticity and .2 probability under the hypothesis and highest (.76) for the feature with low diagnosticity and .8 probability under the hypothesis. Preferences among features may depend in some way on the total probability of the features. To eliminate any such effect, the total probability of the features was held constant at .5. Further, to assess the effect of diagnosticity over a wider range of values, the design included three rather than two levels of diagnosticity.

Table 2
Reported Conditional Probabilities of Six Experimental Features Given the Hypothesized-Trait Category and the Alternative-Trait Category

	Probability of feature under hypothesis			
	More probable under alternative		More probable under hypothesis	
Diagnosticity	*P*(F/H)	*P*(F/A)	*P*(F/H)	*P*(F/A)
Low	.45	.55	.55	.45
Intermediate	.30	.70	.70	.30
High	.15	.85	.85	.15

Note. *P*(F/H) = probability of feature under hypothesized trait; *P*(F/A) = probability of feature under alternative trait.

Method

Subjects

The subjects were 89 undergraduate students from the same population used in Experiment 1.

Procedure

The procedure was identical to that of Experiment 1, except for changes in the reported percentage of occurrence of the features under the two categories. These percentages were designed to establish three levels of feature diagnosticity (low, intermediate, and high) and to present within each level of diagnosticity a feature that is either more probable under the hypothesis than under the alternative or vice versa. The total probability was held constant at .5. Table 2 presents the conditional probabilities for each of the features. These probabilities were presented in bar diagrams as in Experiment 1.

The handwriting features employed in this experiment were (a) emphasis on lower zone, (b) parallel lines, (c) printing, (d) threadlike connections, (e) variability in writing angle, and (f) variation in margin. The assignment of these features to combinations of the experimental variables was counterbalanced across subjects

This possibility seems unlikely because subjects in this and subsequent experiments were repeatedly informed what the hypothesis was and were instructed to test it. Moreover, the preference measure asked subjects to indicate the extent to which they were interested in using each feature for testing the hypothesis.

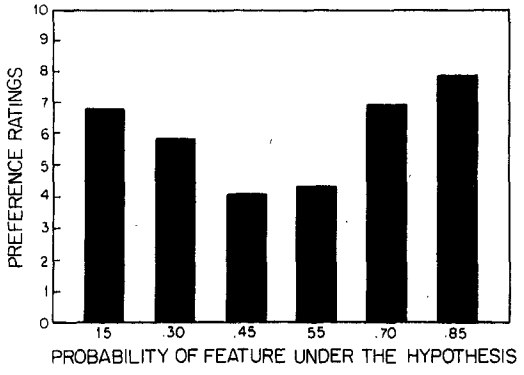

Figure 2. Mean preference ratings of six features as a function of diagnosticity and probability of the feature given the hypothesis (Experiment 2).

according to a 6 × 6 Latin square design. Fifty-nine subjects rated their preference for the features in a random order that was varied across subjects. As in Experiment 1, a separate group of 30 subjects (manipulation-check group) indicated the extent to which the features discriminated between the two categories.

Results and Discussion

In the analysis of variance of the diagnosticity ratings provided by the manipulation-check group, diagnosticity was the only significant source of variation, $F(1, 29) = 64.00$, $p < .001$. These ratings increased from the low-diagnosticity features ($M = 3.21$) through the intermediate-diagnosticity features ($M = 6.15$) to the high-diagnosticity features ($M = 8.59$). Thus, as in Experiment 1, subjects were able to extract diagnosticity information from the diagrams accompanying the features.

Figure 2 presents mean preference ratings of the features as a function of their probability under the hypothesized trait—from a probability of .15 to a probability of .85. The probability of each of the three features on the right side is greater under the hypothesis than under the alternative, whereas the probability of each of the three features on the left side is greater under the alternative than under the hypothesis. For both kinds of features, diagnosticity increases as the probability under the hypothesis becomes more extreme, so that the features located at the center of Figure 2 are the least diagnostic and those located at the extreme ends are the most diagnostic.

It is evident that diagnosticity was the dominant factor in subjects' preferences among the features, $F(2, 116) = 25.24$, $p < .001$. Independent of whether the analytic or intuitive hypothesis was to be tested, subjects' preferences increased from the low- ($M = 4.18$) through intermediate- ($M = 6.42$) to high-diagnosticity features ($M = 7.37$). Figure 2 also shows that within each level of diagnosticity, there is preference for the feature that is more probable under the hypothesized trait, $F(1, 58) = 9.04$, $p < .01$. However, the difference is not significant for low-diagnosticity features and is clearly of secondary importance in comparison to diagnosticity.

These results agree with those obtained in Experiment 1, in that diagnosticity proved to be the prime determinant of the information gathered for testing a hypothesis about the personality of another person. The linkage between the feature and the hypothesized trait had a more consistent effect in Experiment 2, but it still remained of minor importance in controlling subjects' preferences. If features were selected according to their linkage to the hypothesized trait, preference would have been directly related to the features' probability under the hypothesis. However, Figure 2 clearly shows that preferences first decreased as diagnosticity decreased and only then increased, when diagnosticity started increasing. Whether probable or improbable under the hypothesized traits, a diagnostic feature was always preferred to a nondiagnostic one. In this respect, subjects' preferences are consistent with a diagnosing strategy where preference for features probable under the hypothesis does not entail loss in diagnostic power.

Experiment 3

This experiment is concerned with preferences among continuous features. For such features, each trait category can be characterized in terms of the range of values over which it is distributed. For any given pair of ranges of such values, diagnosticity is an increasing function of the interval over which the ranges of the two categories do not overlap. This nonoverlapping interval is a sum of the nonoverlapping invervals under the

two categories (see Figure 3). The magnitude of the nonoverlapping interval under each category, that is, the range of values whose presence confirms the category (and disconfirms the alternative) determines its diagnosability. Furthermore, the nonoverlapping interval of one category can be varied independently of the nonoverlapping interval of the other category. Consider, for example, the two continuous features in Figure 3 where the ranges of the hypothesis and the alternative are partially overlapping. The two features have exactly the same total nonoverlapping interval. However, in Feature A the hypothesis is highly diagnosable but the alternative is not, whereas in Feature B the alternative is highly diagnosable but the hypothesis is not.

According to Snyder and Swann's confirming strategy, people concentrate on the hypothesis, neglecting the alternative, and prefer features that can confirm the hypothesis to features than can disconfirm it. Under this strategy, the information gatherer should prefer Feature A, which can confirm the hypothesis but not disconfirm it, to Feature B, which can disconfirm the hypothesis but not confirm it. In general, among features varying orthogonally in diagnosability of the hypothesis and diagnosability of the alternative, preference should increase with the diagnosability of the hypothesis and be unrelated to the diagnosability of the alternative. In contrast, under the diagnosing strategy, preference depends on total diagnosticity, which combines the diagnosability of the hypothesis and the alternative. Hence, preference should be an additive increasing function of the diagnosability of the hypothesis and the diagnosability of the alternative (see Trope, 1980, for a more detailed discussion of such sources of information).

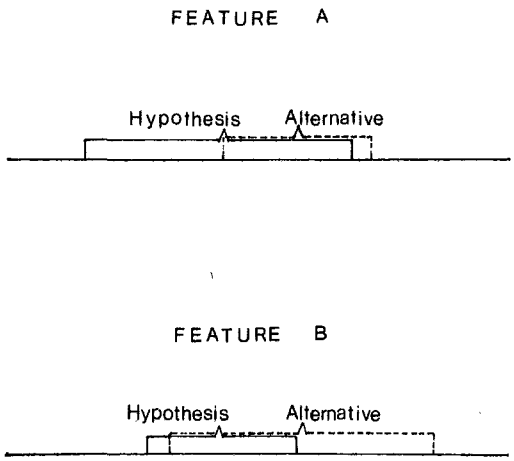

Figure 3. Ranges of scores of the hypothesized and alternative trait categories on two features. (On Feature A, the hypothesized trait is highly diagnosable but the alternative is not. On Feature B, the alternative trait is highly diagnosable but the hypothesized trait is not.)

Table 3
Reported Ranges of Percentages of Nine Features Given the Hypothesized-Trait Category and the Alternative-Trait Category

Diagnosability of hypothesis	Diagnosability of alternative		
	Low	Moderate	High
Low			
Hypothesis	40–77[a]	49–78	44–65
Alternative	43–80[b]	38–75	47–84
Intermediate			
Hypothesis	45–82	40–69	57–78
Alternative	42–71	51–80	38–67
High			
Hypothesis	36–73	53–82	59–80
Alternative	55–76	42–63	40–61

Note. The nonoverlapping intervals under the hypothesized- and the alternative-trait categories are 3, 11, and 19 for low, intermediate, and high trait-diagnosability, respectively.
[a] Reported ranges of the percentages with which the feature appears in the handwriting of people who belong to the hypothesized-trait category.
[b] Reported ranges of the percentages with which the feature appears in the handwriting of people who belong to the alternative-trait category.

Method

Subjects

The subjects were 143 visitors to the Ontario Science Center in Toronto, who volunteered to take part in a study on handwriting and personality.

Procedure

This experiment followed the same procedure as Experiments 1 and 2, except for the use of continuous features and changes in information about them. Specifically, the features were defined in terms of the percentage of times they appeared in the person's handwriting (e.g., the percentage of slanted letters out of all letters examined). Subjects were presented with fictitious information about the percentage of time the feature appeared in the handwriting of analytic and intu-

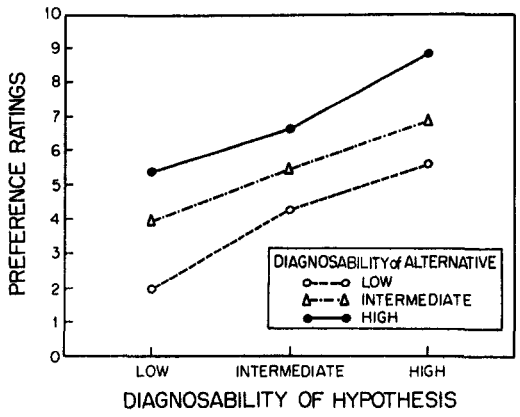

Figure 4. Mean preference ratings of nine features as function of diagnosability of hypothesized trait and diagnosability of alternative trait.

itive people. Specifically, for each trait category, subjects were shown the range of percentage scores within which 90% of the people fell. Diagnosability of the hypothesis and diagnosability of the alternative were manipulated by varying the sizes of the nonoverlapping intervals under the ranges of the respective trait categories. Thus, the nonoverlapping interval under the range of the hypothesized trait was either 3% (low diagnosability), 11% (intermediate diagnosability), or 19% (high diagnosability). Three levels of diagnosability of the alternative were similarly generated. An orthogonal manipulation of these two factors resulted in nine kinds of features. The ranges of the hypothesized and alternative traits on each of the features were presented in diagrams similar to those in Figure 3. Table 3 presents these ranges for each of the nine features. The overall mean and range (from the lowest to the highest limit) of the two distributions were held constant across all features.

The following handwriting features were used in this experiment: (a) connected letters, (b) flourished letters, (c) letters with deviations from standard form, (d) letters with emphasis on lower zone, (e) parallel lines, (f) printed letters, (g) slanted letters, (h) words with correction, and (i) words with unessential elements eliminated. The assignment of these features to levels of the experimental variables was counterbalanced across subjects according to a Latin square design. The positions of the ranges of the hypothesis and the alternative (i.e., the lower range vs. the upper range) were randomly varied across features. The order in which the features were presented was varied randomly over subjects. Half of the subjects tested the intuitive hypothesis and the other half tested the analytic hypothesis. Finally, as in the other two experiments, a separate group of 50 subjects rated the extent to which each feature discriminated between the two trait categories.

Results and Discussion

The ratings provided by the manipulation-check group showed unequivocally that the perceived overall diagnosticity of the features (i.e., the extent to which they discriminate between the two trait-categories) was an additive function of the diagnosability of the analytic-trait category, $F(2, 82) = 39.87$, $p < .001$, and the diagnosability of the intuitive-trait category, $F(2, 82) = 43.00$, $p < .001$. The overall means of the low, intermediate, and high levels of the analytic-trait category were 3.65, 5.01, and 7.15, respectively, and 4.13, 5.30, and 6.95, respectively, for these levels of the intuitive-trait category.

Figure 4 presents mean preference ratings of the nine features. These data bear a striking resemblance to the subjective diagnosticity data. First, preferences for a feature increased quite regularly with diagnosability of the hypothesis, $F(2, 182) = 102.35$, $p < .001$. Second, and in direct contradiction of Snyder and Swann's confirming strategy, preferences increased no less systematically with the diagnosability of the alternative, $F(2, 182) = 86.73$, $p < .001$. That is, although subjects preferred features to the extent that they could show the hypothesis to be true, they also preferred features that could show the hypothesis to be false. In fact, the two kinds of diagnosability seemed to carry the same weight in choice, as evidenced by the fact that there was no instance where subjects preferred a feature in which the hypothesis was more diagnosable than the alternative to a feature in which the alternative was more diagnosable than the hypothesis or where the two were equally diagnosable. The only determinant of choice was the overall diagnosticity of the feature (i.e., the diagnosability of the hypothesis and the alternative combined).

General Discussion

The three studies provide strong evidence that people gather information by what we termed the *diagnosing strategy*. That is, to test the hypothesis that a target person belongs to one trait-category rather than to another, the information gatherer asks about features that maximally discriminate between the two categories. Our experiments found very little evidence for interest in questions about features whose presence tends to confirm rather than disconfirm the hypoth-

esis, as the *confirming strategy* postulated by Snyder and Swann (1978) requires. There was a slight tendency to prefer questions about features that are more probable under the hypothesized trait to questions about features that are more probable under the alternative trait. However, this preference was weak and inconsistent. Even in Experiment 2 where this preference was clearest, it was secondary to the preference for questions about diagnostic features. The findings of Experiment 3 provide perhaps the strongest evidence against the confirming strategy: A question that could only confirm the hypothesis was not preferred to a question that could only disconfirm the hypothesis. Both of these kinds of questions were less attractive than questions that could both confirm and disconfirm the hypothesis.

These data appear to be at variance with Snyder and Swann's studies demonstrating the confirmatory strategy. In his discussion of these studies, Snyder (1981) writes:

> Is there any procedure that will induce individuals to avoid the preferential soliciting of evidence from one behavioral domain? Yes. But, oh, the lengths to which one must go to accomplish this feat. The only procedure that successfully induces individuals to avoid the preferential soliciting of behavioral evidence from one domain is one that provides them with no hypothesis to test. (p. 289)

Obviously, the one procedure that Snyder and Swann did not test was varying the diagnosticity of the evidence. But there are other differences between the procedure of the present study and those employed by Snyder and Swann.[3] First and foremost, in our study subjects received descriptions of both the hypothesized trait and the alternative trait. In contrast, in all of the studies conducted by Snyder, Swann, and their associates (see Snyder, 1981), subjects were provided with a rich, vivid profile of the hypothesized trait but were never provided with a profile of the alternative trait.[4] In fact, Snyder and Swann never even *mentioned* the existence of an alternative trait category.

Second, Snyder and Swann's subjects' attention was further directed to the hypothesis by the instruction to "choose 12 questions that will help you link the general characteristics in the profile with the person's likes, dislikes, and behaviors." These instructions never asked subjects to test whether or not the target was, say, an extrovert. In contrast, the present study explicitly instructed subjects to choose features for the purpose of determining whether or not the target belonged to the hypothesized-trait category.

Third, the kind of questions Snyder and Swann asked subjects to choose from could have further strengthened the impression that their task was not that of testing whether or not the target was, say, an extrovert. Specifically, Snyder and Swann's "extrovert questions" (e.g., "What would you do if you wanted to liven things up at a party?" and "What do you like about living situations in which there are always lots of people around?") do not seem to test whether or not the target is an extrovert, nor do their "introverted questions" (e.g., "What do you dislike about loud parties?" and "Think about times when your shyness in social situations has made you come across as being aloof. Give me an example.") seem to test whether or not the target is an introvert. These questions are clearly biased against the expression of answers that are indicative of the alternative trait. To be indicative of the alternative, the target would have to reject the erroneous assumption on which the question is based (e.g., "Usually, I don't want to liven things up at a party."). Snyder and Swann's questions seem to assume that the target is, for example, an extrovert, without allowing answers that can falsify the assumption. Such questions are nondiagnostic because they force introverts to answer *as if* they were extroverts. This possibility may explain Snyder and Swann's (1978; Experiment 2) provocative finding that subjects judged interviewees who answered predominantly extrovert questions to

---

[3] Thanks are due to Mark Synder for providing us with the experimental materials.

[4] Snyder (1981) reports an unpublished study by Snyder and Swann that provided subjects with both the descriptions of prototypical introvert and prototypical extrovert. However, this study is totally irrelevant to hypothesis testing since subjects were not provided with a hypothesis. Instead, subjects were instructed to "find out if the other person is more like an extrovert or more like an introvert." (Snyder, 1981, p. 286).

be more extroverted than interviewees who answered predominantly introvert questions.

Together, the detailed description of the hypothesized trait without even mentioning the alternative, the explicit instruction to link the description of the hypothesized trait to the target's characteristics, and the biased nature of the questions might have convinced subjects that their task was to concentrate on the hypothesized trait and perhaps check for its manifestations or make discriminations within it. For example, without being told what the alternatives are (if any) to the profile of the prototypical extrovert, subjects might have guessed that the alternatives included one or more levels or types of extroversion that were different from (perhaps more moderate than) the prototypical profile provided. Clearly, for discriminating between levels or types of extroversion, questions that discriminate in the introvert category are useless. Thus, while both the questions called by Snyder and Swann hypothesis-confirming questions and those they called hypothesis-disconfirming questions were undiagnostic for distinguishing between introverts and extroverts, the hypothesis-confirming questions might have been simply more diagnostic for distinguishing among types or levels of the hypothesized trait. It is not surprising, then, that subjects repeatedly preferred these questions.

There is additional evidence that subjects sought diagnostic questions in Snyder and Swann's own studies. Specifically, in addition to extrovert and introvert questions, Snyder and Swann presented five neutral questions that were either classified by judges as irrelevant or could not be classified as extrovert or introvert. Combining Snyder and Swann's four experiments, the mean number of neutral questions selected was 1.94, which is quite high given that at least one neutral question seems quite irrelevant (i.e., "What kind of charities do you like to contribute to?"). Of the remaining questions, one seems highly diagnostic ("What do you think the good and bad points of acting friendly and open are?") and others moderately diagnostic ("What are your career goals?"). We suggest that the frequent choice of these diagnostic, unbiased questions may account for the relative popularity of the neutral questions.

Further support for the preceding analysis was provided by a recent study by Trope and Alon (Note 1), which followed Snyder and Swann's procedure with the exception that the task was to discover in the interview whether the target was an introvert or extrovert and that subjects were free to write by themselves as many questions as they wished for the interview. A total of 286 questions generated by 38 subjects were classified with 92% agreement by two judges into (a) *consistent questions* that asked about a feature derived from the hypothesized trait, (b) *inconsistent questions* that asked about a feature derived from the alternative trait, (c) *biased questions* that already assumed the target possessed one trait and did not allow an answer indicative of the other trait, and (d) *nondirectional questions* that either presented a choice between an extrovert and introvert alternative or were open-ended (e.g., "What are your hobbies?"). Out of the 286 questions, not a single biased question was found. It seems, then, that in their experiments, Snyder and Swann forced a choice among questions that subjects would never spontaneously ask. The great majority of the questions (73%) were nondirectional. Most significant, the difference between the percentage of consistent questions (15%) and inconsistent questions (12%) was negligible. The group that tested the introvert hypothesis did not generate more introvert questions or less extrovert questions than the group that tested the extrovert hypothesis.

The nature of information gathering demonstrated in Trope and Alon (Note 1) and in this article is fundamentally different from that portrayed by Snyder, Swann, and their colleagues. The former studies suggest that information gatherers know to avoid test situations or questions that are biased against the alternative-trait category. In and of itself, asking about features derived from the hypothesis does not necessarily result in confirmation of the hypothesis, since the answers (e.g., absence of features) may disconfirm the hypothesis. Even questions about low-diagnosticity features that are derived from the hypotheses do not necessarily pro-

duce hypothesis confirmation. The use of such questions may result in an inefficient information-gathering strategy in the sense that a large number of questions will be required to obtain the same level of certainty about the validity of the hypothesis. However, the present article shows that even when the questions focus on features that are diagnostic of the hypothesized trait, their absence is diagnostic of the alternative trait. Consequently, not only did our subjects avoid questions biased in favor of the hypothesis, but the cognitive ease of phrasing questions in terms of features derived from the hypothesis did not produce a loss in diagnostic power of the information-gathering procedure.

It should be noted that the hypothetical tasks used in our experiments do not fully represent the complexity of social information gathering (see Festinger, 1954; Goffman, 1959; Jones, 1964). This complexity results not only from the cognitive processes engaged in construction of diagnostic questions, but also from the fact that the information gatherer has to take into account interpersonal constraints, social norms, and the consequences of the question and its answer for the information gatherer, the target, and various others. Thus, information gathering, unlike information processing, frequently entails social action. In this respect, research on information gathering may help bridge the theoretical gap between cognition and action in social situations.

Reference Note

1. Trope, Y., & Alon, E. *Testing hypotheses about the personality of another person.* Unpublished manuscript, University of Toronto, 1980.

References

Ajzen, I., & Fishbein, M. A Bayesian analysis of attribution processes. *Psychological Bulletin*, 1975, *82*, 261–277.

Einhorn, J. J., & Hogarth, R. M. Judgment and decision. *Annual Review of Psychology*, 1981, *32*, 53–88.

Festinger, L. A theory of social comparison process. *Human Relations*, 1954, *7*, 117–140.

Goffman, E. *The presentation of self in everyday life.* Garden City, N.J.: Doubleday, 1959.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

Jones, E. E. *Ingratiation: A social psychological analysis.* New York: Appleton-Century-Crofts, 1964.

Jones, E. E., Davis, K. E., & Gergen, K. J. Role playing variations and their informational value for person perception. *Journal of Abnormal and Social Psychology*, 1961, *63*, 302–310.

Kahneman, D., Slovic, P., & Tversky, A. *Judgment and uncertainty: Heuristics and biases.* New York: Cambridge University Press, 1982.

Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), *Nebraska Symposium on Motivation* (Vol. 15). Lincoln: University of Nebraska Press, 1967.

Kelley, H. H. Attribution in social interaction. In E. E. Jones, D. E. Kanouse, H. H. Kelley, R. E. Nisbett, & B. Weiner (Eds.), *Attribution: Perceiving the causes of behavior.* Morristown, N.J.: General Learning Press, 1972.

Nisbett, R. E., & Ross, L. *Human inference: Strategies and shortcomings of social judgment.* Englewood Cliffs, N.J.: Prentice-Hall, 1980.

Rosenberg, S., & Sedlak, A. Structural representation of implicit personality theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6). New York: Academic Press, 1972.

Ruble, D. N., & Boggiano, A. K. Optimizing motivation in an achievement context. In B. Deogh (Ed.), *Advances in special education.* Greenwich, Conn.: JAI Press, 1980.

Schneider, D. J. Implicit personality theory: A review. *Psychological Bulletin*, 1973, *79*, 294–309.

Snyder, M. Seek and ye shall find: Testing hypotheses about other people. In E. T. Higgins, C. P. Heiman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium on personality and social psychology.* Hillsdale, N.J.: Erlbaum, 1981.

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

Snyder, M., & Swann, W. B. Hypothesis-testing processes in social interaction. *Journal of Personality and Social Psychology*, 1978, *36*, 1202–1212.

Steiner, I. D. Perceived freedom. In L. Berkowitz (Ed.), *Advances in experimental and social psychology* (Vol. 5). New York: Academic Press, 1970.

Strickland, L. H. Surveillance and trust. *Journal of Personality*, 1958, *26*, 200–215.

Thibaut, J. H., & Riecken, H. W. Some determinants and consequences of the perception of social causality. *Journal of Personality*, 1955, *24*, 113–133.

Trope, Y. Inferential processes in the forced compliance situation. *Journal of Experimental Social Psychology*, 1974, *10*, 1–15.

Trope, Y. Seeking information about one's own ability as a determinant of choice among tasks. *Journal of Personality and Social Psychology*, 1975, *32*, 1004–1013.

Trope, Y. Extrinsic rewards, congruence between dispositions and behaviors, and perceived freedom. *Journal of Personality and Social Psychology*, 1978, *36*, 588–597.

Trope, Y. Uncertainty-reducing properties of achievement tasks. *Journal of Personality and Social Psychology*, 1979, *37*, 1505–1518.

Trope, Y. Self-assessment, self-enhancement, and task preference. *Journal of Experimental Social Psychology*, 1980, *16*, 116–129.

Trope, Y., & Brickman, P. Difficulty and diagnosticity as determinants of choice among tasks. *Journal of Personality and Social Psychology*, 1975, *31*, 918–925.

Trope, Y., & Burnstein, E. Processing the information contained in another's behavior. *Journal of Experimental Social Psychology*, 1975, *11*, 439–458.

Trope, Y., & Burnstein, E. A disposition-behavior congruity model of perceived freedom. *Journal of Experimental Social Psychology*, 1977, *13*, 357–368.

Tversky, A., & Kahneman, D. Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), *Progress in social psychology* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1980.

Received December 9, 1980
Revision received October 29, 1981 ■

Manuscripts Accepted for Publication in the Section Attitudes and Social Cognition

Gender Stereotypes as a Product of Inferred Sex Differences in Status: Stereotypes About Social Influence. Alice H. Eagly (Department of Psychology, Purdue University, West Lafayette, Indiana 47907) and Wendy Wood.

Different Levels of Cognitive Dissonance Information Seeking and Information Avoidance. Dieter Frey (University of Kiel, Institut fur Psychologie, 2300 Kiel, West Germany).

Cognitive and Social Bases of Emotional Experience: Outcome, Attribution and Affect in Data from a National Survey. Eliot R. Smith (Department of Sociology, University of California—Riverside, Riverside, California 92521) and James R. Kluegel.

Gender Schema Theory and Self-Schema Theory Compared: A Comment on Markus, Crane, Bernstein, and Siladi's "Self-Schemas and Gender." Sandra Bem (Department of Psychology, Cornell University, Ithaca, New York 14853).

Persistence Following Failure: The Interactive Role of Self-Awareness and the Attributional Basis for Negative Expectancies. Michael H. Kernis (Department of Psychology, University of Rochester, River Station, Rochester, New York 14627), Miron Zuckerman, Alan Cohen, and Susan Spadafora.

An Investigation of the Dimensions of Causal Attribution. Scott Wimer and Harold Kelley (Department of Psychology, University of California—Los Angeles, Los Angeles, California 90024).

Self-Image Versus Impression Management in Adherence to Distributive-Justice Standards: The Influence of Self-Awareness and Self-Consciousness. Jerald Greenberg (Faculty of Management Sciences, Ohio State University, 1775 College Road, Columbus, Ohio 43210).

Integration and Abstraction in Impression Formation. Yaacov Schul (Department of Psychology, Hebrew University of Jerusalem, Jerusalem, Israel).

Person Memory and Causal Attributions. Jennifer Crocker (Department of Psychology, Northwestern University, Evanston, Illinois 60201), Darlene B. Hannah, and Renee Weber.

Prototype Salience and Implicit Personality Theories. David Schneider (Division of Social Sciences, University of Texas, San Antonio, Texas 78285) and Bonnie L. Blankmeyer.

Mediation Among Attributional Inferences and Comprehension Processes: Initial Findings and a General Method. Eliot Smith (Department of Psychology, University of California—Riverside, Riverside, California 92521) and Frederick D. Miller.