Diagnostic and Confirmation Strategies in Trait Hypothesis Testing

Patricia G. Devine, Edward R. Hirt, and Elizabeth M. Gehrke
University of Wisconsin—Madison

The role of diagnostic and confirmation strategies in trait hypothesis testing is examined. The present studies integrate theoretical and empirical work on qualitative differences among traits with the hypothesis-testing literature. Ss tested trait hypotheses from 2 hierarchically restrictive trait dimensions: introversion—extraversion and honesty—dishonesty. In Study 1, Ss generated questions to test trait hypotheses, and diagnosticity was theoretically defined (e.g., questions associated with nonrestrictive ends of trait dimensions). In Study 2, Ss selected questions from an experimenter-provided list in which diagnosticity was empirically defined. In Study 3, Ss chose between 2 equally diagnostic questions. In each of the studies, Ss showed a primary preference for diagnostic information and a secondary preference for confirmatory information. Ss’ preference for diagnostic information suggests that they prefer to ask the most informative questions. The explanation for the confirmation bias is less obvious, and possible reasons for this effect are discussed.

Many researchers within social psychology (e.g., Snyder, 1981; Snyder & Campbell, 1980; Snyder & Swann, 1978; Swann & Giuliano, 1987; Trope & Bassok, 1982, 1983; Trope, Bassok, & Alon, 1984) have investigated the information-gathering strategies that lay interviewers use in testing trait hypotheses about other people (e.g., “Is Fred an extravert?”). This research has revealed two major information-gathering strategies that subjects use when testing trait hypotheses: a confirmation strategy and a diagnostic strategy. However, there has been considerable debate within this literature regarding the relative importance of each strategy. The present research is designed to examine the role of both confirmation and diagnostic strategies in trait hypothesis testing.

Confirmation Strategies

In early hypothesis-testing experiments by Snyder and Swann (1978), subjects were presented with a trait hypothesis (introversion or extraversion) and were asked to select questions from an experimenter-provided list to test the hypothesis. These questions asked about introverted or extraverted features (or were neutral with regard to the introversion–extraversion dimension). Snyder and Swann found that subjects testing the extravert hypothesis, for example, selected significantly more questions that probed for extraversion (e.g., “What would you do if you wanted to liven things up at a party?”) and that subjects testing the introvert hypothesis selected more questions that probed for introversion (e.g., “In what situations do you wish you could be more outgoing?”). Snyder and Swann concluded that this pattern suggested that subjects manifested a hypothesis-confirming strategy. They further suggested that the preference for a hypothesis-confirming strategy was tenacious, persisting even when subjects were promised a $25 reward for accuracy or after subjects received information that it was unlikely that the target person possessed the hypothesized trait (Snyder & Swann, 1978, Experiments 3 and 4).

Findings from a number of other areas also have suggested that there is an overwhelming preference for confirmatory information in hypothesis testing. Writings in philosophy of science (e.g., Lakatos, 1970; Platt, 1964; Popper, 1959, 1970) have argued that scientists tend to design empirical investigations that seek to confirm rather than disconfirm their hypotheses (see also Greenwald, Pratkanis, Leippe, & Baumgardener, 1986). Wason (1960, 1968) found that subjects in a rule discovery task overwhelmingly looked at cases that fit their hypothesized rule. Doherty, Mynatt, Tweney, and Schiavo (1979) and Beyth-Marom and Fischhoff (1983) showed that subjects preferred to seek information about features that were more likely under the hypothesis than under the alternative(s).

Two distinct types of hypothesis-confirming strategies have been identified in the literature. Klayman and Ha (1987) discussed the differences between (a) a positive test strategy and (b) strategies that increase the likelihood that a hypothesis will be protected or preserved (see also Fischhoff & Beyth-Marom, 1983). A positive test strategy involves the tendency to ask questions in which a yes answer would tend to confirm the hypothesis being tested or to choose instances assumed to be true under the working hypothesis (Doherty et al., 1979; Klayman & Ha, 1989; Wason, 1968). However, this strategy does not necessarily lead one to preserve the hypothesis or protect it from falsification. For example, if one is testing the hypothesis that a person is extraverted, a yes answer to a positive test (i.e., extraverted)
question will tend to confirm the hypothesis. However, a no answer (to the same question) will tend to disconfirm the hypothesis.

Hypothesis-preservation strategies, on the other hand, involve asking questions that are likely to lead to the conclusion that the working hypothesis is true. For example, the question might be constraining, so that any answer will necessarily be taken as supporting the hypothesis (Snyder & Swann, 1978). Alternatively, questions might be asked so that the likelihood of a hypothesis-confirming answer is greater than the likelihood of a hypothesis-disconfirming answer (Skov & Sherman, 1986). That is, subjects testing an extraverted hypothesis might prefer to ask a question for which 90% of extraverts and 50% of introverts would answer "yes" to the question, over a question for which 50% of extraverts and 10% of introverts would answer "yes" to the question. For both questions, a yes answer confirms the hypothesis. However, for the 90/50 question, subjects are likely to get a hypothesis-confirming answer 70% of the time, whereas they will get a hypothesis-confirming answer from the 50/10 question only 30% of the time. Thus, the overall likelihood of preserving the hypothesis is greater for the 90/50 question than for the 50/10 question.¹

The Diagnostic Strategy

Trope and Bassok (1982, 1983; Bassok & Trope, 1984) argued that people use a diagnostic strategy rather than a confirmation strategy in trait hypothesis testing. The diagnostic strategy leads subjects to prefer to ask the most diagnostic questions available, those questions that allow subjects to most clearly distinguish between the hypothesis and its alternative(s). In his experiments, Trope provides subjects with a trait hypothesis (e.g., introversion) and asks subjects to select a subset of 12 questions from a list of 24 high and low diagnostic introverted and extraverted questions (6 of each type). If subjects use a confirmation (i.e., positive test) strategy, they should prefer to ask hypothesis-confirming (i.e., introverted) questions regardless of their diagnosticity; if subjects use a diagnostic strategy, they should prefer to ask the high diagnostic questions, regardless of whether they ask about hypothesis-confirming (i.e., introverted) or hypothesis-disconfirming behaviors (i.e., extraverted). The results of these studies illustrated an overwhelming preference for the highly diagnostic questions, suggesting that subjects clearly use a diagnostic strategy in hypothesis testing.

Relative Strengths of the Confirmation and Diagnostic Strategies

Although Trope's work argues convincingly for the diagnostic strategy, several questions are left unresolved concerning the status of the hypothesis-confirming strategy. When the diagnostic strategy was pitted against the hypothesis-confirming strategy, subjects chose the diagnostic strategy. However, the differences in diagnosticity between the high and low diagnostic questions were substantial; thus, all we know for certain is that subjects choose to ask informative questions (cf. Skov & Sherman, 1986). Moreover, in the Trope and Bassok studies, subjects chose a large number of questions (12 out of 24), and the diagnosticity effect was so strong, that it is difficult to

¹ We should note that this analysis does not consider the differential diagnosticity of yes and no answers to the same question. Sliwaczek, Klaiman, Sherman, and Skov (1989) pointed out that yes and no answers to either a 90/50 or a 50/10 question are not equally diagnostic regarding the truth value of the hypothesis. A yes answer to a 90/50 question only weakly supports the hypothesis (because most people would say yes to this question), but a no answer to the same question strongly supports the alternative. Alternatively, a yes answer to a 50/10 question strongly supports the hypothesis, but a no answer only weakly supports the alternative. Sliwaczek et al.'s results suggested that subjects were not sensitive to the differential diagnosticity of answers to such questions; instead, subjects treated the answers as equally diagnostic. These misperceptions of diagnosticity of answers, in addition to subjects' preference for asking questions for which the likelihood of a hypothesis-confirming answer is high, lead to hypothesis-preservation effects.
Exploring Strategy Preferences in the Trait Domain

Both theoretical (Reeder & Brewer, 1979) and empirical (Rothbart & Park, 1986) work on traits suggests that there are qualitative differences among traits that could influence trait hypothesis testing. Attribution theorists have suggested that the relation between behavior and trait inferences varies across traits (cf. Jones & Davis, 1965). Reeder and Brewer attempted to develop a formal framework for specifying how behaviors are related to trait inferences. Specifically, they suggested that the diagnosticity of behaviors for underlying trait dispositions depends on the structure of trait schemata.

Consider the introversion–extraversion trait dimension, which Reeder and Brewer (1979) argued is hierarchically restrictive. Hierarchically restrictive traits are asymmetrical, so that one end of the dimension (extraversion) is associated with a wider range of possible behaviors than is the other end of the dimension (introversion). From this perspective, Reeder and Brewer argued that extraverted persons can engage in both extraverted and introverted behaviors but that introverted persons can only engage in introverted behaviors. Reeder, Messick, and Van Avermaet (1977) presented data consistent with this analysis. That is, subjects perceived introverts to be more behaviorally restricted than extraverts. One consequence of this asymmetry, according to Reeder and his colleagues, is that extraverted behaviors are more diagnostic (allow more confident trait attributions because only extravert can engage in extraverted behaviors) than introverted behaviors.

One can extrapolate from the preceding analysis implications for relative strategy preferences in trait hypothesis testing. Again, consider the trait dimension of introversion–extraversion. If subjects are concerned only with diagnosticity, they should prefer to ask extraverted questions when testing either the introverted or the extraverted hypothesis. This would be revealed by a question type main effect (i.e., preference for extraverted questions). If subjects are concerned only with confirmation, they should prefer to ask questions consistent with the working hypothesis (e.g., introvert questions for the introverted hypothesis). This would be revealed by a significant Trait Tested × Question Type interaction effect with no main effects. However, if Skov and Sherman (1986) were correct, we would expect a strong question type main effect (primary preference for diagnosticity) and a weaker but significant Trait Tested × Question Type interaction (secondary preference for confirmatory information).

In the present research, we explore two different trait dimensions: introversion–extraversion and honesty–dishonesty. The trait dimension of introversion–extraversion has been examined in every trait-hypothesis-testing study to date; thus, if for no other reason, inclusion of this trait dimension is necessary to replicate and extend previous work. In addition, the trait dimension of introversion–extraversion is theoretically interesting because it is hierarchically restrictive and thus has the properties outlined earlier. As a trait dimension replication variable, we examine the trait dimension of honesty–dishonesty, which Reeder and Brewer (1979) also argued is hierarchically restrictive. In contrast to introversion–extraversion, people at the more positive end of the honesty–dishonesty dimension (honesty) are behaviorally restricted, and people at the more negative end (dishonesty) are not behaviorally restricted. Consequently, for the honesty–dishonesty dimension, behaviors associated with dishonesty would be the more diagnostic behaviors. Indeed, empirical data suggest that dishonest (immoral) behaviors are more diagnostic than honest (moral) behaviors (McGraw, 1985; Reeder & Coover, 1986; Skowronski & Carlson, 1987, 1989).

Study 1

The focus of Study 1 is to empirically validate the theoretical implications of Reeder and Brewer's analysis for trait hypothesis testing. In this way, the present work extends previous work in the impression formation and trait attribution literatures, which has suggested that behaviors associated with restricted and nonrestricted ends of trait dimensions are differentially diagnostic (see Reeder, 1985, and Skowronski & Carlson, 1989, for reviews). In Study 1, we examine subjects' spontaneous hypothesis-testing strategies in a question generation task. Two previous studies examined the types of questions subjects spontaneously generate to test introversion and extraversion (Swann & Giuliano, 1987; Trope, Bassok, & Alon, 1984). Subjects were given a trait hypothesis and asked to write questions (the format of which was open) that could be used to test whether a person possessed the particular trait. In Swann and Giuliano's (1987) study, the majority (70%) of questions subjects prepared were posed in a yes/no format. As such, the questions could be identified as hypothesis true (phrased so that a yes answer indicated the presence of the hypothesized trait) or alternative true (phrased so that a yes indicated the presence of the alternative trait). In Trope et al.'s (1984) study, the majority of questions were bidirectional (phrased so that a choice must be made between behaviors consistent with the hypothesized or the alternative trait). However, subjects in the Trope et al. study also generated many hypothesis true and alternative true questions. An interesting pattern is revealed when one examines the yes/no questions generated in both studies. Although, in general, when subjects were testing either introversion or extraversion, they showed a tendency to increase the number of hypothesis true questions they generated (as evidenced in a Trait Tested × Question Type interaction), subjects still showed an overall preference for generating extraverted questions (the question type main effect). Swann and Giuliano (1987) interpreted these findings as supporting confirmatory hypothesis-testing strategies. However, the pattern of their results is also consistent with a diagnostic strategy (e.g., the overall preference for extraverted questions). Moreover, Swann and Giuliano found an identical pattern of question generation for the trait dimension of dominance and submissiveness. This trait was also characterized by Reeder and Brewer (1979) as hierarchically restrictive with submissiveness at the restricted end of the dimension. Note that for both trait dimensions explored by Swann and Giuliano, the restricted ends of the dimensions are the more negative traits. Thus, positivity/negativity is confounded with the restrictiveness/nonrestrictiveness of the trait dimensions.
In the present study, we examine the relative preferences for
diagnostic and confirmation strategies in trait hypothesis test-
ing, controlling for the effects of positivity and negativity. That
is, although as in previous trait-hypothesis-testing research
we use two hierarchically restrictive trait dimensions, the re-
stricted end of one trait dimension is negative (i.e., introversion),
and the nonrestricted end is positive (i.e., extraversion). For the
other trait dimension, the restricted end is positive (i.e., honest)
and the nonrestricted end is negative (i.e., dishonest). Previous
work by Reeder (1985) and Skowronski and Carlson (1987)
would suggest that the restrictiveness/nonrestrictiveness of the
trait dimension, rather than the positivity/negativity, will deter-
mine subjects' question generation.

Method

Subjects. Forty-three University of Wisconsin—Madison intro-
ductive psychology students participated in exchange for extra
credit.

Procedure. Subjects were told that the study concerned impression
formation. They were further told that when a person meets others, he
or she often tries to form impressions of them and that one means by
which a person obtains information about others is to ask questions of
the persons directly or to ask questions of their acquaintances. For each
of the two different traits, subjects were asked to generate at least six
questions they would ask to determine whether another person pos-
sessed the particular trait. Below each trait were a number of boxes in
which subjects were to record their questions.

Subjects were presented with either introversion or extraversion
and either honesty or dishonesty. Half of the subjects got the introversion-
extraversion dimension first, whereas the other half responded to the
honesty–dishonesty dimension first. In all of the studies to be re-
ported, whenever subjects worked with the traits introversion and ex-
traversion, they were provided with the definitions of the typical intro-
vert and the typical extravert used by Snyder and Swann (1978).Subjects
had a very clear understanding of the definitions of honesty and dishonesty,
so we did not include definitions of these traits in any of
our studies.

After generating questions, subjects were debriefed and thanked for
their participation.

Categorization of questions. Two judges coded the questions. Possi-
ble coding categories were suggested to us that were based on the
classification schemes developed by Swann and Giuliano (1987) and
Trope et al. (1984). The coding categories used in the present study
included the following:

Yes/no questions. These questions were phrased such that a respond-
ent could answer yes or no to the questions (e.g., "Are you usually
the initiator in forming new relationships?" "Would you lie in a court of
law?"). Yes/no questions can be identified as hypothesis true or alter-
native true, depending on the hypothesis being tested. Some of the
yes/no questions, however, were ambiguous (e.g., "Do you find others
interesting?"). With ambiguous yes/no questions, the relation of the
question to either the hypothesized or the alternative trait was not clear.

Bidirectional. Bidirectional questions provide respondents with a
choice between identifying themselves as possessing characteristics
associated with the hypothesized or the alternative trait (e.g., "On Fri-
day nights, do you prefer to go to loud parties or to have a quiet conver-
sation with only one or two close friends?").

Open-ended. Open-ended questions ask about features related to
the hypothesized and the alternative trait without specifically men-
tioning them (e.g., "What do you like to do on weekends?").

Irrelevant. Irrelevant questions were judged to have no obvious rela-
tion to the hypothesized or alternative trait (e.g., when generating ques-
tions for honesty, one subject wrote, "What do you like to eat for
lunch?").

Two judges rated the subjects' protocols in different random orders,
and each question was classified by the judges into one of the coding
categories. The judges agreed on 85% of their classifications. Disagree-
ments were resolved by a third judge.

Results

Subjects were encouraged to prepare at least six questions to
test their hypotheses. Overall, there was no difference in the
number of questions generated to test the four traits of interest
(M extraversion = 9.59; M introversion = 10.00; M dishonesty =
9.41; M honesty = 10.18, F < 1, ns). We also examined the
percentage of questions that were classified into each of the
coding categories. Most of the questions (70%) were of the yes/
no type. Overall, 55% of the questions could be identified as
hypothesis true or alternative true yes/no questions, and 15%
were of the ambiguous yes/no type. The percentages in each of
the other coding categories were relatively small (open-ended =
12%; irrelevant = 10%; bidirectional = 8%). Because yes/no hy-
pothesis true and alternative true questions were generated
much more frequently than any of the other question types, our
analyses focused exclusively on these questions. In what follows,
we examine the introversion–extraversion and the honesty–
dishonesty dimensions separately.

Introversion–extraversion dimension. Subjects generated an
equal number of unambiguous yes/no questions whether they
were testing extraversion (M = 5.09) or introversion (M = 4.43),
F(1, 41) = 0.53, ns. The number of introverted and extraverted
questions subjects generated was treated as a repeated measure
and was examined as a function of the trait subjects were test-
ing. As expected, the analysis revealed a significant question
type main effect, F(1, 41) = 19.95, p < .0001. Overall, subjects
prepared a greater number of extraverted (M = 3.47) than intro-
verted (M = 1.30) questions. Thus, consistent with Reeder and
Breuer's (1979) framework, subjects prefer to generate diagnostic
(i.e., extraverted) questions.

However, the question type main effect is qualified by a sig-
nificant Question Type x Trait Tested interaction, F(1, 41) =
16.65, p < .0002. The pattern of this interaction suggests evi-

---

1 In these definitions, introverts are described as shy, timid, distant,
and as preferring to read books over going to loud parties, and so forth.
In contrast, extraverts are described as outgoing, sociable, energetic,
confident, and as being relaxed in social situations and making friends
easily, and so forth. We found in our earliest attempts to do research in
this area that many students did not know what introversion and extra-
version meant. In addition, in most of the research using these traits,
subjects were provided with these definitions; thus, we also included
these definitions, thereby facilitating the comparison of our research
with the existing research.

2 We did not perform an omnibus analysis treating trait dimension as
a variable because all subjects did not test all trait hypotheses. Subjects
were randomly assigned to one of eight possible combinations of trait
hypotheses: introversion then honesty, introversion then dishonesty,
extraversion then honesty, extraversion then dishonesty, honesty then
introversion, dishonesty then introversion, honesty then extraversion,
and dishonesty then extraversion.

3
dence for a confirmatory bias beyond the main effect for diagnosticity. As can be seen in the left panel of Table 1, when testing extraversion, subjects were much more likely to generate extraverted questions than when testing introversion. When testing introversion, subjects were more likely to generate introverted questions than when testing extraversion. These data are consistent with the findings of Skov and Sherman (1986).

**Honesty-dishonesty dimension.** As with introversion and extraversion, overall, subjects generated an equal number of unambiguous yes/no questions when testing honesty \((M = 6.36)\) and dishonesty \((M = 5.64)\), \(F(1, 42) = 0.54, n.s.\). The number of honest and dishonest questions subjects generated was treated as a repeated measure and was examined as a function of the trait subjects were testing. The analysis revealed that subjects generated a greater number of dishonest \((M = 4.39)\) than honest \((M = 1.61)\) questions, \(F(1, 42) = 31.80, p < .00001.\) Again consistent with Reeder and Brewer's (1979) framework, this pattern suggests that subjects prefer to generate diagnostic (i.e., dishonest) questions.

However, this main effect is qualified by a Question Type \(\times\) Trait Tested interaction, \(F(1, 42) = 9.31, p < .004.\) The means for this interaction are presented in the right panel of Table 1. When testing dishonesty, subjects generated more dishonest questions than when testing honesty. When testing honesty, subjects generated more honest questions than when testing dishonesty. Thus, the pattern for the honesty–dishonesty dimension is identical to the pattern observed for the introversion–extraversion dimension.

**Discussion**

The data from Study 1 provide evidence suggesting that subjects use both diagnostic and confirmation strategies in trait hypothesis testing. Reeder and Brewer's (1979) analysis suggested that behaviors associated with the nonrestricted end of trait dimensions are more diagnostic than behaviors associated with the restricted end of trait dimensions (see also McGraw, 1985; Reeder et al., 1977; Skowronsksi & Carlston, 1987). The present results indicate that when testing trait hypotheses, subjects generate a majority of diagnostic (extraverted and dishonest) questions.

Subjects also show a confirmatory bias. That is, when testing extraversion and dishonesty, subjects generate almost all hypothesis true questions; in these conditions, hypothesis true questions are both more diagnostic and confirmatory in nature. Thus, both diagnostic and confirmation strategies would lead to a preference for such questions. When testing introversion and honesty, subjects are clearly influenced by more than diagnosticity. In these conditions, subjects generate a balance of hypothesis true and alternative true questions. For introversion and honesty, hypothesis true questions are confirmatory, but lower in diagnosticity than alternative true questions. Consequently, diagnostic and confirmation strategies are pitted against one another when testing introversion and honesty. In summary, our data are consistent with the notion that both diagnostic and confirmation strategies affect subjects' question generation for both restricted and nonrestricted ends of trait dimensions.

Moreover, the fact that the pattern of results is similar across the introversion–extraversion and the honesty–dishonesty trait dimensions allows us to effectively rule out the positivity/negativity confound in the previous trait hypothesis-testing research (Swann & Giuliano, 1987). In Swann and Giuliano's research, the preferred question types (extraverted questions and dominant questions) represent both the nonrestricted ends of their respective trait dimensions and the positive ends of these dimensions. Thus, it is not possible in their study to determine whether subjects' question generation reflected diagnosticity or positivity effects. In the present study, we used trait dimensions so that the nonrestricted end of one trait dimension was positive (i.e., extraversion) but was negative for the other dimension (i.e., dishonesty). If subjects were guided by positivity in question generation, they should have generated predominantly positive question types (i.e., extraverted and honest questions). This was not the case. Instead, consistent with Reeder (1985) and Skowronsksi and Carlston (1987), subjects in our study were influenced by diagnosticity concerns in their question generation (and, therefore, generated extraverted and dishonest questions).

The results of Study 1 empirically validate Reeder and Brewer's (1979) analysis of hierarchically restrictive trait schemata. In general, the behaviors associated with the nonrestricted end of the trait dimensions appear to be more diagnostic. However, we have no independent assessment of the diagnosticity of restricted and nonrestricted behaviors; that is, we have relied on theoretical rather than empirical definitions of diagnosticity. A much stronger test of the relative strengths of diagnostic and confirmation strategies in trait hypothesis testing would be to manipulate the diagnosticity of questions independently of the restrictive/nonrestrictive end of the trait continuum. This could be accomplished by using a different methodology in which subjects select questions from a list of experimenter-provided questions (cf. Trope & Bassok, 1982; 1983). In this question selection paradigm, questions associated with the restricted and nonrestricted ends of the trait dimensions are empirically equated for diagnosticity.

### Table 1

<table>
<thead>
<tr>
<th>Trait tested</th>
<th>Extravert question</th>
<th>Introvert question</th>
<th>Dishonest question</th>
<th>Honest question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Extraversion</td>
<td>4.59</td>
<td>0.50</td>
<td>4.95</td>
<td>0.68</td>
</tr>
<tr>
<td>Introversion</td>
<td>2.28</td>
<td>2.21</td>
<td>3.82</td>
<td>2.55</td>
</tr>
</tbody>
</table>

The results of Study 1 empirically validate Reeder and Brewer's (1979) analysis of hierarchically restrictive trait schemata. In general, the behaviors associated with the nonrestricted end of the trait dimensions appear to be more diagnostic. However, we have no independent assessment of the diagnosticity of restricted and nonrestricted behaviors; that is, we have relied on theoretical rather than empirical definitions of diagnosticity. A much stronger test of the relative strengths of diagnostic and confirmation strategies in trait hypothesis testing would be to manipulate the diagnosticity of questions independently of the restrictive/nonrestrictive end of the trait continuum. This could be accomplished by using a different methodology in which subjects select questions from a list of experimenter-provided questions (cf. Trope & Bassok, 1982; 1983). In this question selection paradigm, questions associated with the restricted and nonrestricted ends of the trait dimensions are empirically equated for diagnosticity.
As such, if subjects are using a diagnostic strategy, we would expect the question type main effect of Study 1 to disappear. Instead, subjects should prefer to ask empirically defined high diagnostic questions over low diagnostic questions. In addition, by using this paradigm, we can determine whether subjects show a confirmation bias. A confirmation strategy would be revealed by a Trait Tested × Question Type interaction.

Thus, in the following study, we manipulated the diagnosticity of questions presented to subjects and examined their hypothesis-testing strategies in a question selection task (cf. Trope & Bassok, 1983). Once again, subjects tested either the introversion or extraversion hypothesis and either the honest or dishonest hypothesis. The list of questions from which subjects could choose consisted of half high diagnostic questions and half low diagnostic questions. For each trait hypothesis condition, half of both the high and low diagnostic questions were hypothesis true, and half were alternative true.

### Study 2

#### Method

**Question generation and diagnosticity ratings.** The questions subjects generated in Study 1 served as the basis for the questions to be used in the following two studies. After we deleted redundant themes and ensured that all questions were phrased in a yes/no format, subjects' questions were submitted to a second group of 50 subjects to obtain diagnosticity ratings. Twenty-five subjects were asked to rate the honest/dishonest questions for how well answers to the questions would allow the subject to discriminate between an honest and dishonest person, and 25 subjects rated the introverted/extraverted questions for how well the answers would discriminate between an introvert and an extravert. For each question, subjects circled a number between 1 (does not discriminate at all) to 8 (discriminates very well). In addition, in a second rating task, these subjects were asked to indicate whether the action/characteristic described in each question was related to honesty or dishonesty (introversion/extraversion) by circling H or D (I or E). Subjects also rated how strongly the action/characteristic was associated with the circled trait by circling a number from 1 (weakly associated with the circled trait) to 7 (strongly associated with the circled trait). These measures were the same measures used by Trope and Bassok (1983) to determine their high and low diagnostic questions.

From the ratings, we selected 6 high and 6 low diagnostic questions related to each of the four traits of interest. Only questions that were unanimously identified as being related to only one trait and were rated as strongly associated with that trait were identified as acceptable questions. For the honesty–dishonesty dimension, the final list of 24 questions consisted of four equal subsets: (a) honest, high diagnostic questions with a mean diagnosticity rating of 6.41 (ranging from 6.00 to 6.85); (b) honest, low diagnostic questions with a mean diagnosticity rating of 2.73 (ranging from 1.75 to 3.80); (c) dishonest, high diagnostic questions with a mean diagnosticity rating of 6.52 (ranging from 6.30 to 6.80); (d) dishonest, low diagnostic questions with a mean diagnosticity rating of 3.01 (ranging from 2.53 to 3.25). Similarly, there were four equal subsets of questions for the introversion–extraversion dimension: (a) introverted, high diagnostic questions with a mean diagnosticity rating of 6.09 (ranging from 5.72 to 6.52); (b) introverted, low diagnostic questions with a mean diagnosticity rating of 3.03 (ranging from 2.04 to 3.68); (c) extraverted, high diagnostic questions with a mean diagnosticity rating of 6.44 (ranging from 5.96 to 6.84); (d) extraverted, low diagnostic questions with a mean diagnosticity rating of 2.75 (ranging from 1.96 to 3.26). Examples of each of the question categories are presented in the Appendix. The full set of questions was randomly ordered.

**Subjects and design.** Subjects were 120 University of Wisconsin—Madison students from introductory psychology classes who received extra credit for their participation. All subjects responded to one trait from the honesty–dishonesty dimension and one trait from the introversion–extraversion dimension. Thus, the design of these studies was a Trait (1 honest or dishonest) × Trait 2 (introversion or extraversion) × Order (Trait 1 first or Trait 2 first) × Question Type (hypothesis true or alternative true) × Diagnosticity of Questions (high or low) mixed model factorial. Trait 1, Trait 2, and order were between-subjects variables. Question type and diagnosticity of the questions were within-subjects variables. Fifteen subjects were randomly assigned to each of the eight between-subjects cells. One subject did not follow instructions and was dropped from the analyses.

**Procedure.** Following Trope and Bassok (1983), subjects were provided with a booklet and instructions explaining that they would be asked to select a subset of questions (from an experimenter-provided list of questions) that could be used to determine whether another student possessed a particular personality trait. The second page of the booklet introduced subjects to the trait dimension of interest. Page 3 contained the entire list of 24 questions, and subjects were asked to read through all 24 questions before turning to page 4 in the booklet. On page 4, subjects were asked to imagine that they had to test the idea that a person is (trait of interest). Subjects were told that in testing hypotheses about others, one can obtain information from a number of sources, such as from the person in question or from acquaintances. All of the questions were framed as though they would be asked of acquaintances (e.g., "Would this person lie in a court of law?"). This framing is important because to test hypotheses accurately (especially dishonesty), subjects need to be able to assume that they will get truthful answers to the questions.

From the list of 24 questions, subjects were asked to choose the 12 questions that they would most like to include in a questionnaire to test the idea that the person was (trait of interest). Subjects followed the exact same procedure for Trait 1 and Trait 2. When subjects had completed the question selection for both traits, they were debriefed and thanked for their participation.

### Results

As in Study 1, we present the findings for the introversion–extraversion dimension and the honesty–dishonesty dimension separately. For each trait dimension, the questions subjects selected were submitted to a Trait Tested × Order × Question Type × Diagnosticity of Questions mixed model factorial. The order in which subjects responded to the trait hypotheses did not affect any of the findings, and we will not consider this variable any further.

**Introversion/extraversion.** Table 2 presents the number of high and low diagnostic extraverted and introverted questions selected as a function of the hypothesized trait and diagnosticity of the questions. As expected, and consistent with Trope and Bassok (1983), high diagnosticity questions were obtained both for subjects' questions from Study 1 and for some experimenter-generated questions. We found that most of the subjects' own generated questions were not very high or very low in diagnosticity. To ensure that we had questions at both ends of the diagnosticity continuum, the experimenters generated a number of questions that were also rated by judges for their diagnosticity.

---

4 Actually, diagnosticity ratings were obtained both for subjects' questions from Study 1 and for some experimenter-generated questions. We found that most of the subjects' own generated questions were not very high or very low in diagnosticity. To ensure that we had questions at both ends of the diagnosticity continuum, the experimenters generated a number of questions that were also rated by judges for their diagnosticity.
Bassok's (1983) work, subjects selected many more high (M = 9.39) than low (M = 2.57) diagnostic questions, $F(1, 111) = 313.71, p < .00001$. Moreover, as predicted, there was no question type main effect, $F(1, 111) = 0.22, \text{ns}$; subjects did not prefer questions associated with the restricted or nonrestricted end of the trait dimension.

In addition, we obtained a significant Question Type $\times$ Trait Tested interaction, $F(1, 111) = 56.35, p < .00001$, indicating that subjects preferred overall to ask hypothesis true over alternative true questions. That is, when the trait hypothesis was introversion, subjects selected a greater number of introverted questions (M = 7.10) than when the trait hypothesis was extraversion (M = 4.53). When the trait hypothesis was extraversion, subjects selected a greater number of extraverted questions (M = 7.43) than when the trait hypothesis was introversion (M = 4.88).

The analysis also revealed a significant Diagnosticity $\times$ Question Type interaction, $F(1, 111) = 22.59, p < .00001$, and a significant Diagnosticity $\times$ Trait Tested $\times$ Question Type interaction, $F(1, 111) = 28.52, p < .00001$. As can be seen in Table 2, this three-way interaction is merely one of degree; the Question type $\times$ Trait Tested interaction is moderated in the low diagnosticity condition, but the pattern is the same as in the high diagnosticity condition. Thus, for both high and low diagnosticity questions, there is evidence for a confirmatory bias.

**Honesty-dishonesty** Table 3 presents the average number of honest and dishonest questions subjects selected as a function of the hypothesis tested and the diagnosticity of the questions. Overall, subjects preferred to ask high (M = 9.82) to low (M = 2.18) diagnostic questions, $F(1, 111) = 704.55, p < .00001$. There was not, however, a main effect for question type, $F(1, 111) = .71, \text{ns}$. Thus, as with the introversion–extraversion dimension, subjects do not show a preference for selecting questions associated with the restricted or nonrestricted end of the trait dimension when the questions are equated for diagnosticity.

The preference for hypothesis true questions was obtained at both levels of diagnosticity; there was not a significant three-way interaction, $F(1, 111) = 0.72, \text{ns}$. Thus, as was the case for the introversion–extraversion dimension, these data are consistent with a confirmatory bias.

**Discussion**

The data from Study 2 are consistent with Study 1 in that they suggest that subjects use both diagnostic and confirmation strategies in trait hypothesis testing. An important aspect of Study 2 is that we were able to test the diagnostic strategy more directly than in Study 1 because we empirically manipulated the diagnosticity of questions. Thus, in Study 2, diagnosticity was independent of the restrictiveness/nonrestrictiveness of the traits tested. Our differential diagnosticity interpretation of Reeder and Brewer's (1979) analysis (i.e., behaviors associated with nonrestricted ends are more diagnostic than behaviors associated with restricted ends) would suggest that when questions at the restricted and nonrestricted ends of the dimensions were equated for diagnosticity, there would not be a question type main effect. Rather, we should see a preference for the empirically defined diagnostic questions (i.e., a diagnosticity main effect). Consistent with this interpretation, there was no question type main effect, but subjects clearly preferred to ask high diagnostic, compared to low diagnostic, questions. These results argue strongly that subjects are using a diagnostic strategy in their question selection.

The preference for diagnostic information is consistent with the findings of Trope and Bassok (1983). However, we also find strong evidence for a confirmatory bias in subjects' question selection (i.e., a significant Question Type $\times$ Trait Tested interaction). For all traits tested, subjects preferred to ask hypothesis true over alternative true questions.

Overall, these data suggest that both diagnostic and confirmation strategies play a role in trait hypothesis testing. Consistent with Skov and Sherman (1986) and Trope and Bassok (1983), we find that subjects show a primary preference for diagnosticity and a secondary preference for confirmatory information. That is, although both preferences are statistically significant, the preference for diagnostic questions (80.2% of the chosen questions) was much greater than the preference for confirmatory (hypothesis true) questions (56.4% of the chosen questions). However, we agree with Skov and Sherman's interpretation and disagree with Trope and Bassok's (1983) interpretation of this difference. Although Trope and Bassok (1983)
found a statistically significant confirmatory bias, they mini-
mized its importance in hypothesis testing. Consistent with
Skov and Sherman, we argue that the difference in the magni-
tude of these preferences in no way undermines the importance
of the confirmatory bias. The bias emerges despite the fact that
subjects could have asked all 12 of the high diagnostic questions
(6 of which were hypothesis true, and 6 of which were alterna-
tive true). If subjects were interested only in diagnosticity, we
should see no evidence for a confirmatory bias. Instead, in
Study 2, we see clear evidence for a confirmatory bias above
and beyond the effects of diagnosticity.

However, strong conclusions regarding subjects' strategy pref-
dences for diagnostic and confirmatory information must be
advanced with caution. Skov and Sherman (1986) argued that
Trope and Bassok's (1983) and Study 2's methodology, in which
subjects can choose many questions to ask, does not reflect the
constraints under which social perceivers typically test hypothe-
ses (i.e., they typically cannot ask a lot of questions). Fur-
thermore, this methodology limits our ability to study subjects' strategy
preferences. Strategy preferences are most likely to be re-
vealed by the first piece (or pieces) of information subjects seek.
That is, the first information obtained may carry the greatest
weight in evaluating the truth value of the hypothesis (because of
primacy effects) and is likely to influence subsequent infor-
mation seeking (Skov & Sherman, 1986).

To date, none of the existing trait hypothesis-testing studies
have constrained the number of questions subjects could ask
and, thus, have not provided a strong test of the relative
strengths of diagnostic and confirmation strategies. Skov and
Sherman (1986) did constrain the number of questions subjects
could ask (only two questions). However, as previously noted,
the generalizability of their results to the trait domain is ques-
tionable. Thus, a third study was conducted in which subjects
were permitted to ask only one question.

In Study 3, we presented subjects with a trait hypothesis and a
forced choice between two equally diagnostic questions. For
each trait hypothesis, one question was hypothesis true, and the
other was alternative true. If diagnosticity is the only strategy
subjects consider early in information seeking, subjects should
choose the hypothesis true and alternative true questions with
equal frequency. If however, the confirmation bias plays a role
early in the trait hypothesis-testing sequence, this bias is most
likely to be revealed in a forced-choice task between equally
diagnostic hypothesis true and alternative true questions. That
is, subjects' question selections should depart from a 50/50
split, and subjects should show a significant preference for the
hypothesis true question.

Study 3

It is very difficult to equate questions for diagnosticity (cf.
Hirt, Devine, & Evett, 1990). In the present study, we took
questions that were rated as highly diagnostic on our direct
diagnosticity measures and collected data in a series of pilot
studies with the goal of producing questions that were truly
equal in diagnosticity.

In general, in these pilot studies, we presented subjects with a
trait hypothesis (e.g., honest or dishonest) and two questions
phrased in terms of one of the traits (e.g., either two honest or
two dishonest questions). On the basis of pilot testing, we se-
lected two questions to be used to test the introvert/extravert
and two questions to be used to test the honesty/dishonesty
trait hypotheses in the present study. The honest questions we
used were the "return the wallet" and the "admit to breaking
the window" questions (see the Appendix).

When the questions were phrased in terms of dishonesty, the
respective questions asked if this person would keep the wallet
rather than return it to its owner and if this person would keep
quiet and allow someone else to take the blame for breaking the
window. The extraverted questions we used asked if this person
enjoys going places where he or she will have to meet new peo-
ple and if this person finds it easy to strike up a conversation
with people he or she does not know well. When phrased in
terms of introversion, the respective questions asked if this per-
son dislikes going places where he or she will have to meet new
people and if this person finds it difficult to strike up a conver-
sation with people he or she does not know well.

The criterion for inclusion in the present study was that in
the forced choice task, each question was selected 50% of the
time for the hypothesis at each end of the trait dimension (indi-
cating no preference for either question). The 50% selection
criterion had to be true for both phrasings of the questions.
Although this measure may not be ideal, it provides us with
additional evidence that within trait categories, our questions
are equally good for testing both ends of the trait dimension.
This measure suggests that the questions used in the present
study are roughly equivalent in diagnosticity.

Method

Subjects: Overall, 1,377 University of Wisconsin—Madison stu-
dents participated in the present study in exchange for extra credit
in their introductory psychology class. Subjects were presented with
a questionnaire that provided them with a trait hypothesis and two ques-
tions from which they could choose one to test the hypothesis. This
questionnaire was embedded in a series of questionnaires adminis-
tered to all introductory psychology students who volunteered to par-
ticipate in a large survey early in the semester. Three hundred forty-two
subjects tested the introverted hypothesis, 307 tested the extra-
verted hypothesis, 343 tested the honest hypothesis, and 385 tested the
dishonest hypothesis.

Procedure. The procedure was virtually identical to the procedure
of Study 2. However, in Study 3, subjects were given only a single hy-
pothesis to test and could choose only one question to test the hypothe-
sis. Subjects were provided with general instructions, then the specific
trait hypothesis was described. Subjects were instructed to read the
two questions that could be used to test the hypothesis and to place a
checkmark in the blank next to the one question that they would like
to ask to test the hypothesis.

One question was hypothesis true; the other was alternative true. For
example, subjects testing honesty were presented with the "admit to
breaking the window" (i.e., hypothesis true) and the "keep the wallet"

[5] In retrospect, it seems clear why Skov and Sherman (1986) chose to
examine hypothesis-testing strategies in a domain in which the diag-
nosticity of features could be precisely specified. Unfortunately, in the
personality trait domain, the diagnosticity of behaviors is much more
subjective and cannot be defined with precise probabilities. Dealing
with this type of subjectivity, however, appears to reflect the nature of
hypothesis testing in social settings.
Table 4
Frequency (Freq.) of Question Selection as a Function of Trait Tested for Introverted–Extraverted and Dishonesty–Honesty Dimensions: Study 3

<table>
<thead>
<tr>
<th>Trait tested</th>
<th>Extravert question</th>
<th>Introvert question</th>
<th>Dishonest question</th>
<th>Honest question</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Freq.</td>
<td>%</td>
<td>Freq.</td>
<td>%</td>
</tr>
<tr>
<td>Extraversion</td>
<td>202</td>
<td>65.8</td>
<td>105</td>
<td>34.2</td>
</tr>
<tr>
<td>Introversion</td>
<td>162</td>
<td>47.4</td>
<td>180</td>
<td>52.6</td>
</tr>
</tbody>
</table>

(i.e., alternative true) questions or the “allow someone else to take the blame” (i.e., alternative true and the “return the wallet” (i.e., hypothesis true) questions. The order of the questions on the page (e.g., hypothesis true first or alternative true first) was counterbalanced across subjects. Obviously, the hypothesis true questions for honesty are alternative true for dishonesty, and the alternative true questions for honesty are the hypothesis true questions for dishonesty. This pairing of hypothesis true and alternative true questions and counterbalancing of the order of questions presented to subjects were also followed for the introversion and extraversion hypotheses.

Subjects were not debriefed until the end of the semester, when a report describing all the questionnaires and their purposes was distributed to all students who had participated in the survey.

Results and Discussion

Our goal in the present study was to examine whether subjects show a preference for confirmatory information early in information seeking. Subjects were required to choose between equally diagnostic hypothesis true and alternative true questions. A confirmation bias would be revealed if subjects’ question selections departed from a 50/50 split, so that hypothesis true questions were selected significantly more often than alternative true questions. If diagnosticity concerns alone guide subjects’ initial information-seeking strategies, then hypothesis true and alternative true questions should be selected equally often.

The question selection data for all traits are presented in Table 4. Our primary interest was to determine the magnitude of a confirmation bias in subjects’ question selection. Collapsing across all traits, we found that subjects preferred asking hypothesis true questions (56.2%) to asking alternative true questions (43.8%), \( \chi^2(1, N = 1,377) = 10.4, p < .002 \). These data suggest that subjects demonstrate a confirmation bias above and beyond the effects for diagnosticity. Although subjects testing all traits showed a preference for hypothesis true questions, the magnitude of this preference was marginally stronger for the introversion–extraversion dimension (58.9%) than for the honest–dishonest dimension (53.8%), \( \chi^2(1, N = 1,377) = 3.30, p < .07 \).

Within the introversion–extraversion dimension (see Table 4), the magnitude of the confirmation bias differs as a function of the trait being tested. Subjects testing extraversion selected the hypothesis true question more often (65.8%) than subjects testing introversion (52.6%), \( \chi^2(1, N = 649) = 11.0, p < .001 \). A similar analysis on the honesty–dishonesty dimension (see Table 4) indicated that the strength of the confirmation bias (i.e., selection of hypothesis true questions) does not differ as a function of the trait being tested (testing dishonesty = 56.4%; testing honesty = 51.0%), \( \chi^2(1, N = 728) = 1.87, ns. \)

In Study 2, we did not find a question type main effect in question selection. In Study 3, because we had empirically defined diagnosticity, we did not expect questions associated with the nonrestricted end of trait dimensions (extraversion and dishonesty) to be selected more often than the questions associated with the restricted end of the dimensions (introversion and honesty). However, in Study 3, we unexpectedly found that subjects were more likely to select questions associated with the nonrestricted ends of the trait dimensions (54.4%) than questions associated with the restricted ends of the dimensions (45.6%), \( \chi^2(1, N = 1,377) = 5.15, p < .03 \). Thus, question type preferences were obtained, even though the questions were equally diagnostic. We speculate about possible reasons for this in the General Discussion.

General Discussion

The goal of the present series of studies was to examine the relative contribution of diagnostic and confirmation strategies in trait hypothesis testing. The pattern of findings across the three studies supports the findings of Skov and Sherman (1986), indicating that both diagnosticity and confirmation play an important role in determining the information subjects seek when testing hypotheses. Diagnosticity clearly plays the primary role in trait hypothesis testing. Subjects showed an overwhelming preference for highly diagnostic information in this research. Thus, our results, like those of Trope and Bassok (1983) and Skov and Sherman, suggest that subjects are sensible hypothesis testers and select the most informative questions available to them.

However, when subjects are presented with equally diagnostic hypothesis true and alternative true questions, subjects also show a preference for hypothesis true questions, indicating a secondary but consistent confirmation bias. Moreover, this secondary preference for confirmatory evidence also appears to be pervasive. We observed this preference in three studies using different methodologies. That is, subjects showed the confirmation bias when they generated their own questions to test trait hypotheses (Study 1), when they selected many questions from an experimenter-provided list (Study 2), and when they selected a single question in a forced-choice task (Study 3). In Study 1, diagnosticity was defined theoretically (cf. Reeder & Brewer, 1979), and in Studies 2 and 3, diagnosticity was defined empirically.

Our studies are the first to examine the relative strengths of diagnostic and confirmation strategies in trait hypothesis testing. The prior trait hypothesis-testing studies have tended to pit
one strategy against the other (Trope & Bassok, 1983) or have focused exclusively on only one strategy (Swann & Giuliano, 1987). Trope and Bassok (1983) concluded that diagnosticity is the dominant strategy in hypothesis testing and suggested that a confirmation strategy (defined as the seeking of high or low diagnostic hypothesis true questions over high diagnostic alternative true questions) rarely occurs. Swann and Giuliano (1987) did not examine diagnosticity but argued that their question generation data were consistent with a confirmation strategy.

However, a closer examination of the findings from each of these studies suggests that the results are, in fact, consistent with the conclusions of both the present research and Skov and Sherman (1986); that is, in each of these studies, there is a primary preference for diagnosticity and a secondary preference for confirmatory information. For example, Trope and Bassok (1983) found a strong diagnosticity main effect but also found a statistically significant Question Type X Trait Tested interaction (as we found in Study 2). Using our own definition of a confirmation strategy, this interaction indicates a confirmation effect (i.e., a preference for hypothesis true questions) in addition to the diagnosticity effect. Although Swann and Giuliano (1987) did not directly assess diagnosticity, the theoretical analysis of Reeder and Brewer (1979) regarding hierarchically restrictive traits can be applied to the trait dimensions they explored. Overall, subjects generated a greater number of questions associated with the nonrestricted ends of the trait dimensions (extraversion and dominance). As we argued in Study 1, behaviors associated with the nonrestricted ends of trait dimensions are more diagnostic than behaviors associated with the restricted ends of trait dimensions. This question type main effect, we argue, illustrates a diagnosticity effect. Swann and Giuliano (1987) also obtained a significant Question Type X Trait Tested interaction, suggesting a confirmation strategy. Thus, Swann and Giuliano's results are consistent with both diagnosticity and confirmation effects.

**Origins of the Confirmation Bias**

Given the consistency with which we have observed both diagnosticity and confirmation effects, important questions remain concerning the reasons why subjects show these preferences. The preference for diagnosticity appears to be straightforward: Subjects prefer the information that will allow them to best discriminate between the hypothesis and the alternative. However, the reasons underlying the preference for confirmatory information are not as obvious. As we argued previously, the preference for hypothesis true questions (i.e., positive test strategy) does not necessarily lead to preservation of the current hypothesis. Subjects could receive either confirmatory or disconfirmatory feedback to these questions. Thus, the motivation to preserve one's current hypothesis does not serve as an adequate explanation for a confirmation bias.

However, we can suggest some possible (though yet untested) reasons for why subjects show a preference for confirmatory information. Subjects may find hypothesis true information easier to process than alternative true information (cf. Skov & Sherman, 1986; Slowiacek, Klayman, Sherman, & Skov, 1989). That is, if a subject finds out that someone performs a behavior consistent with the hypothesis (i.e., yes to hypothesis true question), it may be easier to interpret that behavior and draw a trait inference than if a subject finds out that someone did not perform a behavior inconsistent with the hypothesis (i.e., no to alternative true question). Clearly, an extra stage of processing is required to interpret the alternative consistent behavior in terms of its implications for the current hypothesis. Thus, the additional effort required to process feedback to alternative true questions may lead subjects to prefer the easier-to-process hypothesis true questions.

Alternatively, subjects may choose hypothesis true questions because they (mis)perceive them as more diagnostic than alternative true questions. Past research (Beyth-Marom & Fischhoff, 1983; Doherty et al., 1979; Skov & Sherman, 1986; Slowiacek et al., 1989), for example, demonstrated that subjects consider only the likelihood of a feature or behavior specified in a question under the hypothesis, rather than considering the relative likelihood of the feature (behavior) under the hypothesis versus the alternative(s). As a result, only questions with extreme values under the hypothesis are perceived as diagnostic. This effect has been labelled pseudodiagnosticity.

In our experimental situations, subjects might manifest pseudodiagnosticity by only considering the implications of the questions they select for the hypothesis tested and not for the alternative. For example, subjects testing an introvert hypothesis and asking a hypothesis true question, such as "Do you feel ill at ease in social situations?", may only consider the likelihood that introverts would do this behavior and not the likelihood that extraverts would do this behavior.

A third possibility for why subjects show a preference for confirmatory information is that the specific hypothesis being tested may make behaviors consistent with that hypothesis more accessible in memory. For example, asking subjects to test the introvert hypothesis may cue behaviors associated with introversion and could lead to a preference for hypothesis true questions. Although the present research was not designed to differentiate among these possibilities, it is clear that future research will need to address questions regarding the reasons why subjects demonstrate a confirmation (positive test) strategy.

**Implications of Trait Schemata for Hypothesis Testing**

Across the three reported studies, we obtained converging evidence that both diagnostic and confirmation strategies influence subjects' information gathering. However, the results of Study 3 also indicated that subjects preferred to ask questions associated with the nonrestricted ends (extraversion and dishonesty) rather than the restricted ends (introversion and honesty) of the trait dimensions. This finding is especially puzzling because the questions provided to subjects in Study 3 were equal in empirically established diagnosticity. Thus, the question type main effect in Study 3 cannot be attributed to the differential diagnosticity of the behaviors associated with the restricted and nonrestricted ends of the trait dimensions. Moreover, our results demonstrate that social desirability is not a viable explanation for this effect because subjects asked more questions about dishonest behaviors (low in social desirability) as well as more questions about extraverted behaviors (high in social desirability).

However, it seems possible that this question type main effect could simply be the result of the natural confounding of question type with diagnosticity in most real-life situations. That is, as Reeder and Brewer (1979) suggested, behaviors associ-
cialized with the nonrestricted ends of hierarchically restrictive trait dimensions tend to be more diagnostic and informative than behaviors associated with the restricted ends. Thus, perhaps this question type main effect merely reflects an overextension of a diagnostic strategy heuristic (choose behaviors associated with the nonrestricted end because they are the most diagnostic), a typically useful strategy, to an atypical situation in which diagnosticity and question type are unconfounded.

Alternatively, it could be argued that behaviors associated with the nonrestricted ends of trait dimensions tend to reflect the presence of certain abilities and behaviors (going to parties, doing dishonest acts), whereas the restricted ends of trait dimensions tend to reflect the absence of those same abilities and behaviors (not going to parties, not doing dishonest acts). The work on the feature positive effect (e.g., Fazio, Sherman, & Herr, 1982; Newman, Wolff, & Hearst, 1980) suggests that subjects find it much easier to think about the presence of features than the absence of features. That is, the presence of features may be more accessible than the absence of features. Thus, subjects' preference for behaviors associated with the nonrestricted ends of dimensions may reflect the ease with which subjects can think about the presence of behaviors or abilities versus the absence of behaviors or abilities.

Final Conclusions

Perhaps the most important contribution of the present research is that it serves to integrate extant hypothesis-testing research with theoretical and empirical work in the areas of impression formation and trait attribution. As we have shown, qualitative differences among traits determine the diagnosticity of behaviors associated with each end of trait continuua and ultimately affect the types of questions subjects select to test their hypotheses. In addition, focusing on these qualitative differences raises a new set of questions concerning the ways in which features of the specific trait that is tested (or, more generally, the content of hypotheses) influence the hypothesis-testing process. For example, are answers to some question types more influential in making trait inferences than answers to other question types? Alternatively, is less evidence required for subjects to make confident trait attributions about some traits than about others?

Personality trait hypothesis testing is an inescapable part of our everyday life and plays a critical role in the initiation and development of social relationships. However, the domain of trait hypothesis testing does not afford as much control over variables such as diagnosticity as is possible in a domain in which subjects have no a priori expectations about or experience with the hypothesis content (e.g., Skowronski, 1987). We would argue that the methodological challenges encountered in trait hypothesis testing are necessary evils if this important and complex phenomenon is to be understood. We believe that exploring these issues not only will help us to understand the dynamic processes by which people test trait hypotheses, but also will help us to better understand the underlying representations of trait dimensions in the minds of our subjects.

References


Sample High and Low Diagnostic Questions

**High Diagnostic Honest**

1. If this person found a wallet with identification, would he or she return it to its owner without taking anything?
2. If someone was being blamed for breaking the window this person broke, would this person admit that he or she had done it?

**High Diagnostic Dishonest**

1. Would this person lie in a court of law?
2. Does this person cheat on his/her girlfriend/boyfriend?

**High Diagnostic Introverted**

1. Does this person enjoy doing things alone more than with others?
2. Does this person dislike going places if he or she knows that he or she will have to meet new people?

**High Diagnostic Extraverted**

1. Can this person strike up conversations with people that he or she doesn't know well?
2. Does this person usually organize social activities (e.g., parties)?

**Low Diagnostic Honest**

1. Is this person religious?
2. Does this person always pay his or her bills on time?

**Low Diagnostic Dishonest**

1. Does this person gossip?
2. Does this person tape movies from television?

**Low Diagnostic Introverted**

1. Does this person watch PBS?
2. Does this person study on weekends?

**Low Diagnostic Extraverted**

1. Does this person like beer?
2. Is this person attractive?