The Role of the Hypothesis and the Evidence in the Trait Hypothesis Testing Process

SOPHIA R. EVETT AND PATRICIA G. DEVINE

University of Wisconsin at Madison

EDWARD R. HIRT

Indiana University

AND

JANA PRICE

University of Wisconsin at Madison

Received: September 1, 1993; revised: February 18, 1994; accepted: March 31, 1994

Two experiments were conducted to examine (1) whether a trait hypothesis about a target person sets up expectancies for the type of evidence that will be received and (2) how subjects use a trait hypothesis and actual evidence about the target to draw conclusions about the validity of the hypothesis and, thus, to form an impression of the target person. In Experiment 1, regardless of the type of question selected (i.e., hypothesis true or alternative true), subjects expected to receive hypothesis consistent evidence. In Experiment 2, subjects' impressions of the target were a joint function of the initial trait being tested and the type of evidence received. Hypothesis consistent evidence led to dispositional attributions that the target possessed the hypothesized trait. Consistent with Reeder and Brewer's (1979) analysis of trait schemata, hypothesis inconsistent extraverted evidence led to dispositional attributions of extraversion, whereas hypothesis inconsistent introverted evidence led to hedging and situational attributions. Dis-

Portions of this paper were presented at meetings of the Midwestern Psychological Association, Chicago, April 1991 and 1992. The authors thank Michelle Bruess, Wendy Coomer, Jason Howard, Tami Kelber, Julie Scholz, and Dana Tepp for their input in developing this research and for their assistance with data collection and coding. The authors also thank three anonymous reviewers for their helpful comments on an earlier version of this manuscript. Correspondence and reprint requests should be addressed to Sophia R. Evett or Patricia G. Devine, Department of Psychology, 1202 West Johnson Street, University of Wisconsin, Madison, WI 53706.
discussion focuses on the limitations of the question selection methodology that predominate in the trait hypothesis testing literature and on the importance of considering the properties of traits to more fully understand the process of testing trait hypotheses. © 1994 Academic Press, Inc.

Many social psychologists have likened the process of forming an impression to that of testing a scientific hypothesis (e.g., Darley & Gross, 1983; Snyder & Swann, 1978). That is, people generate a hypothesis about the characteristics of a target person (e.g., is this person shy?) and proceed to test the hypothesis by seeking out information about the person. An issue that has garnered a great deal of attention among trait hypothesis testing researchers concerns how perceivers’ initial trait hypotheses influence the way in which they gather information about the target person (e.g., Bassok & Trope, 1984; Devine, Hirt, & Gehrke, 1990; Pennington, 1987; Snyder & Swann, 1978; Swann & Giuliano, 1987; Trope & Bassok, 1982, 1983; Trope, Bassok, & Alon, 1984). To explore information gathering strategies, trait hypothesis testing researchers have relied predominantly on a single methodology: examining the types of questions hypothesis testers select to test hypotheses about the personality of a target person. Although this question selection methodology has been very useful in examining what type of information people initially seek when testing a hypothesis, it is rather limited in what it can reveal about the process of actually testing a trait hypothesis.

**DIAGNOSTIC AND CONFIRMATORY STRATEGIES IN QUESTION SELECTION**

In typical trait hypothesis testing studies, subjects are provided with a trait hypothesis (e.g., introversion) and select questions that could be asked of the target person from a list provided by the researchers. Information gathering strategies have been inferred from characteristics of the preferred questions. These characteristics include: (1) the diagnosticity of a question (i.e., how well it differentiates between a hypothesized trait and its alternative), (2) the phrasing of the question (i.e., whether it asks about characteristics associated with the hypothesized trait or the alternative trait), and (3) the likelihood of receiving a hypothesis confirming or hypothesis disconfirming answer to the question.

Research using the question selection methodology has revealed two major information gathering strategies: the diagnostic strategy and the confirmatory strategy. The diagnostic strategy, proposed by Trope and Bassok (1982, 1983; Bassok & Trope, 1984), suggests that people are objective hypothesis testers who are motivated to seek out the most “diagnostic” information available. The diagnostic strategy leads people to prefer questions that most clearly distinguish between the hypothesis and its alternative(s). According to this strategy, the specific hypothesis being
tested (e.g., whether it is "introversion" or "extraversion") is considered relatively less important in determining subjects' question selection preferences. In contrast, the confirmatory strategy suggests that rather than being objective hypothesis testers, people tend to be biased toward confirmation of the initial hypothesis\(^1\) (e.g., Bassok & Trope, 1984; Devine et al., 1990; Skov & Sherman, 1986; Snyder & Cantor, 1979; Snyder & Swann, 1978; Swann, Giuliano, & Wegner, 1982). The confirmatory strategy is revealed by two types of question preferences: (1) a preference for questions that ask about features associated with the presence of the hypothesized trait rather than the alternative trait (called the "positive test strategy") and (2) a preference for questions for which the probability of a hypothesis confirming answer is greater than the probability of a hypothesis disconfirming answer (called the "hypothesis preservation strategy"). Both types of confirmatory strategies argue that the hypothesis plays an important role in determining what types of questions are preferred.

Although there has been considerable debate regarding the relative importance of the diagnostic and confirmatory information-seeking strategies, the question selection literature suggests that the strategies may operate simultaneously. Specifically, hypothesis testers show a primary preference for highly diagnostic questions and a secondary preference for questions of the confirmatory or hypothesis preserving variety (Devine et al., 1990; Skov & Sherman, 1986). Thus, conclusions about hypothesis testing strategies have been based solely on the characteristics of the questions selected. We would argue, however, that the question selection methodology is inherently limited in what it can reveal about hypothesis testing strategies. Testing a hypothesis involves not only asking questions, but also receiving and evaluating evidence in order to draw conclusions about the validity of the hypothesis (Pyszczynski & Greenberg, 1987).

Because researchers have predominantly relied on the question selection methodology, insufficient attention has been devoted to two questions that we would argue are critical in fully understanding the nature of hypothesis testing strategies.\(^2\) First, do hypotheses set up expectations for

---

\(^1\) Researchers have proposed both motivational and cognitive explanations of the confirmation bias ranging from an inherent belief in the truth of the hypothesis to the greater ease of processing positive (hypothesis confirming) versus negative (hypothesis disconfirming) information. Determining which of these mechanisms is operating (or if both are) is not of concern to the present research.

\(^2\) It is important to note that we are not claiming that all trait hypothesis testing researchers have focused only on the question selection methodology and ignored the remainder of the hypothesis testing process. Both Pennington (1987) and Swann and Ely (1984), for example, examined subjects' impressions of extremely introverted or extraverted targets following extensive question and answer periods (e.g., receiving elaborate answers to 15 questions). Regardless of whether the evidence was consistent or inconsistent with subjects' initial
how the target will answer questions? Second, how do hypothesis testers use the hypothesis and the evidence provided by the target person's answers to questions to draw conclusions about the validity of the hypothesis and, thus, to form an impression about the target? Hypothesis testers may evaluate evidence in an objective fashion, or, alternatively, they may be biased in their evaluation of evidence in a hypothesis confirming manner. The primary goal of the present research was to investigate the roles of both the hypothesis and the evidence in the trait hypothesis testing process by examining whether a trait hypothesis sets up expectations for how a target will answer questions selected by the subjects (Experiment 1) and how subjects use initial evidence to evaluate a trait hypothesis and to make a judgment about a target's disposition (Experiment 2).

THE ROLE OF HYPOTHESES AND EVIDENCE IN THE HYPOTHESIS TESTING PROCESS

In most hypothesis testing studies, the initial hypothesis is presented to subjects in a fairly arbitrary fashion and is not typically based on preexisting knowledge about the target. As such, the hypothesis conveys no information about the target; it merely represents a potential personality characteristic of the target that could be either supported or disconfirmed. If subjects recognize the arbitrary nature of the hypothesis, they should disregard it as a source of information and their impressions should be based on the evidence received. However, several hypothesis testing researchers have suggested that the initial hypothesis provides a framework within which subjects seek and evaluate information (Klayman & Ha, 1987; Skov & Sherman, 1986; Snyder, 1981; Snyder & Swann, 1978). For example, Skov and Sherman (1986) argued that by providing subjects with a hypothesis to test, features and attributes consistent with the hypothesized trait become highly accessible. In testing the hypothesis, subjects seek out information regarding these accessible features and, as a result, their information seeking is biased toward hypothesis consistent information. Similarly, Snyder (1981) suggested that the task of testing a hypothesis appears to be one of building a case in support of the hypothesis. Given a particular hypothesis to test, subjects seek out and may expect to receive hypothesis consistent evidence.

If a hypothesis does indeed set up expectancies, it is likely to affect hypothesis (or expectancy), subjects drew accurate conclusions about the target's personality (see Slowiaczek, Klayman, Sherman, & Skov, 1992, for a similar finding outside of the trait domain). Although these studies indicate that evidence plays an important role in the hypothesis testing process, they do not permit an examination of how a hypothesis influences subjects' evaluation of evidence early in the process of testing a hypothesis. Our empirical strategy focuses on hypothesis testers' reactions to initial evidence they receive. It is at these early stages that a hypothesis may exert its strongest influence. We will return to this issue in the discussion.
how subjects process and evaluate evidence (Darley & Gross, 1983; Hastie, 1984; Hastie & Kumar, 1979; Lau & Russell, 1980; Pyszczynski & Greenberg, 1981; Pyszczynski, LaPrelle, & Greenberg, 1987; Srull, Lichtenstein, & Rothbart, 1985; Stangor & McMillan, 1992; Wong & Weiner, 1981). According to Darley and Gross (1983), expectancies serve as working hypotheses that are then tested in a biased fashion. They empirically demonstrated that subjects selectively attended to and remembered hypothesis consistent (or expected) evidence. Moreover, hypothesis consistent evidence was more readily accepted as valid. Evidence that was inconsistent with the hypothesis was distorted, discounted, or otherwise explained away. As a result of this biased evaluation of evidence, subjects were able to "find" confirmation for their hypotheses from a mixed set of evidence. Similar processes may operate when subjects test trait hypotheses. If hypotheses lead subjects to expect evidence that is consistent with the hypothesized trait (cf. Skov & Sherman, 1986; Snyder, 1981), such evidence should be readily accepted and should lead subjects to make dispositional attributions that the target possesses the hypothesized trait (Gilbert, 1989; Hastie, 1984; Jones & Davis, 1965; Kulik, 1983). However, when the evidence is inconsistent with the initial hypothesis, subjects may be surprised by the evidence and scrutinize it more closely (Hastie, 1984; Hastie & Kumar, 1979; Srull et al., 1985; Stangor & McMillan, 1992). This scrutiny could lead them to distort or explain away the evidence. If so, subjects may conclude that a hypothesis is supported even when the evidence does not warrant such a conclusion. Alternatively, some hypothesis inconsistent evidence may resist distortion and may not be as easily explained away. Such evidence is likely to undermine an initial hypothesis.

Reeder and Brewer's (1979) theoretical analysis of trait schemata, although not developed in the hypothesis testing domain, provides insight into how subjects' increased scrutiny of hypothesis inconsistent evidence may influence their evaluation of the evidence and their inferences about the target person. Reeder and Brewer posited that for some trait dimensions, which they labeled "hierarchically restrictive," the trait at one end of the trait dimension is associated with a broader range of behaviors than is the trait at the other end of the dimension (see also Reeder, Messick, & Van Avermaet, 1977). Consider, for example, the hierarchically restricted trait dimension introversion–extraversion. Reeder and Brewer argued that although extraverts typically behave in an extraverted manner, they also occasionally engage in introverted behaviors (e.g., sometimes acting shy with a stranger). As such, introverted behaviors are clearly within the behavioral repertoire of extraverts. In contrast, the range of behaviors for introverts is much more restricted, with extraverted behaviors generally outside their behavioral repertoire. Thus, introverts behave almost exclusively in an introverted manner, only very rarely
engaging in extraverted behaviors. An implication of this asymmetry is that whereas extraverted behaviors are highly diagnostic (because they indicate the presence of the trait extraversion), introverted behaviors are less diagnostic (because they are not uniquely associated with either end of the trait dimension).

The key argument we derived from Reeder and Brewer's (1979) analysis is that when evidence is unexpected, the overall diagnosticity of the behaviors should determine the types of trait inferences subjects make. For example, because the behavioral repertoire of introverts does not include extraverted behaviors, extraverted evidence is likely to undermine an introverted hypothesis and subjects should draw extraverted dispositional inferences. Moreover, because extraverted behaviors are highly diagnostic of extraversion, inferences drawn from these behaviors should be made with certainty. In contrast, because the behavioral repertoire of extraverts includes introverted behaviors, introverted evidence is not likely to seriously challenge an extraverted hypothesis nor will it provide strong support for it. This ambiguity should create uncertainty when making inferences about the target's internal disposition. In light of Reeder and Brewer's analysis, then, we propose that the nature of the evidence may determine whether hypothesis inconsistent evidence will be distorted and explained away (cf. Darley & Gross, 1983) or accepted when making judgments about the target person.

In sum, we believe that to more fully understand strategies involved in the process of trait hypothesis testing, and the role of the initial hypothesis in this process, it is crucial to move beyond examining question selection preferences. To this end, we conducted two experiments. Our first step in moving beyond the question selection methodology was to examine whether providing subjects with an arbitrary trait hypothesis sets up expectancies for the type of evidence that will be provided by the target. Thus, in Experiment 1, subjects were given a hypothesis to test, they were permitted to select a question to ask of the target, and they reported their expectations concerning how the target would answer the question. Our second step was to examine how subjects responded to evidence, in the form of a target's answer to their question. Thus, in Experiment 2, subjects were given a hypothesis to test, received actual evidence that was either consistent or inconsistent with the initial hypothesis, and then reported judgments about the target's disposition.

**EXPERIMENT 1**

Experiment 1 employed the traditional question selection methodology with one simple extension: In addition to receiving a trait hypothesis to test and selecting a question to test that hypothesis, subjects then described what answer (if any) they expected to receive to their question. The combination of the phrasing of the selected question and the expected
answer revealed what type of evidence subjects expected. For example,
selecting the introverted question “Is this person shy around strangers?”
and expecting a Yes answer indicated anticipating introverted evidence,
which would be hypothesis consistent for the introverted hypothesis and
hypothesis inconsistent for the extraverted hypothesis. A control condition
was included to determine the nature of subjects’ expectations in the
absence of a specific trait hypothesis to test.

Method

Subjects and Design

One hundred and sixty-six students in an introductory psychology course participated in
a study on impression formation as part of a class demonstration. Subjects were randomly
assigned to the hypothesis conditions. Two-thirds of the subjects tested one of two specific
hypotheses (i.e., “introverted,” N = 55, “extraverted,” N = 56). The remaining subjects
did not test a specific hypothesis. Half of these subjects were asked to determine if the
target was either “introverted or extraverted,” and the other half were asked to determine
if the target was either “extraverted or introverted.” The order of the traits was counter-
balanced to control for primacy effects. Because the order of the traits showed no effect,
these two conditions were collapsed to form a single no specific hypothesis control condition,
N = 55.

Procedure

Subjects were told that they would be selecting questions that would help them learn
whether an individual possessed a particular personality trait. Subjects read the definitions
of introversion and extraversion originally used by Snyder and Swann (1978) to ensure that
everyone was familiar with these traits. Next, subjects read twelve highly diagnostic Yes/No
questions taken from Devine et al. (1990). Six questions were phrased in terms of introverted
characteristics (e.g., Is this person shy?) and six were phrased in terms of extraverted
characteristics (e.g., Is this person outgoing?). The questions were presented in a randomly
determined order.

Subjects selected the question that they would most like to ask to determine whether a
person was ______ (the hypothesized trait). Next, they were asked to state how they antici-
ipated the person would answer the selected question and to explain their reasoning. An
open-ended response format was chosen to allow subjects to state the exact nature of their
expectations, if they had any, rather than to force them to choose from a restricted set of
possible responses.

Dependent Measures and Coding

Three dependent measures were examined: (1) subjects’ selection of either an introverted
or extraverted question; (2) the type of answer subjects’ expected to receive; and (3) subjects’
written explanations for the answer they anticipated. The expected answers and the expla-
nations were content coded by two judges. Subjects’ reports of the answers they expected
were coded into one of the following four categories: (1) Yes, (2) No, (3) either Yes or
No (indicating that the subject did not have a specific expectation), or (4) Other. The judges
agreed on 87% of the classifications, with discrepancies resolved by a third judge.

Subjects’ written explanations for the answer they anticipated were coded into one of the
following four categories: (1) exclusively introverted reasons, such as, “I chose this question
because introverts have difficulty talking with strangers;” (2) exclusively extraverted reasons,
such as, “Extraverts enjoy meeting and talking with strangers;” (3) reasons that focused
on both ends of the trait dimension, such as, “This question focuses on a key distinction between introverts and extraverts. Introverts are uncomfortable talking with strangers, whereas extraverts love to meet new people;” and (4) all other reasons were classified as Other and not analyzed. Judges agreed on 78% of the items; discrepancies were resolved by a third judge.

Results

Our main analyses focused on the types of questions subjects in the specific hypothesis conditions selected and their expectations regarding how targets would answer their questions. Because all of the data involved frequencies, they were analyzed using either the $\chi^2$ goodness-of-fit test or the Fisher’s exact test. The Fisher’s exact test is analogous to the $\chi^2$ test of independence, except that it provides the exact probability for any pattern of frequencies in a $2 \times 2$ table, assuming that chance is operating (Hays, 1988).

No Specific Hypothesis Condition

In the no specific hypothesis condition, subjects selected questions that would help them determine if the target was “either introverted or extraverted” (or “extraverted or introverted”). Responses in this condition provided us with baseline data against which the questions selected and the type of evidence expected in the specific hypothesis conditions could be compared.

Overall, 63.6% of subjects selected introverted questions and 36.4% selected extraverted questions. This pattern of question selection is somewhat surprising given that the questions were preselected to be highly and equally diagnostic (cf. Trope & Bassok, 1983). However, our key interest in the no specific hypothesis condition concerned subjects’ expectations for the answers they would receive to the questions they selected. Only 3.6% of subjects’ open-ended responses could not be coded into one of the three categories of interest. The codable responses revealed that over half of the subjects (51.0%) reported that the target might

---

3 However, upon closer inspection, it was clear that the preference for introverted questions was accounted for by a single question. This question, “Does this person enjoy doing things alone more than with other people?” was selected by 34.5% of the subjects in the no specific hypothesis condition. No other question approached this one in popularity. This question is the only one in the list of 12 that mentioned characteristics associated with both ends of the trait dimension. Although it was not intended, a target’s answer to this question actually provides direct information about both ends of the trait dimension, and, therefore, may have been particularly attractive to those whose goal was to test both ends of the trait dimension simultaneously. When this question was excluded from the analysis, there was no significant question type preference, with 44.4% selecting introverted questions and 55.6% selecting extraverted questions, $\chi^2 (1, N = 38) = .44$, n.s. Subjects in the specific hypothesis condition did not show the same bias toward this question. In both the introverted and the extraverted hypothesis conditions, several other questions were selected with equal or greater frequency.
answer "either Yes or No," indicating that they had no specific expectations for how the target would answer the question. When subjects did report that they expected a specific response, they were equally likely to indicate that they expected an introverted (24.5%) as an extraverted (24.5%) response.

Further support that subjects had no specific expectations can be found in the reasons they gave for the answer they anticipated receiving. Subjects' reasons were coded as focusing on introversion, extraversion, or both traits. Eighty percent of subjects' responses fell into these three categories. Of the codable responses, the majority of subjects' reasons (61.4%) focused on both introversion and extraversion. For the remaining responses, 22.7% of the responses focused on introversion and 15.9% of the responses focused on extraversion. These results suggest that the majority of subjects in the no specific hypothesis condition did not have concrete expectations about the target person's disposition and were focused on both ends of the trait dimension.

Specific Hypothesis Conditions

In examining the question preferences of subjects in the specific hypothesis conditions, it was useful to label the introverted and extraverted questions as they relate to the specific hypothesis being tested. In accord with the previous literature, we designated the questions as hypothesis true and alternative true (Devine et al., 1990). Introverted questions are hypothesis true for subjects testing an introverted hypothesis and alternative true for an extraverted hypothesis. Extraverted questions are hypothesis true for an extraverted hypothesis and alternative true for an introverted hypothesis. The percentage of subjects selecting a hypothesis true or an alternative true question are presented in Table 1. Subjects' question selection preferences appear to be guided by the specific hypothesis being tested. Consistent with previous trait hypothesis testing research, the majority of subjects selected hypothesis true (67.6%) rather than alternative true (32.4%) questions, \( \chi^2 (1, N = 111) = 13.70, p < \)
.001. The preference for hypothesis true questions was similar for both the introverted and extraverted hypothesis conditions.

Our key question, however, concerned the answers subjects anticipated receiving to the questions they selected. For the specific hypothesis conditions, analyses were conducted on the 93.6% of responses that were codable. Across the specific hypothesis conditions, only 9.6% of subjects' open-ended responses indicated that they did not have a specific expectation about how the target would answer the question, which is substantially lower than the 51.0% in the no specific hypothesis condition. Indeed, most subjects (90.4%) expected the target to provide either a "Yes" or a "No" answer; analyses were based on these responses. As can be seen in Table 2, regardless of the type of question selected, an overwhelming majority of the subjects expected to receive evidence consistent with their initial hypothesis, Fisher's exact \( (N = 93) \ p < .0001 \). Specifically, 92.3% of the subjects who selected hypothesis true questions anticipated Yes answers and, thus, expected to receive hypothesis consistent evidence. In addition, 96.6% of the subjects who selected alternative true questions anticipated No answers, which indicates that they too expected to receive hypothesis consistent evidence.4

Table 2

<table>
<thead>
<tr>
<th>Question selected</th>
<th>Expected answer</th>
<th>Yes</th>
<th>No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis true</td>
<td></td>
<td>92.3%</td>
<td>7.7%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(60)</td>
<td>(5)</td>
</tr>
<tr>
<td>Alternative true</td>
<td></td>
<td>3.4%</td>
<td>96.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(28)</td>
</tr>
</tbody>
</table>

*Note. These data were collapsed across the introverted and extraverted hypothesis conditions.*

Subjects' explanations for the answer they anticipated receiving revealed a strong influence of hypothesis condition. Whether they were testing introversion or extraversion, subjects showed an overwhelming tendency to provide reasons that focused only on the hypothesized trait, rather

---

4 Subjects did the exact same task for the trait dimension honesty-dishonesty. Regardless of whether subjects selected hypothesis true or alternative true questions, the majority of subjects (90.6%) expected to receive the answer that would confirm the hypothesized trait. It would appear that hypotheses set up expectancies for hypothesis consistent evidence for at least two trait dimensions.
than focusing on either the alternative trait or both traits. Eighty-seven percent of subjects responses could be coded into these three categories. Of these codable responses, 77.3% of subjects’ reasons were coded as focusing only on the hypothesized trait. For example, when testing extraversion, regardless of the question type selected, the majority of subjects discussed the implications of the anticipated answer for the trait extraversion but never discussed the implications of the answer for the trait introversion. Only 4.1% of subjects’ reasons were coded as focusing on the alternative trait. The remaining 18.6% of subjects’ reasons mentioned both ends of the trait dimension. These results further support the notion that an initial trait hypothesis leads most subjects to think only in terms of the hypothesized trait, and that this effect is present regardless of the type of question selected.

Discussion

The primary goal of Experiment 1 was to extend the analysis of hypothesis testing beyond question selection to examine subjects’ expectations regarding how the target person would answer their questions. Several sources of data suggest that the hypothesis provides a frame that guides not only subjects’ question selection preferences but also their expectations for what evidence they will likely receive (Skov & Sherman, 1986; Snyder & Swann, 1978). First, subjects’ question selection preferences replicated the well-established finding that subjects generally prefer hypothesis true to alternative true questions. Second, subjects’ expectations for how the target would answer the questions indicated that the overwhelming majority of subjects expected hypothesis consistent evidence. This was true regardless of which specific hypothesis subjects tested and whether they had selected a hypothesis true or an alternative true question.

The expectation of hypothesis consistent evidence by subjects who selected alternative true questions is particularly interesting. Just as the question selection literature has classified the preference for hypothesis true questions as a “confirmatory” strategy, the preference for alternative true questions has been classified as a “disconfirmatory” strategy. Viewed in the context of subjects’ expectations, it is misleading to classify subjects who selected alternative true questions as using a disconfirmatory strategy because the majority of these subjects expected that the target’s response would confirm their hypothesis. Thus, like their counterparts who selected hypothesis true questions, these subjects also showed a confirmatory bias. These data help to highlight the theoretical benefits of extending the analysis of hypothesis testing beyond the question selection methodology to consider the types of answers subjects expect. Subjects’ information gathering strategies should be defined in terms of the joint function of the question selected and the answer anticipated. As such, these data
suggest that the magnitude of the confirmatory bias so often discussed in the literature appears to be even stronger than has previously been imagined. Finally, subjects' explanations for the answers they expected to receive also appeared to be guided by the hypothesis being tested. Subjects discussed what the answer could reveal about the presence or absence of the hypothesized trait without considering the implications of the answer for the presence or absence of the alternative trait.

In summary, Experiment 1 suggests that the hypothesis affects subjects' information gathering strategies by influencing the type of questions subjects select and their expectations for how the target will answer their questions. It has yet to be determined, however, whether and how such expectations will affect subjects' responses to evidence that is consistent with or challenges the initial hypothesis. These issues are investigated in Experiment 2. The design of Experiment 2 enables us to explore how the initial hypothesis together with actual evidence that was either consistent or inconsistent with this hypothesis (and with subjects' expectations) would affect subjects' impression ratings of a target person and their explanations of these ratings. In order to detect the extent to which subjects accepted or questioned the evidence in forming an impression of the target person, we employed two key measures. First, after receiving evidence, subjects made trait ratings of the target person ranging from "very introverted" to "very extraverted." Second, subjects provided written explanations of those trait ratings which were content coded for the degree of certainty subjects had about the internal disposition of the target person. Certainty was revealed by the presence of dispositional attributions about the target's behavior in the explanations, whereas doubt or uncertainty was revealed by subjects (1) qualifying their dispositional statements with hedging and uncertainty or (2) explaining away the target's behavior with situational attributions.

EXPERIMENT 2

Method

Subjects and Design

One hundred and twenty-six Introductory Psychology students participated in exchange for extra credit. Eight subjects were dropped from all analyses because of their failure to correctly define the traits introversion and extraversion, leaving a total of 118 subjects. Subjects were randomly assigned to one of twelve cells in a Hypothesis ("introverted," "extraverted," "either introverted or extraverted") × Question Type (introverted, extraverted) × Answer (yes, no) between-subjects factorial design. Subjects participated individually.

Procedure

Subjects were brought to a room with a computer that was allegedly connected to a computer in an adjoining room. Subjects were led to believe that there was another participant of the same sex in the adjoining room. They were told that the purpose of the
experiment was for one participant to form an impression of the other by asking and receiving answers to questions. They were further told that they would be using the computers as the mode of communication to encourage honesty in answering the questions and to prevent the participants from using nonverbal cues and physical appearance in making their judgments. In actuality, there was no other participant, but rather, subjects “interacted” with a preprogrammed question and answer sequence, designed using the Micro Experimental Laboratory (MEL) system (Schneider, 1988, 1990). MEL recorded subjects’ responses and response times.

While the subject read and signed a consent form, the experimenter left the room to allegedly check on the other participant, returning with a signed consent form that was conspicuously placed in a folder with the subject’s form. The experimenter then activated the program and left the room for the remainder of the experiment.

The computer displayed instructions that informed subjects that they would be forming an impression of the other person and asked them to test either a specific hypothesis (i.e., determine if the person is “introverted” or “extraverted”) or no specific hypothesis (i.e., determine if the person is “either introverted or extraverted”). Subjects were instructed to determine if the target person possessed that trait by selecting one question from a set of four questions to be answered by the target person. Each set was comprised of either all highly diagnostic introverted questions or all highly diagnostic extraverted questions. The questions in the two sets were identical except for the introverted or extraverted phrasing. For example, the introverted question “If you were sitting next to a stranger on a plane, would you keep to yourself?” was rephrased to read “If you were sitting next to a stranger on a plane, would you strike up a conversation?” in the extraverted question set. Thus, subjects were not given the opportunity to choose between hypothesis true and alternative true questions, as was the case in Experiment 1, but rather chose from questions of only one type.

After subjects selected a question, they were made to wait 30 s before the target’s “answer” (Yes or No) appeared on the screen for three seconds. All subjects received either introverted evidence (e.g., an introverted question and a Yes answer) or extraverted evidence (e.g., an introverted question and a No answer); for subjects testing a specific hypothesis, this evidence was either consistent or inconsistent with the initial hypothesis.

Subjects were asked to rate the target on a scale from 1 (“very introverted”) to 9 (“very extraverted”) on the basis of “the information that you have at this point.” Subjects were then asked to rate how confident they were in this trait judgment on a scale from 1 (“not at all confident”) to 9 (“very confident”). Subjects were also asked to provide an open-ended written explanation of their trait rating.5

Next, subjects were told they would have the opportunity to ask the target another question. The computer displayed a list of 10 questions from Experiment 1, five of which were introverted questions and five of which were extraverted questions, and instructed subjects to choose one. As in Experiment 1, subjects were asked what response they expected to receive and to explain their reasoning. Finally, subjects described the typical introvert and extravert. They were then debriefed and dismissed.

---

5 Although confidence ratings were taken, there was not sufficient variability to draw any systematic conclusions. Most of the ratings were at the middle of the scale. It is likely that the exact phrasing of this question was a bit unusual because we asked subjects to base their trait rating on the evidence they had received, which was rather limited by any standards. The open-ended responses provided much more insight regarding subjects’ certainty in their evaluations of the target person. Thus, we focus on the open-ended responses as our measure of certainty in the text.
Coding

All of the open-ended responses were content coded by two judges, and discrepancies were resolved by a third judge.

*Explanations of trait ratings.* Explanations of the trait ratings were coded according to how certain subjects were about their inferences of the target's disposition. The presence of *dispositional attributions* of the target's behavior in the explanations indicated certainty about the ratings (e.g., "I rated the person in this way because only introverts are shy with strangers"). These dispositional attributions were further characterized as indicating that the target was either "dispositionally introverted" or "dispositionally extraverted." Doubt or uncertainty was revealed by (1) subjects making *situational attributions* of the target's behavior (e.g., "Even outgoing people sometimes act shy with a stranger"), (2) subjects qualifying their dispositional attributions with expressions of hedging or uncertainty (e.g., "I'm not sure," "perhaps") or (3) subjects using both dispositional and situational attributions (e.g., "Maybe he's really shy or maybe he just doesn't feel like talking today") in the same explanation (see Wright & Mischel, 1988, for similar coding of uncertainty). Judges agreed on 82% of the items.

*Expected answer to second question.* The explanations subjects gave for selecting a second question were also coded, using the coding scheme from Experiment 1. Subjects' responses were coded according to what answer they expected the target to give to the second question ("Yes," "No," "either Yes or No," or "Other"). Judges agreed on 97% of the items.

Results and Discussion

To examine the role of the hypothesis in determining subjects' impressions of the target person, it is important to examine first how subjects made judgments in the absence of a specific hypothesis to test. This condition provides a baseline against which we can compare subjects' judgments when they are testing a specific hypothesis. Thus, we will present the data from the no specific hypothesis condition before turning our attention to the specific hypothesis conditions. For all hypothesis conditions, we obtained similar results for the two types of introverted evidence (i.e., introverted question and Yes answer, extraverted question and No answer) and for the two types of extraverted evidence (i.e., extraverted question and Yes answer, introverted question and No answer). Therefore, we collapsed the 2 (Question Type) × 2 (Answer Type) design to create a single two level factor: Evidence Type (introverted, extraverted).

*No Specific Hypothesis Condition*

Subjects in the no specific hypothesis condition based their evaluations of the target on the evidence provided by the target's answer to the question. Subjects who received introverted evidence (e.g., an introverted question with a Yes answer) rated the target as generally introverted ($M = 3.35$), whereas subjects who received extraverted evidence (e.g., an introverted question and a No answer) rated the target as generally extraverted ($M = 7.18$), $t(37) = 12.92, p < .001$. Interestingly, subjects'
explanations of their trait ratings suggested that those who received introverted evidence were less certain about the target’s internal disposition than were subjects who received extraverted evidence. Of the 89.7% of explanations that were codable, 61.9% of the subjects who received extraverted evidence generated dispositional attributions for the target’s behavior. In contrast, 71.4% of the subjects who received introverted evidence generated situational attributions or qualified their dispositional attributions with hedging and uncertainty. These data are consistent with Reeder and Brewer’s (1979) analysis which suggests that extraverted behaviors are more diagnostic (and permit more confident trait attributions) than introverted behaviors.

Subjects were given the opportunity to pose a second question to the target person. Their question selection preferences and the answer they expected to receive indicated how subjects’ expectations could be shaped by early evidence about the target. Although subjects in the no specific hypothesis condition did not show any systematic preferences for extraverted or introverted questions, their expectations for how the target would answer the question appeared to be determined by how the target answered the first question (i.e., with introverted or extraverted evidence). Of the 76.9% of responses that could be coded as “Yes” or “No” answers, 83.5% of the subjects expected that the target would answer the second question in a manner consistent with his or her answer to the first question. Interestingly, this tendency was somewhat stronger when the target provided extraverted evidence (94.7%) than when the target provided introverted evidence (72.7%), again supporting Reeder and Brewer’s perspective. However, this difference was not statistically significant, Fisher’s exact test \( (N = 30) \ p = .13 \). Both types of evidence appeared to lead subjects to develop a hypothesis that sets up expectancies for how the target person will respond to further questioning.

Specific Hypothesis Conditions

Impression ratings. Subjects’ impression ratings in the no specific hypothesis condition appeared to be driven by the type of evidence received. The specific hypothesis conditions provided us with the opportunity to examine the joint effect of the evidence and the initial hypothesis on subjects’ ratings of the target person and their explanations for those judgments. Subjects’ trait ratings revealed a main effect for the type of evidence received, \( F(1, 75) = 274.56, p < .001 \). Specifically, subjects who received introverted evidence rated the target as generally introverted \( (M = 3.32) \), whereas subjects who received extraverted evidence rated the target as generally extraverted \( (M = 6.74) \). No other main effects or interactions were significant. Taken at face value, these data suggest that subjects’ impressions are primarily based on the evidence received. That is, subjects’ trait ratings of the target were correspondent with the be-
havioral evidence regardless of whether this evidence was consistent or inconsistent with the initial hypothesis.

We believe, however, that caution must be exercised in relying exclusively on the trait rating data. Subjects were instructed to make their trait ratings based on the information they had gathered up to that point in the experiment. Thus, subjects' trait ratings may reveal their attempt to do what the experimenters asked of them—provide a rating of the target person based on minimal evidence (cf. Miller, Schmidt, Meyer, & Colella, 1984). Moreover, the scale itself, with only the endpoints labeled, does not provide a clear way to indicate one's judgmental uncertainty (Devine, 1989; Miller & Rorer, 1982; Miller et al., 1984). Devine (1989), for example, found that, under such circumstances, subjects may make extreme attributions about a target's disposition even when they lack certainty about their judgments. Therefore, we believe that to completely understand subjects' ratings, we need to examine their explanations for those ratings. It is in such explanations, rather than in the ratings themselves, that subjects may reveal their judgmental uncertainty (if any exists).

Explanations of impression ratings. Examination of subjects' explanations suggests that the hypothesis does affect how subjects respond to hypothesis consistent and inconsistent evidence. The percentage of subjects generating dispositional attributions (i.e., expressing certainty) and those either hedging or generating situational attributions (i.e., expressing uncertainty) are presented in Table 3. Only 16.5% of the responses could not be coded into these categories; analyses were done on the 83.5% of responses that were codable. Not surprisingly, the majority of subjects (78.8%) who received hypothesis consistent evidence generated dispositional attributions for the target's behavior. Such evidence was expected (as shown in Experiment 1) and led to attributions correspondent with both the evidence and the initial hypothesis. This pattern was identical for both introverted and extraverted hypotheses, resulting in a non-significant Fisher's exact test \((N = 33) p = .22\).

Quite a different pattern emerged in the explanations of subjects who were confronted with hypothesis inconsistent evidence. Recall that, overall, these subjects made trait ratings that were correspondent with the introverted or extraverted evidence they received. However, subjects' certainty in their judgments, as revealed by the type of attributions they

\(\textbf{Note:}\) It is not clear what a rating near the midpoint of the scale would mean to subjects. It may indicate uncertainty about the nature of the target's disposition or it may indicate that the target's disposition is neither very introverted nor very extraverted. In addition, Miller et al. (1984) argued that subjects may believe that experimenters view judgments near a scale's midpoint as evidence of incompetence rather than uncertainty. Their empirical evidence suggests that subjects were less favorable in their impression ratings of others who made moderate ratings than others who made more extreme ratings.
TABLE 3

PERCENTAGE OF DISPOSITIONAL ATTRIBUTIONS VERSUS HEDGING AND SITUATIONAL ATTRIBUTIONS IN TRAIT RATING EXPLANATIONS AS A FUNCTION OF INITIAL HYPOTHESIS AND TYPE OF EVIDENCE RECEIVED IN EXPERIMENT 2

<table>
<thead>
<tr>
<th>Type of evidence received</th>
<th>Hypothesis consistent</th>
<th>Hypothesis inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introverted hypothesis</td>
<td>Extraverted hypothesis</td>
</tr>
<tr>
<td>Dispositional explanations</td>
<td>87.5% (14)</td>
<td>70.6% (12)</td>
</tr>
<tr>
<td>Hedging/situational</td>
<td>12.5% (2)</td>
<td>29.4% (5)</td>
</tr>
</tbody>
</table>

Note. All dispositional explanations were correspondent with the evidence received.

generated, depended on the specific trait being tested, Fisher’s exact test ($N = 33$) $p = .006$. As can be seen in Table 3, when subjects tested an introverted hypothesis and received extraverted evidence, they were highly likely to generate dispositional attributions for the hypothesis inconsistent evidence (93.7%). Specifically, these subjects made dispositional attributions that were correspondent with the extraverted evidence and not with the introverted hypothesis. In fact, more subjects in this condition generated dispositional attributions than in any other condition. However, the percentage of dispositional attributions in this condition was not significantly greater than in the conditions in which subjects received hypothesis consistent evidence (Fisher’s exact $p$’s $> .10$). In stark contrast, when subjects tested an extraverted hypothesis and received introverted evidence, the majority of subjects (64.7%) generated situational explanations for the hypothesis inconsistent evidence or qualified their dispositional attributions with hedging and uncertainty. Thus, although these subjects rated the target as introverted, their explanations revealed uncertainty about the target’s disposition.

The full pattern of the trait rating and explanation data suggest that subjects were not simply influenced by their hypothesis nor were they influenced exclusively by the evidence received. Rather, their judgments were determined jointly by the hypothesis and the evidence. For example, although subjects readily accepted introverted evidence when it was expected (i.e., when testing the introverted hypothesis), when such evidence was unexpected (i.e., when testing the extraverted hypothesis), subjects’ explanations revealed that they felt uncertain about their trait ratings. This pattern makes sense when one considers the range of behaviors associated with extraversion and introversion (Reeder & Brewer, 1979). Because introverted behaviors are within the behavioral repertoire of
extraverts, such evidence did not undermine an extraverted hypothesis. Instead, subjects found reasonable ways to make sense of the unexpected introverted behavior by generating situational explanations or by hedging dispositional explanations. In contrast, when the unexpected behavior was extraverted, subjects appeared to abandon their introverted hypothesis and make confident dispositional extraverted attributions, correspondent with the evidence received. Subjects appeared to be sensitive to Reeder and Brewer's notion that extraverted behaviors are outside the behavioral repertoire of introverts, and, thus, are diagnostic of an extraverted personality.

Second question and expected answer. Subjects' second question selection opportunity provided us with information about how their expectations changed after receiving hypothesis consistent or inconsistent evidence. Experiment 1 demonstrated that before receiving evidence, the majority of subjects select hypothesis true questions and expect to receive hypothesis consistent evidence. This pattern may change when subjects actually receive hypothesis consistent or inconsistent evidence and may depend on the specific trait being tested. Recall that, in the present study, subjects in the no specific hypothesis condition showed no systematic preference for introverted or extraverted questions, and their expectations were guided by the type of evidence they initially received. However, subjects in the specific hypothesis conditions did have question selection preferences that were influenced by both the initial hypothesis and the type of evidence received. The percentage of subjects selecting introverted and extraverted second questions are shown in Table 4. Not surprisingly, subjects who received hypothesis consistent evidence generally selected a hypothesis true question for the second question. The pattern was somewhat stronger for subjects testing an extraverted hypothesis who received extraverted evidence (71.4% selected an extraverted question) than for

<table>
<thead>
<tr>
<th>Type of evidence initially received</th>
<th>Hypothesis consistent</th>
<th>Hypothesis inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introverted hypothesis</td>
<td>Introverted hypothesis</td>
<td></td>
</tr>
<tr>
<td>Introverted question</td>
<td>57.9%</td>
<td>19.0%</td>
</tr>
<tr>
<td></td>
<td>(11)</td>
<td>(4)</td>
</tr>
<tr>
<td>Extraverted question</td>
<td>Extraverted hypothesis</td>
<td></td>
</tr>
<tr>
<td>Extraverted question</td>
<td>28.6%</td>
<td>50.0%</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>(9)</td>
</tr>
<tr>
<td>Extraverted question</td>
<td>42.1%</td>
<td>81.0%</td>
</tr>
<tr>
<td></td>
<td>(8)</td>
<td>(17)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(9)</td>
</tr>
</tbody>
</table>
TABLE 5

PERCENTAGE OF SUBJECTS EXPECTING TO RECEIVE AN INTROVERTED VERSUS AN EXTRAVERTED RESPONSE TO THE SECOND QUESTION AS A FUNCTION OF INITIAL HYPOTHESIS AND INITIAL EVIDENCE RECEIVED IN EXPERIMENT 2

<table>
<thead>
<tr>
<th>Expect to receive</th>
<th>Hypothesis consistent</th>
<th>Hypothesis inconsistent</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Introverted hypothesis</td>
<td>Extraverted hypothesis</td>
</tr>
<tr>
<td>Introverted evidence</td>
<td>62.5%</td>
<td>5.6%</td>
</tr>
<tr>
<td></td>
<td>(10)</td>
<td>(1)</td>
</tr>
<tr>
<td>Extraverted evidence</td>
<td>37.5%</td>
<td>94.4%</td>
</tr>
<tr>
<td></td>
<td>(6)</td>
<td>(17)</td>
</tr>
</tbody>
</table>

subjects testing an introverted hypothesis who received introverted evidence (57.9% selected an introverted question), Fisher's exact test \((N = 40) \ p = .06\) (see left half of Table 4). However, as we have argued, to understand subjects' strategies it is important to examine not only what type of question subjects selected, but also how they expected the target would answer their question.

The type of evidence subjects expected to receive to their second question is presented in Table 5. Analyses are based on the 84.8% of subjects' responses that were coded as Yes or No answers. Given the high diagnosticity of extraverted evidence, we expected that most subjects who were initially testing an extraverted hypothesis and received extraverted evidence would expect an extraverted response to the second question. Indeed, 94.4% of these subjects expected to receive extraverted evidence. Consistent with the question selection data above, this effect was weaker for subjects initially testing an introverted hypothesis who received introverted evidence, Fisher's exact test \((N = 34) \ p = .006\). Only 62.5% of these subjects expected to receive introverted evidence to the second question. These different patterns of selected questions and expected answers for the introverted and extraverted hypothesis conditions may indicate that some subjects testing introversion recognized that the introverted evidence they received did not rule out the possibility that the target was extraverted, and thus, considered extraversion as a plausible alternative hypothesis.

Subjects who received hypothesis inconsistent evidence also differed in their question selection preferences and expected answers. The question selection data for subjects who received hypothesis inconsistent evidence are shown in the right half of Table 4. Many of the subjects initially given an introverted hypothesis who received extraverted evidence appeared to switch to an extraverted hypothesis, as revealed by the 81.0% of subjects
who selected an extraverted second question (which is hypothesis true for an extraverted hypothesis). The subjects initially given an extraverted hypothesis who received introverted evidence, and who had generally revealed uncertainty in the explanations of their trait ratings, showed no question type preferences, with 50.0% selecting an introverted question and 50.0% selecting an extraverted question. These different question selection preferences suggest that the impact of receiving hypothesis inconsistent evidence depends on the specific nature of the trait being tested, Fisher's exact test \((N = 39) \ p = .04\).

The expectations of subjects who initially received hypothesis inconsistent evidence were also in line with Reeder and Brewer's (1979) analysis of hierarchically restricted trait dimensions (see the right half of Table 5). Of subjects initially testing an introverted hypothesis who received extraverted evidence, 84.2% expected an extraverted response to their second question, suggesting that these subjects may have switched to an extraverted hypothesis. Again, this effect was much weaker when the hypothesis inconsistent evidence was introverted, Fisher's exact test \((N = 33) \ p = .0006\). Only 64.3% of subjects initially testing an extraverted hypothesis who received introverted evidence to the first question expected to receive introverted evidence to the second question as well. Thus, it appears that receiving one introverted response results in some ambiguity about how the target might respond to future questions, regardless of whether the evidence is consistent or inconsistent with the initial hypothesis. A single piece of extraverted evidence, however, strongly supports an extraverted hypothesis and seems to undermine an introverted hypothesis, leading subjects to respond as if they have switched to an extraverted hypothesis.

**GENERAL DISCUSSION**

The primary goal of the present research was to move beyond question selection to more fully understand subjects' trait hypothesis testing strategies. Specifically, we conducted two experiments to determine (1) whether hypotheses set up expectancies for the type of evidence that will be received and (2) what the roles are of the hypothesis and the evidence in subjects' impressions of a target person. Experiment 1 revealed that hypotheses do indeed set up expectancies for what type of evidence will be provided by the target person. Regardless of whether subjects were testing an introverted or extraverted hypothesis and whether they selected a hypothesis true or alternative true question, subjects expected to receive evidence that was consistent with the hypothesized trait.

Experiment 2 revealed that subjects' overall impressions of a target person were a joint function of the initial trait hypothesis and the type of evidence received. When subjects received evidence that was consistent with the hypothesis (and with subjects' expectations), their trait ratings
of the target were correspondent with the hypothesized trait. In addition, subjects indicated in their explanations that the target’s response reflected his or her underlying disposition. When the evidence was inconsistent with the hypothesis (and with subjects’ expectations), their trait ratings of the target were correspondent with the hypothesis inconsistent evidence. However, subjects’ explanations of their trait ratings revealed important differences between the introverted and extraverted hypothesis conditions. Hypothesis inconsistent extraverted evidence undermined the introverted hypothesis, leading subjects to conclude that the target was dispositionally extraverted. In contrast, hypothesis inconsistent introverted evidence did not fully undermine the extraverted hypothesis. Subjects who received such evidence revealed their uncertainty in their trait ratings by making situational attributions for the target’s response or qualifying dispositional attributions with hedging.

The results of Experiment 2 are best interpreted in light of the analysis we derived from Reeder and Brewer’s (1979) discussion of the properties of traits. Reeder and Brewer characterized the trait dimension introversion-extraversion as “hierarchically restrictive” such that extraverts can engage in a wide range of behaviors (from extraverted to introverted behaviors) whereas introverts engage in only a narrow range of behaviors (only introverted behaviors). Extraverted evidence undermined an introverted hypothesis because extraverted behaviors are outside the behavioral repertoire of introverts. Introverted evidence, however, did not fully undermine an extraverted hypothesis because introverted behaviors are within the behavioral repertoire of extraverts. Introverted evidence does not seriously challenge nor does it provide support for an extraverted hypothesis, leading subjects to express uncertainty about the target’s internal disposition.

It is of interest to note that although subjects appeared to seek out and anticipate hypothesis consistent evidence (Experiment 1), they did not maintain a “confirmatory bias” in their evaluation of evidence (Experiment 2). That is, when their expectations were challenged by hypothesis inconsistent evidence, subjects did not strongly bias their evaluation of the evidence so as to maintain their original hypothesis (Darley & Gross, 1983). Indeed, when the hypothesis inconsistent evidence was highly diagnostic (i.e., extraverted evidence), subjects abandoned their original hypothesis and concluded that the target possessed the alternative trait. When the hypothesis inconsistent evidence did not permit confident trait attributions (i.e., introverted evidence), subjects manifested appropriate levels of uncertainty.

It appears that subjects evaluate evidence in light of their expectations and the range of behaviors likely, given the presence of the hypothesized trait. Expected (hypothesis consistent) evidence, whether strongly correspondent with the hypothesized trait or not, is accepted and leads to
strong dispositional inferences (cf. Gilbert, 1989). In contrast, evidence that is inconsistent with initial expectancies is given further scrutiny and is more elaborately processed (e.g., Hastie, 1984; Hastie & Kumar, 1979; Srull et al., 1985; Stangor & McMillan, 1992). To guide their inferences about the target, subjects appear to use their knowledge of the properties of the trait being tested (i.e., behavioral restrictiveness) to evaluate the likelihood of the behavior given the hypothesized trait. In light of this analysis, it may not be appropriate to characterize evidence in discrete terms (i.e., hypothesis consistent or inconsistent). Rather, evidence might be better characterized in terms of its likelihood given a particular hypothesized trait. If subjects do indeed view evidence about traits in these probabilistic rather than discrete terms, researchers will need to be aware of this perspective when developing experimental materials and when interpreting the results of new research on hypothesis testing and the impression formation process.

Undermining Hypotheses about Behaviorally Unrestricted Traits

In the context of this discussion, one might conclude that an extraverted hypothesis would never be abandoned in light of introverted evidence. We want to be clear that our focus has been on the initial stages of testing a hypothesis, when only limited information is available (i.e., one Yes or No answer to a question). Pennington (1987) and Swann and Ely (1984), for example, found that in the face of extensive evidence that challenges an initial hypothesis, subjects will conclude that their hypothesis is incorrect, even if the hypothesis is from the behaviorally unrestricted end of a trait dimension, such as extraversion (see Footnote 2). Although this research suggests that hypotheses can be undermined, it is not specific about how the process of disconfirming an initial hypothesis occurs (i.e., subjects received extensive information about the target before rendering any trait judgments). How much evidence is required for subjects to abandon an initial hypothesis or to conclude an initial hypothesis is definitely correct? Do the processes of confirmation and disconfirmation differ for introversion and extraversion? Following Reeder and Brewer’s analysis (1979; see also Rothbart & Park, 1986), subjects may require more inconsistent evidence to abandon the extraverted than the introverted hypothesis. Similarly, subjects may require more consistent evidence to strongly confirm the introverted than the extraverted hypothesis. Rarely are social perceivers confronted with uniform evidence (i.e., all consistent or all inconsistent). How is the hypothesis testing process affected by the presentation of mixed evidence? And does the order of receiving consistent versus inconsistent evidence matter? These questions suggest rich avenues for future research that will provide a more complete picture of the processes by which social perceivers test hypotheses and make judgments about the personalities of others.
Moving beyond Question Selection

We argued that it was important to go beyond question selection in order to more fully understand subjects' strategies in the hypothesis testing process. The findings from the present research help to highlight some of the limitations of relying exclusively on the question selection methodology to study hypothesis testing strategies. For example, the data from Experiment 1 make it clear that strategies cannot be determined on the basis of question selection alone, as has been done in the previous trait hypothesis testing literature. Rather, information gathering strategies are best revealed by considering both the type of question selected and the answer expected. The "confirmatory bias" has been characterized in the question selection literature by the tendency to select hypothesis true questions. We found that most subjects who selected hypothesis true questions expected to receive a hypothesis confirming Yes answer. However, the question selection literature seems to have misclassified subjects who selected alternative true questions as taking a "disconfirmatory" strategy. Our data clearly suggest that subjects who select alternative true questions expect to receive the hypothesis confirming No answer. Indeed, by looking at both the type of question selected and the expected answer, it appears that the magnitude of the confirmation bias (before any evidence is received) is stronger than previously thought. Realizing that subjects expect to receive hypothesis consistent evidence, independent of the question selected, puts hypothesis testing researchers in a better position to make predictions about how they will respond to expected and unexpected evidence.

Perhaps the most significant limitation of the question selection methodology for understanding trait hypothesis testers' strategies is that it does not provide subjects with the opportunity to receive and evaluate evidence. It has been argued (and supported by the present research) that hypotheses set up expectations and that these expectations can influence subjects' responses to evidence (Fiske & Taylor, 1991; Snyder, 1981; Snyder & Swann, 1978). Thus, the role of the hypothesis extends beyond the question selection phase of hypothesis testing. For this reason, it is critical that the trait hypothesis testing field move beyond question selection to examine the process of testing and evaluating a hypothesis. This process consists of many stages including seeking out and evaluating evidence, reassessing the truth value of the hypothesis, seeking out and evaluating more evidence until a final judgment can be made about the target's internal disposition (see Pyszczynski & Greenberg, 1987). The goal of the present research was to understand how the hypothesis influences subjects' evaluation of the initial evidence they receive. Future research will need to examine the later stages of the hypothesis testing process, as subjects integrate several pieces of evidence and come to a conclusion about the
validity of their hypothesis. This step by step examination may be of particular interest to trait hypothesis testing researchers because, as the present research revealed, subjects are sensitive to qualitative differences among traits that have important implications for the extent to which evidence can challenge an initial hypothesis.

One important advantage of moving beyond question selection and considering the role of hypotheses in evaluating evidence is that it highlights the natural connections between the hypothesis testing, attribution, and impression formation literatures. Indeed, our analysis of how subjects would respond to hypothesis consistent and inconsistent evidence was derived from a consideration of both the attribution and the impression formation literatures. Developing these connections will most assuredly enrich the hypothesis testing literature and, we believe, will reveal a coherence among these areas that has not heretofore been fully appreciated.

REFERENCES


Swann, W. B., Jr., Giuliano, T., & Wegner, D. M. (1982). Where leading questions can


