

1-1-1981

The evidence and mechanisms employed in intuitive hypothesis testing.

Daniel Patrick Blyth
University of Massachusetts Amherst

Follow this and additional works at: https://scholarworks.umass.edu/dissertations_1

Recommended Citation

Blyth, Daniel Patrick, "The evidence and mechanisms employed in intuitive hypothesis testing." (1981).
Doctoral Dissertations 1896 - February 2014. 1386.
<https://doi.org/10.7275/cqy5-hk41> https://scholarworks.umass.edu/dissertations_1/1386

This Open Access Dissertation is brought to you for free and open access by ScholarWorks@UMass Amherst. It has been accepted for inclusion in Doctoral Dissertations 1896 - February 2014 by an authorized administrator of ScholarWorks@UMass Amherst. For more information, please contact scholarworks@library.umass.edu.

312066013546914

THE EVIDENCE AND MECHANISMS EMPLOYED IN
INTUITIVE HYPOTHESIS TESTING

A Dissertation Presented

By

DANIEL PATRICK BLYTH

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 1981

Psychology

THE EVIDENCE AND MECHANISMS EMPLOYED IN
INTUITIVE HYPOTHESIS TESTING

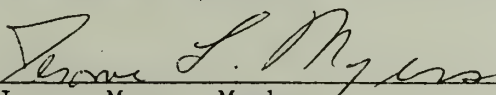
A Dissertation Presented

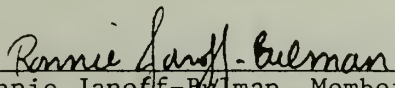
By

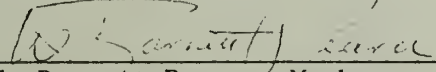
Daniel Patrick Blyth

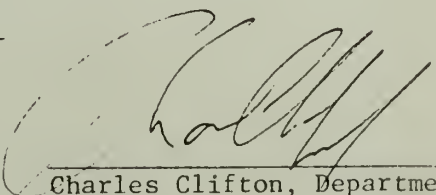
Approved as to style and content by:


Icek Ajzen, Chairperson of Committee


Jerome Myers, Member


Ronnie Janoff-Bulman, Member


W. Barnett Pearce, Member


Charles Clifton, Department Head
Psychology

Acknowledgements

This work grew out of my interest in understanding how people come to their inferences about others. Icek Ajzen promoted that interest and helped me in learning how to conceptualize and test my ideas.

Jerry Myers was, throughout this project, very generous with his time, knowledge, and patience. After each of my meetings with him, I felt more enthusiastic about my work and more confident of my ability to carry it out.

This dissertation is the culmination of my graduate career, a career that has been enriched by my association with Linda Lang-Gunn. She shared her journals and knowledge, and freely gave her friendship and support.

Finally, and most especially, I thank Lynn. I'll remember this dissertation as a long and arduous project that often left me confused, bored, dispirited, angry, and generally a pain in the ass to be with. Lynn graciously endured all this and, even more, provided the love, stability, and encouragement that enabled me to rise above it.

Both as a personal life event and as an academic achievement, the importance of this dissertation will diminish in the years to come. I hope, though, that I remember that it represented my honest effort to do the best I could. I'm proud of what I've accomplished.

ABSTRACT

The Evidence and Mechanisms Employed In
Intuitive Hypothesis Testing

(February, 1981)

Daniel Patrick Blyth, B.A., University of Colorado

M.A., University of Massachusetts, Ph.D., University of Massachusetts

Directed by: Professor Icek Ajzen

This paper examines and critiques experimental approaches to the study of intuitive reasoning. The insufficiency of earlier attempts to characterize human judgment via mathematical or probabilistic models was traced to three sources: the inability of the probability calculus to adequately typify causal relationships, the presumption of analytic cue integration to the exclusion of other (e.g., analogic) processing, and the failure of mathematical formulations to incorporate the rich store of archival information an individual brings to any judgment.

The more recent attempts to describe intuitive reasoning in terms of cognitive heuristics were also examined. While this approach holds forth the promise of a more adequate description of the processes actually used by individuals, its assumption of discrete and task-specific judgmental strategies will obscure and delay the exploration of the fundamental principles guiding the way in which information is used enroute to any decision or inference.

Both the probabilistic and the heuristic models of judgment have identified several errors in intuitive reasoning. One of these is the "confirmatory error"--the tendency to use predominately confirming evidence in evaluating the plausibility of a hypothesis. An experimental study was conducted to better determine the source of this error. Is it due to memory or interpretive errors? Does the hypothesis enhance the retrievability of confirming evidence while masking or suppressing the recall of disconfirming data? Or is the interpretation of retrieved evidence altered such that more data comes to be viewed as supportive of the hypothesis? The obtained results suggest that both the retrieval and the interpretation of evidence is affected by the hypothesis, and these effects are to some extent dependent upon the instructions or orientation brought to the judgmental context. If individuals are asked to review the information they consider relevant, there is a tendency to focus more on confirming than disconfirming evidence. It is suggested that this tendency arises because the instructions encourage the use of a comparison-to-prototype decision strategy, and the prototype is weighted in favor of the hypothesis. Under more critical instructions when the decision maker is prompted to first review all available evidence about the issue, and then to evaluate each bit of that evidence for its implications for the hypothesis, the confirmatory error is less likely to occur. The findings also suggest that under either orientation there is a tendency to interpret ambiguous or ambivalent evidence as supportive.

The paper concludes with the challenge to investigators of social cognition to resolve an apparent paradox: The experimental literature has shown intuitive reasoning processes to be seriously and pervasively flawed, yet the impression of reality is that individuals, and the species, reason appropriately--they generally manage their daily affairs well, avoiding harm to themselves and others, they learn from their mistakes, and they adapt their decisions accordingly in light of past experience. One of the two impressions is misleading. Either the optimistic view of reality is yet another example of cognitive errors or the pessimistic impression created by the literature is overstated.

TABLE OF CONTENTS

| | |
|--|-----|
| ACKNOWLEDGEMENTS | ii |
| ABSTRACT | iii |
| Chapter | |
| I. PREDICTIVE MODELS OF JUDGMENT | 1 |
| Irrational Reasoning or Unequivalent Information | 3 |
| Modelling Causal Reasoning | 16 |
| II. DESCRIPTIVE MODELS OF JUDGMENT | 21 |
| Cognitive Heuristics | 22 |
| The Theory of Lay Epistemology | 24 |
| Testing Hypothesis: The Evidence Used | 26 |
| Testing Hypotheses: The Process Used | 31 |
| III. ERRORS AND BIASES IN INTUITIVE HYPOTHESIS-TESTING | 34 |
| IV. THE MECHANISMS INVOLVED IN INTUITIVE HYPOTHESIS- TESTING: DIFFERENTIAL RECALL & DIFFERENTIAL RELEVANCE | 46 |
| Method | 53 |
| V. RESULTS | 58 |
| Equivalence of the Prior Information | 58 |
| Manipulation Check: Ambivalence of the Prior Information | 58 |
| No Effect of the Prior Probability Manipulation | 63 |
| Judgments of Job Suitability | 63 |
| Effect of the Hypothesis on Information Recalled | 66 |
| Effect of the Hypothesis on the Interpretation of Evidence | 67 |
| VI. DISCUSSION | 75 |
| FOOTNOTES | 86 |
| REFERENCES | 94 |
| APPENDIX A - Instructions and Narrative | 102 |
| APPENDIX B - Test of Factual Knowledge | 109 |

LIST OF TABLES

| | | |
|----|--|----|
| 1. | Ambivalence of the Prior Information | 61 |
| 2. | Job Suitability Judgments. | 65 |
| 3. | Proportions of Confirming and Disconfirming Evidence Recalled | 68 |
| 4. | Subjective Interpretations of Recalled Evidence. . . | 71 |
| 5. | Contribution of Each Type of Confirmatory Error To the Total Amount of Confirmatory Errors. . . . | 73 |

C H A P T E R I

PREDICTIVE MODELS OF JUDGMENT

Social judgment, in its broadest sense, is concerned with the way in which people develop beliefs, make inferences, generate predictions, and evaluate conclusions about other people or events. Defined this globally, the field has a long history of research and includes such diverse areas as attribution theory, models of decision-making, and interpersonal judgment. Although wide-ranging in terms of the particular judgments and contexts examined, these different content domains share a common objective, the accurate description of intuitive reasoning.

An obvious difficulty is that researchers cannot observe the reasoning process per se. The way in which a decision is made can only be inferred from the decision itself. Under this constraint much of the initial research on social judgment adopted the procedure of comparing informal or intuitive judgments against those which would be prescribed by normative or statistical models of information usage. Normative models, based on statistical and probability theory, are an attractive standard of comparison because they represent "rational" inferential strategies, rational in the sense that the informational parameters obey certain rules, and the entire process is information-bound. Judgments are based solely upon available information and they are therefore unaffected by "irrational" factors such as the arbitrary prejudices or desires of the particular judge making the decision. In addition to their freedom from

motivational biases, rational decisions are similarly unencumbered by cognitive failings. They are not, for example, victimized by inefficient organization in memory nor by memory losses.

A great many comparison-to-normative studies have been conducted, but the obtained results are conflicting. Peterson and Beach (1967) reviewed over one hundred such studies and concluded that "normative models provide a good first approximation for a psychological theory of inference. Inferences made by subjects are influenced by appropriate variables in appropriate directions" (pp. 42-43). Slovic and Lichtenstein (1971) conducted a similar review and came to a similar conclusion. Linear statistical models were found to account for most of the variance in subjective inferences. Not all researchers, however, share these conclusions. Simon (1957) decided that people have a limited capacity for rational (i.e., normative) thinking and operate instead on a principle of "bounded rationality." This bounded rationality "is not even approximately optimal" (p. 198). Slovic (1972) reviewed more than twenty-five studies of subjective decision-making and concluded that intuitive processes "show large and consistent biases" (p. 41). Slovic, Fischhoff, and Lichtenstein (1976) found that people "systematically violate the principles of rational decision-making when judging probabilities, making predictions, or otherwise attempting to cope with probabilistic tasks" (p. 169). Hammond (1974) examined decisions made about social policy and came to the pessimistic conclusion that "man's cognitive capacities are not adequate for the tasks which confront him" (p. 4).

Irrational Reasoning or Unequivalent Information

These different conclusions create the impression that sometimes people reason rationally and sometimes they don't. This need not suggest, however, that intuitive judgmental strategies are capriciously invoked, that people choose to obey statistical rules when they feel like it and avoid them when, for arbitrary reasons, they do not. Comparison-to-normative studies do not illuminate what people actually do with information. For the most part, they can only show whether or not the intuitive scientist has reached the normative solution. When discrepancies between intuitive and statistical judgments occur, the researcher cannot know, in most cases, whether the discrepancy arose because the individual used different information enroute to the decision or because he or she employed an inappropriate decision rule. Unfortunately, this distinction has often been overlooked. For the most part, such discrepancies have been cited as evidence of irrational processes when the alternative explanation is more likely. Another look at the literature suggests that as statistical models include more of the information subjects perceive to be relevant to their decisions, they more accurately predict subjective judgments.

Consider, first of all, the studies which show an agreement between statistical and intuitive judgments. In the main, this concordance in judgment has been found when the stimuli involved have been non-social (e.g., bundles of sticks, poker chips, flashing lights, etc.). On the other hand, disagreements between the

"statistical man" and the "intuitive scientist" tend to be more common and most pronounced when the stimulus materials portray events with which the subject already has some familiarity (e.g., people's personalities, naturally occurring event sequences, and so on). When the statistical man and the intuitive scientist begin with the same prior information--that is no prior information--their decisions are in greater agreement. However, when the intuitive judge brings to the experimental session some assumptions and beliefs about the class of events portrayed in the stimulus materials, and when those assumptions are not shared by the statistical man, their judgments diverge.

In the typical procedure in normative studies of judgment the experimenter constructs stimulus materials to vary certain parameters. This manipulation allows the researcher to calculate the statistically optimal use of the information provided. From the experimenter's point of view, the final decision is a function of the information he or she provides to the subjects. However, the subjects themselves are not similarly constrained solely to whatever information is made available during the experiment. The "illusory correlation" studies provide a good example. Chapman and Chapman (1967) paired Rorschach percepts (responses) with two statements of the clinical symptoms of the purported patient who was alleged to have made the response. Subjects were shown thirty such pairs, each representing a different patient. After viewing all thirty pairs, subjects were given a symptom and asked to recall the most prevalent Rorschach percept by

patients with that symptom. The normatively correct procedure for estimation of covariance consists of computing a correlation coefficient based on equal attention to both diagonals in the 2×2 contingency matrix. In this experiment (Experiment 2), the stimulus materials were prepared in such a way that no correlation actually existed between percepts and symptoms. While the relationship was not actually present in the experimental materials, it was consistent with people's typical assumptions (as measured with pretest subjects) about the correlation between certain symptoms and particular traits. Subjects' estimates of the correlation in the stimulus materials were thus strongly affected by their previous beliefs about what they believed to be true in the general population, regardless of what was said to be true during this one experimental session. Berman and Kenny (1976) reported very similar findings. Inferences about the personality characteristics of a fictitious person were found to be substantially influenced by the assumed covariances of those characteristics in people at large, independent of their correlation in a sample of stimulus materials. This reliance on prior beliefs rather than only upon the empirical relationships displayed in contrived experimental information has been replicated in a variety of contexts (cf. Ajzen, 1977; Einhorn & Hogarth, 1978; Jenkins & Ward, 1965; Snyder & Uranowitz, 1978; Smedslund, 1963; Ward & Jenkins, 1965; Wason, 1960, 1968, 1969).

In numerous other examples, biases or errors were attributed to subjects when the experimental results do not necessarily support

that view. In a recent series of investigations, Ross and his colleagues have examined the persistence of intuitive beliefs. Of particular interest was the continued reliance on those beliefs when the evidence presented argued against their validity. Although a variety of experimental settings, tasks, and methods for discrediting evidence have been employed, the basic procedure in all these studies has been fairly consistent.

Subjects receive false feedback after each trial in a novel discrimination, judgment, or performance task. After this manipulation of subjects' impressions of their abilities, the experimenter "discredits" the evidence upon which the impression was based. Discrediting consists of a debriefing session in which the subject learns that the feedback was totally arbitrary and unrelated to his or her actual performance. Other discrediting procedures have included providing subjects with an alternative explanation of their performance. For example, in a "logical problem solving task" discrediting was accomplished by informing subjects that good or poor teaching methods, rather than their own abilities, were the cause of their success or failure (Lau, Lepper, & Ross, 1976). Following this discrediting session in which they acknowledge their understanding of the experimental deception, subjects are asked to predict their performance on similar tasks in the future. The consistent finding is that the false feedback "perseveres" and continues to strongly influence subjects' predictions despite its extensive and acknowledged discrediting.

These results, that predictions about performance in the future are consistent with knowledge about prior performance, can be explained without assuming an unwarranted reliance on invalidated beliefs. Since subjects' use of a priori understandings was not directly assessed, it is not certain that such understandings were actually invoked. It is possible that subjects (or at least some of them) had no idea of the reasons for their success or failure in these novel tasks. In estimating their performance on subsequent tasks, they may have adopted the reasonable strategy which prescribes that the best predictor of future ability is past ability. Although the evidence about their ability was questioned, it could, especially in the absence of any other predictive base, continue to serve as the best estimate of their subsequent performance.

In more recent work, Ross and his associates have extended this research to examine more directly the survival of intuitive beliefs or theories themselves. In one experiment, theories were provided for subjects (Anderson & Ross, 1978). At the beginning of the experimental session subjects were led to believe that a relationship existed between a personality characteristic (riskiness) and performance (success or failure as a fireman). After informing subjects of the "theory," concrete examples ostensibly confirming the theory were presented. As in the previous studies, subjects were then debriefed, i.e., they were told that the experimenter had made up both the theory and the examples. Following this debriefing session, subjects were asked to predict personality traits (riskiness)

on the basis of performance information--a test of their continued reliance on discredited theories. The results indicated that subjects continued to believe that the initially presented theory characterized the true functional relationship. Their predictions were in accordance with the fictitious theory.

In another experiment reported by Anderson and Ross (1978), subjects were not provided with an a priori theory, but were instead given information about the personalities and performances of two persons and were asked to discover the relationship between traits and performance. After the functional relationship had been discovered, the debriefing procedures were employed. Again, the theories survived the debriefing sessions and were used as the basis for subsequent predictions.

These results are said to demonstrate that once a theory is accepted as a plausible explanation for the observed relationship between antecedents and consequents, it survives the erosion or discrediting of its evidential base. Whether the theory was provided by the experimenter or induced by subjects themselves, it continued to influence their predictions even after the theory and its supporting data were acknowledged to have been fabricated by the researcher. However, this finding reflects an unwarranted use of disconfirmed theories only if the theories have no other support than that provided, and discredited, during the experiment.

Consider, for example, a subject who is told that he or she has been successful in discriminating between genuine and bogus suicide notes (Ross, Lepper, & Hubbard, 1975). Since no "theory" explaining the causes of such successful discrimination was provided, the subject is free to call upon his or her own beliefs. To explain this apparent success the subject can invoke a generalized theory of performance in any task, reasoning, perhaps, that success arises from a combination of sufficient innate ability and practice. The subject then searches for evidence attesting to his or her possession of these causal properties. (One subject attributed his success to his prior reading of the works of a suicidal author---evidence, perhaps, of practice.) This theory explaining success stands as a generalized rule. It has accounted for success or failure in countless previous tasks. There is no reason to suppose, and indeed it would be irrational if, this validated theory were abandoned in the face of the single experimental instance in which it may or may not have been confirmed. In this study, the evidence was questioned, but it was not denied. Despite the fact that after debriefing the subject is unsure of his or her actual performance, the theory relating ability and/or practice to performance is still valid, as is the historical evidence for his or her possession of the causal factors indicated by the theory.

In the second series of experiments by Ross and Anderson, the experimenter provided the theory. For some subjects, riskiness was said to be positively related to a fireman's performance. For others, the converse was said to be true. In either case, the subjects found

riskiness to be a plausible cause of either success or failure. Anderson and Ross report that subjects in both conditions were able to explain why riskiness was related to competence in a fireman. Riskiness is a plausible cause of competence because examples of instances in which its presence or absence co-varied with success or failure are available in memory. Both theories are thus supportable by historical evidence. Even though the experimenter said he concocted it himself, subjects were able to support the theory with remembered instances. They could therefore continue to use the theory as a predictor. It is easy to imagine that quite different findings would have been obtained if the theory provided by the experimenter were implausible, i.e., if it were not supportable by evidence obtained outside the experimental session. If, for instance, the theory had suggested a functional relationship between hair color and fire-fighting ability, the perseverance effect would probably not occur. Such a theory would be quickly abandoned after debriefing.

Ross et al suggest that their research demonstrates "the survival of theories in the absence of evidence" (Anderson & Ross, 1978). That is true only if evidence is limited to that presented in experimental sessions. As in many of the studies cited previously, it denies the evidence subjects bring with them to the experiment.

Studies of this kind make it clear that individuals have and use a rich store of historical knowledge to guide their inferences and decisions. They do not base their judgments solely on whatever information is currently available, but freely and confidently draw

upon their stored understandings of similar people and events. As long as statistical models do not take into account historical knowledge, discrepancies between intuitive and normative judgments are to be expected. A principal source of such discrepancies can thus be traced to differences in information and not to fallacious or irrational reasoning on the part of the intuitive decision-maker. Peterson and Beach (1967) recognized this point after their review: "If the statistical man were to incorporate subjects' assumptions, his inferences would be more descriptive of those made by subjects" (p. 43).

Prior beliefs and assumptions can be included in statistical decision-making. Bayesian models, for example, specify the way in which new information should be reconciled with existing or prior information. The latter is represented in Bayes' theorem as the prior probability or expected value. A large number of studies have shown that the inclusion of prior information decreases the discrepancy between normative and subjective judgments. Although individuals often tend to be conservative in their probability revisions, Bayes' theorem has been shown to be a reasonably good predictive model of many judgments, including trait inferences (Peterson, Ulehla, Miller, Bourne, & Stilson, 1965), impression formation (Cohen, 1973), interpersonal relations (McNeel & Messick, 1970), and attribution processes (Ajzen, 1971).

However, Kahneman and Tversky (1972, 1973) have found that people do not judge in a manner consistent with normative principles even when prior odds are included in the normative model. On the basis of a series of experiments they concluded that people "do not appear

to follow the calculus of chance or the statistical theory of prediction" (1973, p. 237) and that "man is not a conservative Bayesian; he is not a Bayesian at all" (1972, p. 450).

In one experiment, subjects were given brief descriptions of five people. One description, for example, described Jack as a 45 year old man with a wife and four children. Jack was said to be conservative, careful, ambitious, with no interest in politics or social affairs, and was said to spend his free time on carpentry, sailing, and mathematical puzzles. Prior odds were manipulated by varying the sample size from which the descriptions were drawn. Subjects in one condition were told that the sketch was randomly chosen from summaries prepared by a panel of psychologists who interviewed 30 engineers and 70 lawyers. Subjects in the other base rate condition read that the psychologists had interviewed 70 engineers and 30 lawyers. After presentation of this information, subjects were asked to judge whether Jack was an engineer or a lawyer. Kahneman and Tversky found that subjects in both conditions decided that Jack was more likely to be an engineer. The Bayesian procedure prescribes that the decision should be a function of two factors: the prior odds, i.e., the prior probability that any person in the sample is a lawyer or an engineer, and the probability that the description provided fits a lawyer or an engineer. In this case, the prior odds were given by the sample sizes. In the high-engineer condition, the prior probability was 70/30. Conversely, in the low-engineer (or high-lawyer) condition, the prior odds were 30/70. Given the marked difference in prior probabilities, subjects in the high-engineer

group should predict that Jack is more likely to be an engineer, while the other subjects should be similarly influenced by their prior odds and judge that Jack is a lawyer.

In another demonstration, Tversky and Kahneman (1974) told their subjects about a certain town having both a large and a small hospital. Forty-five babies are born each day in the larger hospital and fifteen in the smaller. When asked which hospital recorded more days on which more than 60% of the babies born were male, most subjects confidently answered that there was no difference, that each hospital would have recorded an equal number of such days. According to sampling theory, the smaller hospital should have been chosen. Because the number of births is less in the smaller hospital, the variance in the ratio of boys and girls will, most likely, be higher. The chances are therefore greater that the smaller hospital recorded more days during which the number of boys was significantly greater than the number of girls. The subjects in this experiment, however, did not let this consideration influence their predictions.

Judgments by subjects in both of these problems reflect a lack of attention to base rates. In the engineer example, the optimally correct judgment is determined, in part, by the respective sample sizes. Similarly, in the second example, the choice of hospitals is, normatively, dependent upon the sample size. Results of these and similar studies (Bar-Hillel, 1975; Hammerton, 1973; Lyon & Slovic, 1976; Nisbett & Borgida, 1975; Nisbett, Borgida, Crandall, & Reed, 1976) have been interpreted as evidencing a bias

in intuitive reasoning, a bias in that relevant information about the statistical parameter is not weighted by the size of the sample from which the parameter was derived.

However, it is not the case that people simply do not attend to sample sizes or base rates. Tversky and Kahneman (1973) did find that the effect of the manipulation of base rates was significant in their study; it was, however, overshadowed by the specific information about Jack. They also found that when no other information about the person is provided, prior probabilities guide subjects' judgments. Also, Bar-Hillel (1975) found that inferences and predictions are appropriately influenced by base rates when those base rates are extreme. Some recent research on this issue has helped explain why prior probabilities and base rates are relatively ignored in some cases and appropriately relied upon in others.

Ajzen (1977) asked his subjects to assess the likelihood that an hypothetical student had passed a final exam. A brief description of the student was provided, as was base rate information. Two different forms of the base rate information were used. Some subjects were told that the hypothetical student was a member of a class in which 75% of the students passed (or, in another condition, failed) the exam. Subjects in the other base rate group read that a psychologist chose to interview a sample of which 75% had passed (or failed), and that the hypothetical student was a member of the chosen sample. Clearly, the base rates in the first condition imply something about the difficulty of the exam, a factor causally relevant in an

estimate of the likelihood of any student's success on the exam. The second base rates carry no such causal implications; they merely summarize the interviewing preferences of the psychologist. Ajzen's data indicated that subjects in both conditions took into account the base rates provided, but causal base rates much more strongly influenced subjects' predictions. Note that the Bayesian solution to this problem would be the same in both base rate conditions--the numerical values of these prior probabilities are identical. However, it is their connotative implications, and, more accurately, their causal implications that render them more or less relevant to the judgment. If base rates provide causally relevant information, they appropriately influence inferences. If, on the other hand, they merely summarize the frequency distribution of non-causal variables, they are given little or no weight in subjective decision-making. Base rates in the form of distributional frequencies impart no causal information. They serve only to express highly generalized expectancies about any member of the sample selected at random. In these experimental examples, and indeed in most inferential tasks, people are required to make predictions about particular individuals or events, not about any person or event of a certain kind taken at random.

If subjects in Tversky and Kahneman's studies had been asked the same question about people's occupations or about other pairs of hospitals, and if they continued ignoring base frequencies and sample sizes, they would, in the long run, err more often than they were correct. However, it may not be fair to condemn subjective

reasoning as fallacious simply because it would render erroneous results in the long run. It may be better to evaluate subjective decisions on logical grounds rather than on purely statistical grounds. Subjects in Tversky and Kahneman's studies followed a logically reasonable course; they invoked their intuitive beliefs to assess whether any of Jack's or the hospitals' characteristics were of the kind known to them to cause (or co-vary with) the suggested outcomes.¹

Modelling Causal Reasoning

The studies conducted by Ajzen show that not all information is equally relevant in subjective judgments. Although sample sizes may be critically important to the statistical man, they are usually less important to the intuitive scientist. Ajzen's results point to the predominant role of causality in human reasoning. People think in causal terms, and this has been unambiguously documented in varied contexts. Michotte (1963) has shown that individuals tend to perceive causality merely from the observation of temporarily contiguous events. Piagetian theory (Piaget, 1972; Piaget & Inhelder, 1969) holds that understandings about causality are an inherent part of and concomitant with the process of cognitive development. Numerous learning studies have documented that people rarely see events as randomly determined (cf. Peterson & Beach, 1967; Smedslund, 1963; Ward & Jenkins, 1965). Because events in their day-to-day lives are the product of identifiable causes, people do not regard chance as a

suitable explanation for events in experimental settings. Subjects believe that patterns are present in experimental materials and that the underlying regularities can be discovered, often despite the researchers instructions to the contrary. (In many cases, the subjects are correct. Jones and Myers (1966) have shown that sequences experimentally randomized in short block are often not truly random at all and that subjects can outguess the experimenter. See Myers, 1966, for a review.) Heider (1958, Kelley (1967, 1973), and other attribution theorists (cf. Jones & Davis, 1965) have demonstrated the prevalence and impact of causal reasoning in person perception (see Janoff-Bulman's review, 1977). Ajzen's (1977) data and results of similar studies by Tversky and Kahneman (in press) show that people look for and prefer causal antecedents even when they are, from a statistical point of view, no more diagnostic than other information. In addition to the tendency to prefer causal evidence upon which to base a prediction about the future, people prefer causal explanations for events that have already occurred, even when other reasons (e.g., sampling variability, regression to the mean) are available as sufficient alternative explanations.

Normative statistical models, however, cannot accommodate this fundamental feature of human reasoning, largely because they do not define the temporal ordering of events. In a recent paper, Tversky and Kahneman (in press) describe a set of experiments originally conducted by Turoff (1972). Subjects were asked to estimate the likelihood of one event in light of another event. The information provided to subjects in one study was the following (as summarized by Tversky

and Kahneman):

Let C be the event that within the next five years Congress will have passed a law to curb mercury pollution. Let D be the event that within the next five years the number of deaths attributed to mercury poisoning will exceed 500. Let \bar{C} and \bar{D} indicate the negations of C and D, respectively.

After presentation of this information, subjects were asked to estimate which member of two pairs of conditional probabilities was most likely:

$$1. \quad p(C/D) \quad \text{or} \quad p(C/\bar{D})$$

and

$$2. \quad p(D/C) \quad \text{or} \quad p(D/\bar{C})$$

Most subjects estimated that Congress is more likely to restrict mercury pollution if the death toll rises than if it does not. The majority of subjects also judged that the death toll is less likely to rise if the law is enacted than if it is not. That is, subjects estimated:

$$1. \quad p(C/D) > p(C/\bar{D})$$

and

$$2. \quad p(D/C) < p(D/\bar{C})$$

While these are intuitively reasonable judgments, they violate the normative statistical conventions.

Event sequences can be linked statistically only through conditional probability statements. The normative judgment in this problem, given by the rules of conditional probability, is that

if $p(C/D) > p(C/\bar{D})$ (as subjects in this experiment estimated)
then $p(D/C) < p(D/\bar{C})$

This discrepancy between intuitive and statistically optimal judgments arises because conditional probabilities are temporarily ambivalent; neither event is assumed to precede or cause the other. Subjective decisions, however, routinely make assumptions about the temporal order and causal relationships between events. The difference between normative and informal estimates in judgments of this kind does not reflect irrationality or even bounded rationality on the part of subjects. The subjective decision can be shown to obey logical and probabilistic rules if the probability calculus is expanded to include the temporal order of events. Suppes (1970) has suggested one way to describe causal relationships in probabilistic terms. His work does not constitute a full probabilistic theory of causality, and given the prevalence of causal reasoning in human judgment, much more work is needed. Nevertheless, even the formative bases of such a theory provide a better description of intuitive causal reasoning than that afforded by conditional probability statements.

Given two events, C and D, D is the cause of C if:²

$$p(C_t, / D_t) > p(D_{t'}, / C_t)$$

and,

$$p(C_t, / D_t) > p(C_t, / D_t) > p(C_t, / \bar{D}_t)$$

where,

$$t < t' \quad (t \text{ precedes } t')$$

and,

$$p(D) > 0$$

If the first two statements are assumed to be true, as they apparently were, and reasonably so, by subjects in Turoff's study, they imply

$$p(D/C) < p(D/\bar{C})$$

which is the judgment Turoff's subjects rendered. Although contrary to the rules of conditional probability, it is logically correct. The fallacy here does not rest with subjects' reasoning, but with the attempt to model that reasoning via an inappropriate statistical formula.

C H A P T E R I I

DESCRIPTIVE MODELS OF JUDGMENT

The inability of the probability calculus to accomodate causal principles, given that causality plays such a central role in human reasoning, represents a serious limitation of the adequacy of statistical models of the inference process. Furthermore, even when statistical models are shown to be predictive of intuitive judgments, there is no suggestion that statistical reasoning bears any resemblance to the actual processes people employ in their decisions. Recent research on social judgment has turned away from statistical models as analogues and has aimed toward a more descriptively accurate explanation of intuitive reasoning. Most current research has acknowledged the importance of previously stored information--the knowledge that individuals have accumulated over time--and recognizes that they do not, neither in their daily lives nor in experimental tasks, judge each new instance as if it were wholly independent from any other previous experience. People bring with them remembered examples of similar instances and evaluate new occurrences in light of previous ones. Any descriptive theory of judgment should provide an explanation of how people invoke their stored knowledge enroute to inferences about the particular people or events at hand. This explanation would necessarily include assumptions about the way in which information about prior instances is represented in memory, and about how that information is accessed and used.

Cognitive Heuristics

In the series of articles cited above, Kahneman and Tversky (1972, 1973; Tversky & Kahneman, 1974) proposed that people have developed a limited number of judgmental "heuristics". These heuristics are described in terms of cognitive rules-of-thumb, short-cut methods of using stored beliefs to infer or predict something about new people or events. According to Kahneman and Tversky, people use different heuristics in different judgmental tasks.

The "representativeness" heuristic, for example, is said to be used in determining whether or not a particular instance is an example or a member of a superordinate category. In judgments of this kind, people consider the extent to which the essential features of the new instance are representative of or similar to the defining characteristics of the category. To return to the lawyer/engineer example, Kahneman and Tversky (1973) suggest that subjects inferred Jack's occupation by comparing his attributes to those of the most typical or most representative lawyers and engineers. Because Jack's traits and preferences were more like those of the stereotypical engineer, subjects assigned him to that category.

The "availability" heuristic is said to guide judgments of frequency or probability. Judgments of the likelihood of occurrence are based on the ease with which similar occurrences can be recalled from memory. In one demonstration of this heuristic, subjects were asked whether the letter K appeared more often in the first or third position of English words. Because words beginning with K are

more easily retrieved from memory than words which have a K as the third letter, subjects guessed, erroneously, that K is more likely to appear in the first position (Tversky & Kahneman, 1973).

The "anchoring and adjustment" heuristic is said to be employed in judgments of numerical values when some initial value is made available and salient. For example, subjects in one experiment were initially given a randomly generated number. They were then asked to indicate whether the random number was higher or lower than the number of African nations in the United Nations. Finally, subjects were asked to estimate the actual number of African members in the United Nations. Despite the fact that the initial value was known to be arbitrarily drawn, it markedly affected subjects' final estimates. When the random number was high, subjective estimates of African members were high. When the initial value was low, subjects' guesses were correspondingly lower. It appeared that subjects followed a strategy of adjusting their estimates from the initial anchoring value, even though that anchor was clearly irrelevant to the question (Tversky & Kahneman, 1974).

Tversky and Kahneman's work on heuristics is an important step in the development of a theory of intuitive reasoning. However, there are some troublesome aspects of a conceptual framework which posits discrete and disconnected judgmental strategies. Particularly problematic is the assumption of task-specific judgmental processes. This implies that there may be as many discrete heuristics as there are types of decisions, with the attendant implication that a theory of subjective

judgment would consist of little more than a catalogue of qualitatively distinct tasks, each with its matching heuristic. Secondly, even within a single type of task, different heuristics can compete as explanations of the processes employed. For example, perhaps subjects judged Jack to be an engineer, via the representativeness heuristic, by assessing the similarities between his characteristics and those of the stereotypical engineer. But couldn't the same judgment be explained in terms of the availability heuristic by assuming that examples of engineers who possessed characteristics like Jack's were more available in memory than examples of lawyers who possess Jack's traits?

The Theory of Lay Epistemology

Kruglanski and his colleagues have offered an alternative framework by which to analyze and understand intuitive reasoning. The theory of lay epistemology (Kruglanski, Hamel, Maides, & Schwartz, 1978; Kruglanski & Ajzen, Note 1) views the process as a sequential series of steps by which an individual uses both historical knowledge and current information to generate inferences and make judgments about a particular person, entity or event. This theory assumes that there exists a single fundamental judgmental process, one that is relatively invariant across different tasks and that is stable regardless of the kind of knowledge relevant to the question at hand. Individuals are likened to scientists in their pursuit of answers to their questions. According to the theory, the course of the "epistemic episode" leading to any inference or judgment involves the following

stages.

I. Problem Initiation

In everyday affairs, questions requiring a decision or judgment emerge from or are initiated by the individual's experiences. In psychological experiments, the problems are posed by the researcher.

II. Problem Formulation

Whether self-generated or asked by someone else, the question, once formed, leads to the generation of potential answers or hypotheses. Like their scientific counterparts, intuitive hypotheses represent the alternative plausible answers to the particular question. To borrow Kruglanski and Ajzen's example, a person may want to know why his car failed to start on a cold morning. The hypotheses consist of those factors known or assumed to be associated with the normal functioning of his engine--the battery, ignition, etc. The number of hypotheses generated is determined by the knowledge and creativity of the intuitive scientist. Someone who is totally ignorant about automobiles would be hard pressed to generate many hypotheses to explain why his car isn't working. On the other hand, a trained mechanic would be able to list a variety of potential causes of the engine's failure.

In other contexts, the rival hypotheses may be provided by someone else. In experiments, for example, the researcher may pose the initial question and may accompany it with several potential answers.

III. Problem Resolution

During this stage in the epistemic sequence, the individual evaluates the relative plausibility of the competing hypotheses by considering whether they are consistent or inconsistent with all the available evidence. Consistent evidence increases the individual's confidence in a hypothesis and inconsistent evidence weakens that confidence.

IV. Termination

It is possible, in principle, to have an indefinite number of alternative hypotheses for any given question. Intuitive hypothesis-testing could, then, proceed indefinitely. However, the epistemic sequence is, in practice, bounded by several factors. The intuitive scientist comes to accept one hypothesis and to reject the remainder on the basis of the importance of the question, his or her tolerance of ambiguity, the relative strength of support of one hypothesis over others, plus a variety of motivational influences.

Testing Hypotheses: The Evidence Used

Perhaps the majority of the research on social judgment has been concerned with the problem resolution phase of the epistemic sequence, the stage during which the intuitive scientist tests the plausibility of rival hypotheses. In order to assess those hypotheses the individual draws from his or her total store of knowledge that subset which would constitute evidence for or against them. According to Kruglanski et al, evidence is that knowledge or that set of beliefs

that is relevant to the proposition under consideration. They propose that relevance can be defined as material implication. More formally, some belief X is relevant to some hypothesis H

$$\text{IFF } \{X \rightarrow H\} \text{ or } \{H \rightarrow X\} \text{ or } \{X \rightarrow \bar{H}\}$$

It is not clear that the concept of psychological relevance should be made equivalent to the proposition of material equivalence as it is defined in the calculus of formal logic. However, this definition is sufficient to impart the reasonable assumption that two occurrences are relevant if they co-vary and/or share common properties. Knowledge of the traits and preferences of lawyers is relevant to the hypothesis that Jack is a lawyer while the size of the group from which he came is not because the former has co-varied with occupational standing but the latter has not.

The "if and only if" condition of the definition serves to circumscribe the boundaries of relevant knowledge. However, within those boundaries a wide range of beliefs are potentially relevant and could be used to evaluate the strength of one or more hypotheses. Evidence relevant to the hypothesis that Jack is a lawyer could include (in addition to whatever is known about Jack) the traits of a single known lawyer, the traits of several lawyers, and/or those of a stereotypical lawyer. Each meets the criterion of evidential relevance and each could be used as the comparative base against which to judge Jack's characteristics. In most cases, each would yield quite different judgments. Kruglanski's relevance definition does not aid in predicting which type of evidence would be recalled from memory in any judgmental task, nor is there a consensus in other theoretical views

of the judgmental process. Although both specific examples and stereotypical representations are assumed to be available in memory, theories differ in suggesting which would be used as evidence.

In the cognitive literature considerable attention has been devoted to explaining how individuals judge whether a given instance or case belongs in one category or another. In the terminology of lay epistemology, this is an evaluation of the hypothesis that a particular person, object or event is a member of a certain group of like persons, objects or events. One school of thought holds that this decision is made by comparing the particular case not against some known member of the category but against a cognitive summary of the distinctive characteristics of the category. It is assumed that, as a function of experience with multiple exemplars of a group, people abstract its defining features and generate a prototype which embodies those definitive aspects. There is little consensus as to how a prototype should be precisely defined, but it is clear that it represents the central tendency of the category. For groups identified by qualitative attributes (i.e., groups which are defined by color, form, etc.), the prototype consists of the modal values along those attributes. When the exemplars of a category assume values on quantitative dimensions, the prototype represents the mean or average value on those dimensions. Differential salience or importance of attributes is accommodated by a weighting parameter. Thus, the prototype exists as some sort of mathematical combination of the attributes of component exemplars, averaged over all exemplars in the category. According to prototype models, people recognize and

classify new cases by comparing them to the stored prototype.

Most of the research on prototypes has employed non-social stimuli (dot patterns, geometric forms, etc.) and considerable support for this family of models exists in the cognitive literature (cf. Bransford & Franks, 1971; Franks & Bransford, 1971; Posner & Keele, 1968). Prototypes have also been introduced as an explanatory concept in social judgments as well. Tversky and Kahneman's representativeness heuristic is, essentially, a judgment by prototype model. Likewise, Abelson's (Abelson, 1976; Langer & Abelson, 1972; Schank & Abelson, 1977) "scripts" are assumed to be cognitive representations of the diagnostic features of types of social event sequences. More recently, two investigations have examined the utility of the prototype model as an explanation of social categorization. Cantor and Mischell (1977) and Tsujimoto (1978) explored how people use stored information to make decisions about other people's traits. In both studies, results appeared to support a comparison-to-prototype explanation.

An alternative interpretation of the way in which people use acquired information to make decisions about new instances is provided by the "feature" family of models. They differ from prototypes in that they assume that much more information about exemplars, such as their component features, is used during judgmental tasks. In prototype models, the component dimensions are assumed to be independent, and the prototype is aggregated over all exemplars. Thus, the individual category members themselves play no role in determining the classification of new cases. Feature models challenge

this assumption with the argument that such judgments are based on the retrieval of information about the features of specific exemplars.

Included within the family of feature models are the "proximity" model (Reed, 1972), the "context theory of classification" (Medin & Schaffer, 1978), and the "property set" model (Hayes-Roth & Hayes-Roth, 1977). Note, too, that Tversky and Kahneman's availability heuristic proposes that many judgments are based on the retrieval of specific examples that are similar to the instance under consideration. Brooks (1978) also suggests that, under certain learning and task conditions, people reason by comparing the specific case at hand against its most similar exemplar and not against a stereotypical representation of the entire category of similar events.

In short, there are several, and often competing, views of the evidence people recall in testing their intuitive hypotheses. The use of one type over another is surely dictated by several factors, perhaps the most important of which is the variability of the hypothesized class of people or events. In homogeneous groups with exemplars which differ little from one another, a prototype will suffice as an accurate and economical summary of the characteristics defining group membership. However, for "ill-defined" categories composed of multiple exemplars which vary widely, the prototype can only embody the more abstract commonalities. As the prototype becomes more abstract it is a less adequate representation of any single exemplar, it is less adequate as a summarization of the entire category, and it is less useful as comparative evidence in making predictions about any new instance. In other words, while both

prototypes and memory for specific exemplars may be relevant as evidence in testing hypotheses, their degree of relevance is limited by the variability of the categories to which they refer. Evidence can be ordered in terms of its relevance by adding another term in Kruglanski's definition of relevance which substitutes "to the extent that" for "if and only if." This additional statement provides for a ranking of different evidence in terms of their probative values.

Testing Hypotheses: The Process Used

Having gathered the relevant evidence, the task of the intuitive scientist is to test the competing hypotheses in light of the assembled evidence. The way in which this testing proceeds is not well articulated. It is often proposed, for example, that individuals assess the "match" between the evidence and the current instance or that they "compare" its evidentiary features with those of the problematic case. The actual mechanisms involved in "matching" or "comparing" are not clearly defined. In the theory of lay epistemology, it is proposed that people accept or reject hypotheses by applying the "principle of consistency," by considering whether the evidence relevant to a hypothesis is consistent with that hypothesis. Unfortunately, the consistency principle is not developed within the theory. It stands as a general rule and is defined as equivalent to logical consistency: Evidence and a hypothesis are consistent to the extent that they mutually imply each other. An example used to illustrate this concept is the following. The hypothesis that John is a good student would be held on the evidence that he does his homework

because doing homework is consistent with being a good student.

Clearly, empirical research is needed to lend more precision to the principle of psychological consistency. At best, the consistency rule refers to a generalized mode of problem resolution. A handful of theoretical positions have been proposed as more definite explanations of the comparison process. These models may be viewed as types or components of judgments of consistency.

Both prototype and feature models assume that people compare the characteristics of the new case against those of the hypothesized category by computing a metric distance between the two. According to these models, the attributes of the new instance and those of the recalled instance (either a prototype or a specific exemplar) can be thought of as occupying some psychological space. In assessing the similarities between the two entities, the individual is said to calculate their psychological distance, and this distance is a measure of their congruence. In prototype models, the recognition of new instances is proposed to be a function of their distance from prototype. In feature models, the distance of the new case from the exemplar is assumed to determine the classification judgment.

Tversky (1977) has offered a non-metric model of the way in which people assess similarity. According to his contrast model, similarity between entities is a function of the ratio of common to non-common attributes.

Brooks (1978) has proposed that in many tasks, individuals adopt an "analogical" rather than an analytical strategy of comparison.

However, he has not specified exactly how that analogical judgment might be made.

C H A P T E R I I I

ERRORS AND BIASES IN INTUITIVE HYPOTHESIS-TESTING

The theory of lay epistemology, as well as a great many other approaches, likens the naive problem-solver to an intuitive scientist--generating hypotheses and then assessing the strength of their evidential support. While both the goals and the process may be fundamentally the same for the intuitive and the formal scientist, vast differences exist in their problem-solving methods. In teasing out the effects of hypothesized causal factors, or in controlling extraneous effects, the researcher can hold some variables constant while manipulating others. The intuitive scientist has no such measure of control. He or she must disentangle the effects of causal factors sheerly by cognitive activity alone, by trying to figure out what the results of such disentanglement might have been. This distinction is a difference between active experimentation and "passive cognition" (Hammond, 1978). Furthermore, the research scientist can suspend a judgment to accept or reject an hypothesis until empirical evidence, collected under relatively pristine conditions, is assembled. The intuitive scientist must usually make his or her decision much more quickly and his or her evidence typically consists not of a newly acquired set of unbiased data, but rather of what he or she can remember. Thus, since his or her evidence is retrieved from memory, the intuitive scientist's hypothesis-testing is vulnerable to all the factors known to influence memory storage and recall. The

evidence is often incomplete, it may have been originally acquired in haste, or incidentally, and it will erode over time. Under all these disadvantages, judgmental errors are bound to occur.

In most of the social judgment literature, "bias" and "error" have been used interchangeably to refer to any source which intrudes upon the subjective inference process and yields a judgment discrepant from the normative solution. Kruglanski and Ajzen (Note 1) argue that it is useful to preserve a distinction between the two terms. Here, bias refers to a tendency to prefer, on arbitrary grounds, one hypothesis over another. Errors, on the other hand, occur when information is unwittingly used incorrectly, i.e., when the subjective weight attached to some evidence does not conform to its "objective" weight.³ Biases arise from the motivations of the decision-maker and reflect his or her wishes or desires. Errors are traceable to the information used enroute to the decision. According to this distinction, biases are the result of a deliberate and active selection of preferred evidence; errors are the result of unintentional flaws in the storage or recall of relevant evidence.

Biases and errors can occur at any stage of the epistemic sequence. "Wishful thinking," for example, may lead one to prefer one potential hypothesis over other alternatives and to seek its supporting evidence while ignoring information supportive of the unattractive hypotheses. In instances such as this, the intuitive judge has both self-serving and discrediting evidence available, but chooses to select, or attach unwarranted weight to, that evidence which tends to confirm the desired hypothesis. In their paper, Kruglanski

and Ajzen discuss this and other motivational biases and their intrusion into the reasoning process. (See also Beckman, 1970; Freize & Weiner, 1971; Heider, 1958; Johnson, Feigenbaum, & Weiby, 1964; Jones & Davis, 1965; Kelley, 1967; Miller & Ross, 1975; and Ross, 1977 for other discussions of motivated biases, particularly as they occur in attributional judgments.)

According to the present interpretation, cognitive errors occur because of the information used in testing hypotheses. For any given hypothesis, a wide variety of evidence is potentially available in memory. However, not all that information is accessed during the epistemic process. Retrieval is influenced by factors which increase the "availability" of some information, factors such as salience, primacy, recency, and so on (cf. Tversky & Kahneman, 1973). While these may be unintentional constraints on the retrieval of evidence, their effects nevertheless produce inaccurate judgments.

In addition to these factors which are known to influence the accessibility of stored information, there may be other unintentional intrusions into the epistemic process. In some recent research, Snyder and his colleagues have proposed the idea that the hypothesis itself may create subtle criteria for the selection of relevant evidence. The basic premise is that formulated hypotheses tend to be tested by a "confirmatory strategy." Rather than attending to both supportive and non-supportive evidence, people tend to assess the plausibility of a hypothesis by selectively retrieving and attending to more evidence which confirms it. This tendency is not assumed to be motivated by the desires of the individual (it has been shown to

occur even when subjects are given incentives to be accurate), but is considered to be an unintentional error in intuitive hypothesis-testing.

In one study in this research program subjects were led to believe that they would conduct interviews with other students (Snyder & Swann, 1978). Subjects were told that their task was to use the interviews to gather information to test a hypothesis. Some subjects were asked to find out if the interviewee was an extravert. Other subjects were to assess if the interviewee was an introvert. After presentation of the respective hypotheses, subjects were given a profile describing the traits, preferences, and behaviors of typical extraverts (or introverts), and a list of potential questions that could be asked of the interviewee to test the given hypothesis. Some suggested questions were ones that would, according to the judgment of pre-test subjects, be asked of people known to be extraverts, some would be asked of known introverts, and some were neutral. From these potential questions subjects were asked to select 12 which would enable them to link the interviewee's characteristics to those described in the profile.

The results showed that subjects chose to ask questions that solicited hypothesis-confirming evidence about twice as often as they chose questions that solicited disconfirming evidence. Subjects in the extravert condition tended to select questions characteristically asked of persons already known to be extraverts, while subjects in the introvert condition preferred questions typically asked of known introverts. Although the disconfirming evidence should be important in any decision to accept or reject the stated hypothesis, subjects

overwhelmingly favored the confirmatory evidence.

The tendency to adopt this confirmatory strategy was replicated under several conditions in several experiments. It occurred when the origin of the hypothesis was varied, i.e., whether it was indicated by the results of a valid personality test or simply emerged from an informal character sketch (Snyder & Swann, 1978, Experiment 1). The confirmatory strategy was also unaffected by the likelihood of the hypothesis, where likelihood was manipulated by base rates (Experiment 3), and was similarly unaffected by substantial incentives for judgmental accuracy (Experiment 4). Furthermore, even when subjects were testing hypotheses about themselves, they tended to choose confirmatory evidence even when the hypothesis was not necessarily flattering (Snyder & Skrypnik, Note 2). The only condition under which the effect was not obtained was when no hypothesis was presented. In the absence of either personality profile, subjects were more likely to select questions from both the Introvert and Extravert domain (Snyder & Swann, Note 3).

While these investigations appear to offer fairly convincing support for the pervasive confirmatory error in intuitive hypothesis-testing, certain of the procedures used undermine the validity of the findings. Specifically, the instructions to subjects may have provided strong demand characteristics. Recall that they were asked to choose questions that would enable them to link the characteristics of the interviewee with those in the profile of the typical extravert (or introvert). Under these instructions, subjects may have interpreted

their task not as one of unbiased hypothesis-testing, but rather as one of matching the profiled characteristics with those of the interviewee. If the task were seen as such, it would be reasonable to select those questions for which a match could be obtained, i.e., questions which would provide confirming evidence.

In another series of experiments, the confirmatory strategy was further explored. In this series (Snyder & Cantor, 1979), two important changes were made. First, different instructions and procedures were employed. Secondly, subjects tested the hypothesis by retrieving evidence from memory rather than by gathering new evidence once the hypothesis was presented.

Snyder and Cantor prepared a four page narrative about one week in the life of a woman named Jane. The account was structured to include examples of Jane behaving in an extraverted fashion as well as instances in which she acted in an introverted manner. (On the basis of this account, pretest subjects judged Jane to be both moderately extraverted and moderately introverted.) In the first experimental session, subjects read the narrative and answered 20 questions about specific factual details of the story. In the second session, conducted two days later, subjects were randomly assigned to either the Extravert or Introvert hypothesis condition. Subjects in the former were asked to judge how well-suited Jane was for the job of real estate salesperson. Those in the Introvert condition were to assess Jane's suitability for the position of research librarian. All subjects were given a profile of the ideal person for the respective positions.

The extravert profile indicated that the ideal real estate salesperson was outgoing, talkative, bold, and so on. The ideal librarian was described as reserved, quiet, studious, soft-spoken and discrete.

Before making their judgments of job suitability, subjects were asked to write down all the facts from the story they considered to be relevant to assessing Jane's qualifications for the job in question. Finally, subjects were to rate, on a six-point scale, how well-suited Jane was for the hypothesized position.

The results of the first experiment in this series showed no difference in judgments of job suitability between the Extravert and Introvert conditions. However, subjects in both conditions tended to report hypothesis-confirming evidence as relevant more often than disconfirming evidence. In a second study, the idealized profile was omitted and subjects were free to use their own notions of the best real estate salesperson or the ideal research librarian. Another dependent measure was also added in this second experiment. After deciding Jane's suitability for the hypothesized position, subjects were asked to judge her suitability for the other job as well. Under these conditions, Snyder and Cantor found that subjects rated Jane as more suitable for the hypothesized job than for the other position, even though the story about Jane provided equal evidence for both hypotheses.

Before proceeding to the explanations proffered for these obtained effects, the distinction between this and some similar research should be made clear. Several studies have investigated the

influence of "ancillary" information on the recall and interpretation of previously-learned information. Snyder and Uranowitz (1978), for example, presented subjects with a case history of a woman called Betty. Later, and depending upon the experimental condition to which they were assigned, subjects were told that she had pursued either a lesbian or a heterosexual life style. Subjects were then given a test of their recall of the initial case history (which was, according to pretest subjects, equally supportive of either kind of life style). Snyder and Uranowitz found that when subjects erred in recall, their errors were in the direction of the labelling accomplished via the ancillary information. That is, subjects who were later told that Betty was a lesbian misremembered that her life's events were consistent with stereotypical beliefs about that life style. Subjects in the heterosexual labelling condition also misremembered what they had originally read about Betty such that they recalled her life's events in a manner consistent with typical assumptions about heterosexual women.

In a similar study, Spiro's (1977) subjects read about an engaged couple. The man was said to be opposed to having children and, after discussing this with his fiancée finds that she enthusiastically agrees (or, for some subjects, unequivocally disagrees) with him. In the ancillary information presented later, subjects learn either that the couple eventually were married or that they broke off their engagement.

When subjects in this experiment were tested for their memory of

the first passage, Spiro found that they consistently erred by discounting or forgetting facts that were inconsistent with the final outcomes and over-emphasizing those that were consistent with the knowledge they later acquired. For example, subjects who were originally told that the fiancée vehemently disagreed about having children and later learned that the couple proceeded to get married forgot that the fiancée disagreed, or remembered that she was only mildly opposed to her husband's opinion.

Both these studies evidence the reconstructive facility of memory. Information is not simply stored as it is presented, but is actively interpreted and changed by other information. Subjects in these experiments reconstructed their initial knowledge to make it consistent with the information they later obtained. The difference between this work and that of Snyder and Cantor is that in the former, the ancillary information was presented as a factual outcome. Despite what they had learned in the original passage, subjects came to know that the woman was either a lesbian or a heterosexual. Likewise, Spiro's subjects knew that regardless of their agreement or disagreement on the issue of children, the couple eventually either got married or broke up. In Snyder and Cantor's studies, the ancillary information is not a factual bit of knowledge. It is simply a hypothesis or a possibility that must be evaluated in light of their previously-stored knowledge. When later information is presented as fact, there is apparently a substantial tendency to reconcile it with previous beliefs, to make a coherent whole out of all related

knowledge, regardless of apparent discrepancies in parts of that knowledge. However, when the subsequent proposition is stated not as a fact but as a possibility, there should be no such tendency. Snyder and Cantor's results, however, appear to show that the same kind of errors are made.

Snyder and Cantor suggest that their findings point to a fundamental feature of intuitive reasoning. In testing hypotheses about other people, individuals tend to adopt the confirmatory strategy, asking, in effect, "What do I know about this person that would enable me to support my hypothesis?" rather than asking, "What do I know that would enable me to confirm or disconfirm my hypothesis?" This strategy produces an error in that it tends to increase the likelihood that the hypothesis will be accepted when, as in these studies, both supportive and contradictory evidence is potentially available.

Snyder and Cantor point out that this tendency may arise as a result of either of two mechanisms. First, the error may occur during the retrieval of stored evidence. The statement of the hypothesis itself may induce the selective recall of stored instances or facts that support it. This selective recall could occur if memory is organized in terms of superordinate conceptual categories, and, under them, their defining exemplars. Given such a cognitive structure, subjects in these experiments may have ordered Jane's behaviors under conceptual categories denoting introversion and extraversion as they read the narrative. The presentation of the hypothesis activates the conceptual categories implied by that hypothesis and the task becomes

one of finding the goodness of fit between Jane's behaviors and exemplars ordered under the superordinate conceptual category. For subjects in the Extravert condition, for instance, the task is to match behaviors stored as specific examples of Jane's extraversion against those stored as generalized exemplars of that superordinate concept. Disconfirming evidence is less easily accessed because it is stored as examples of a different category. This interpretation would be consistent with models of memory which view stored information as a network of hierarchical links and nodes. According to these models, memory is interrogated in a sequential fashion, beginning with higher-order generalities and traversing downward until specific behaviors or characteristics are located. (For a discussion of models of this type, see Anderson's HAM theory, 1976, and Collins and Loftus' "Spreading activation" Model. Hastie and Kumar, 1979, among others, propose that memory for persons may particularly be organized and scanned in this manner.)

The confirmatory strategy may be due to an altogether different mechanism. Both confirming and disconfirming evidence may be equally retrievable and recalled in the same way, but greater diagnostic weight may be attached to the former. This would represent more of an active and deliberate strategy on the part of the intuitive scientist and would reflect the subjective belief that the presence of the hypothesized traits and characteristics are worth more as evidence than the presence of alternative traits or characteristics. In other words, confirming evidence is not more available than disconfirming

evidence, but is perceived to be more relevant in testing hypotheses. Several of the "illusory correlation" (cf. Smedslund, 1965) findings have been interpreted as demonstrating this subjective belief. Einhorn and Hogarth (1978) provide a good discussion of features of people's ordinary lives and the way in which they acquire information which promote this belief and lead to its unchallenged persistence.

Snyder and Cantor note that their experiments do not resolve the competing interpretations. Their obtained results may be due to either "differential retrieval" or "differential relevance." The research proposed below is designed to explore the contributions of these two mechanisms in intuitive hypothesis-testing.

CHAPTER IV

THE MECHANISMS INVOLVED IN INTUITIVE HYPOTHESIS-TESTING: DIFFERENTIAL RECALL AND DIFFERENTIAL RELEVANCE

This research was designed to gain further knowledge of the apparent confirmatory error, and the conditions which promote or suppress its occurrence. Snyder and his colleagues have shown, in a succession of related studies, that when they are asked to list what they consider to be relevant in evaluating a proposition or hypothesis, individuals reliably report more confirming than disconfirming evidence. There are also some data which suggest that the tendency to favor confirming evidence while testing a hypothesis translates to a tendency to unjustifiably accept it. But the evidence for this effect of a confirmatory strategy on the final judgment is unconvincing. After making a decision about one hypothesis, subjects were asked to turn around and use the same information to test an alternative hypothesis. Subjects in this experiment (Snyder & Cantor, 1979, Experiment 2) were much more likely to accept the first than the second of the two hypotheses. It is not surprising that after making one considered judgment, individuals do not then immediately argue against it by affirming its alternative to be more true.

There is then ample evidence suggesting that the confirmatory strategy intrudes on intermediate cognitive processes used enroute to a final judgment about an hypothesis, but little evidence that it affects that final judgment itself. There is also little indication why that

intrusion occurs. The suggested mechanisms are that the hypothesis becomes a retrieval cue prompting recall for favorable evidence, but providing no such retrieval prompt for disconfirming evidence. The second proposed explanation suggests that the error comes not primarily during the retrieval of available information, but during the evaluation of that information. According to this view, recalled evidence is evaluated in light of the hypothesis such that confirming evidence is seen as more favorable, disconfirming evidence is viewed as less inconsistent and (at least some) irrelevant information comes to be interpreted as supportive of the hypothesis. The research described here included dependent measures of both the recall and interpretive processes to better untangle the source of the confirmatory error. The experimental design also varied, through different instructions, the orientation participants brought to the hypothesis-testing situation to determine under what conditions the confirmatory error is most likely to occur.

One group of subjects (Uncritical Orientation) participated in a replication of Snyder and Cantor's (1979) study: They were to test one of two opposing hypotheses and, before making their judgments about it, listed all the evidence they considered relevant. Under these instructions, the confirmatory error is most likely to occur. The single hypothesis can serve as a retrieval cue enhancing the recall of hypothesis-consistent information while providing no such organizing schema for disconfirming evidence. Furthermore, and apart from any potential effects of the hypothesis on retrieval, members of this group

were free to evaluate whatever evidence they recalled in light of the hypothesis, interpreting "relevant" as "confirming" evidence.

Two other groups of participants tested hypotheses under conditions designed to lessen the biasing effects of the hypothesis. After the hypothesis was stated, members of these groups were asked to list all the information they could recall, and not just a subset of "relevant" knowledge. Although the hypothesis was still available as a potential organizing schema for selective retrieval of favorable evidence, their instructions to recall all their pertinent information prompted these subjects to retrieve disconfirming evidence that might not be organized under the hypothesis schema, as well as any confirming evidence that is. These recall instructions should diminish one of the possible sources of the confirmatory error--the potential for the hypothesis to exclude from recall much of the disconfirming evidence.

The use of a confirmatory strategy in judgment may best be regarded as another cognitive rule-of-thumb, a short cut method for quick and intuitive hypothesis-testing. Its presence may, therefore, be limited to instances in which individuals give relatively cursory and uncritical attention to the proposition under consideration. (If this is the case, its use would be most expected when the hypothesis deals with subjectively unimportant people or events.) In other judgment situations, when individuals are prompted, either because of the salience of the issue or because of specific instructions to do so, they may be much more deliberate in the way they evaluate the pertinent evidence at hand, and errors in their hypothesis-testing process may then be less frequent

and less pronounced. The present study incorporated these more critical conditions. In addition to being prompted to retrieve all their available knowledge, participants in these two groups (Critical Orientation) were asked to rate each item of recalled evidence for its implications for the hypothesis.

The instruction to be deliberate both in recall and in evaluation of evidence should diminish, but not necessarily eliminate, the tendency to adopt a confirmatory strategy in hypothesis-testing. If that strategy is used, members of the two Critical Orientation groups may, like those in the Uncritical Orientation group, recall more confirming than disconfirming evidence. If there is a tendency to interpret evidence in light of the hypothesis, that tendency will be reflected in the relevance ratings these subjects assigned to the items of evidence they recalled. Since this study includes measures of what individuals retrieve from memory and how they interpreted that evidence, the source of the confirmatory error can be more accurately determined.

The two Critical Orientation groups differed in the number of hypotheses tested and in the prior probability of those hypotheses. Members of one group (Critical Orientation-Single Hypothesis) tested one of two contrasting hypotheses (Jane's suitability for either the Introvert or the Extravert occupation) and that hypothesis carried one of three levels of prior probability. The critical mode of evaluating evidence may be adopted in some cases but not in others. When the decision to accept, reject or estimate the likelihood of the

hypothesis has important personal ramifications or when individuals are, as subjects in this group were, specifically asked to do so, they will be more inclined to be systematic in their hypothesis-testing and will be less likely to rely on a short cut approach. Other features of the hypothesis-testing situation may also promote or suppress the tendency to follow a confirmatory strategy. The prior probability of the hypothesis, independent of its current evidential support, would seem to be a potentially important factor and one which could act on either the retrieval process or on the evaluation of the evidence. Recall that Snyder and Swann (1978, Experiment 3) also suspected that this variable may influence the process of testing a hypothesis. However, they chose to manipulate likelihood by varying base rates. Low prior odds were created by informing some subjects that Jane was a member of a sorority of which only seven of 30 members were extraverts (or introverts). High prior probabilities were created by informing the remainder of the subjects that 23 of the 30 members of Jane's sorority were extraverts (or introverts). In preceding sections it has been shown that base rates such as these are not equivalent to subjective prior probabilities since they merely summarize the frequency distributions of people in categories but impart no explanation for that distribution. The fact that Snyder and Swann found no effect of this manipulation is, therefore, not surprising. In this study a potentially more effective manipulation was employed.

Prior odds were varied by telling participants in this group (Critical Orientation-Single Hypothesis) of the judgments of others who had previously informally tested the same hypothesis. Some learned that others had judged the hypothesis to be true; other members of this group learned that the hypothesis was, in the judgment of others, more likely to be false. For a third group of participants, the hypothesis carried no prior probability. They received no information about the previous judgment of others.

The second of the groups given the Critical Orientation tested both of the alternative hypotheses. When two contrasting propositions are considered simultaneously, evidence which tends to confirm one tends to disconfirm the other. If the hypothesis serves as an organizing schema, the presentation of both alternatives should make available in memory evidence which supports both hypotheses as well as evidence which disconfirms both. The retrieval process should not, therefore, favor one of the alternatives over the other.

Like those who tested a single hypothesis under the critical orientation, these subjects who tested both hypotheses were asked to attempt to recall all the prior information they had learned, and were instructed to be deliberate in evaluating whatever they recalled, assigning relevance ratings to each item of evidence.

To summarize the experimental conditions, one group (Uncritical Orientation) tested either the Introvert or the Extravert hypothesis and, before doing so, listed that evidence they considered relevant to their judgment. No measures of the interpretation of evidence were

taken from members of this group. A second group (Critical Orientation-Single Hypothesis) also tested one of the two hypotheses, and that hypothesis carried either a low prior likelihood, a high prior likelihood, or no prior probability. To diminish any tendency to retrieve a preponderance of confirming evidence, members of this group were asked to attempt to recall all their archival knowledge and to critically evaluate each bit of that evidence before deciding to accept or reject the hypothesis. The third group of participants (Critical Orientation-Both Hypotheses) tested both the Introvert and the Extravert hypotheses. These participants had no information about the prior likelihood of the hypotheses they were considering. Before making their decision about the hypotheses members of this group were also asked to think back and recall all the pertinent knowledge they had learned and to critically evaluate that evidence for its implications for both hypotheses.

With these three groups it is possible to determine more about the source of the confirmatory strategy. If it is due to the effects of the hypothesis on the retrieval of evidence from memory, all participants will recall more confirming than disconfirming evidence. However, if the instructions to attempt to recall all available evidence are effective in making available in memory disconfirming as well as confirming evidence, those who tested their single hypothesis under critical orientation should retrieve proportionately more disconfirming evidence than those who were not prompted with specific recall instructions.

If the confirmatory error operates during the interpretation of recalled evidence, the relevance ratings provided by those in the Critical Orientation groups will reflect that source of bias. If members of these two groups, despite their instructions to be careful in evaluating the evidence and despite the fact that the prior knowledge offers the same amount of confirming and disconfirming evidence, interpret most of what they recall as supportive of the hypothesis, the hypothesis will have been shown to powerfully affect the way in which information is construed.

Method

Participants. One hundred thirty-five male and female undergraduates enrolled in introductory psychology courses participated, in small groups, for course credit.

Procedure. The present study employed materials and procedures based on those used by Snyder and Cantor (1979). It was conducted in two sessions. During the first, participants were provided with an archival store of information. In the second, they used that information to test hypotheses under different conditions.

First Session. To provide the prior information, all participants read an identical account of events in the life of a woman named Jane. Two days later they used that information to test one or two of the complementary hypotheses: that Jane was well-suited for a job that required the personal attributes of the prototypic introvert (Research Librarian) or that she was well-suited for a job

that required the personal attributes of a prototypic extravert (Real Estate Salesperson). The narrative about Jane provided considerable support for either hypothesis: In different situations and at different times Jane was as likely to behave an introverted or extraverted fashion. The descriptors of Jane's behavior were formed using trait terms which, by themselves, were shown by Cantor and Mischell (1977) to be normatively associated with introversion or extraversion. The narrative contained eleven bits of evidence characterizing Jane as an introvert ("Jane was shy and timid at the supermarket") and nine instances in which her behavior was decidedly more extraverted ("Jane remained friendly and outgoing while jogging despite her long day at work").

Errors in judgment may occur because, in light of the hypothesis, confirming evidence is seen as more confirming and disconfirming evidence is seen as less disconfirming. This type of error is testable in the current experimental design by examining the relevance ratings assigned by participants in the two Critical Orientation groups (Single and Both Hypotheses) who were asked to rate each item of evidence for its relevance to the hypothesis. In addition to this tendency to interpret the positive and negative evidence in a manner consistent with the hypothesis, irrelevant evidence may also be interpreted in light of the hypothesis. To explore this potentially biasing effect of the hypothesis, the narrative about Jane was modified to include ten instances in which her behavior was irrelevant to either the introvert or extravert hypothesis. Irrelevant evidence was

created using descriptors which were, in Cantor and Mischell's study, shown to be unrelated to the introversion-extraversion dimension (e.g., "Jane was patient and courteous at work").

After reading the four page narrative (a copy of which appears in Appendix A), all participants completed a brief test of their recall of the narrative. None of the questions in this test concerned Jane's personality or job suitability. The test was included to ensure that all participants, regardless of the group to which they were assigned, had equivalent prior knowledge available when later testing hypotheses about Jane.

Second Session. Two days later all participants were randomly assigned to test one hypothesis, the contrasting hypothesis, or both hypotheses, and to do so under one of two instructions. One group of 30 participants (Uncritical Orientation) took part in essentially a replication of Snyder and Cantor's second experiment. They tested either the Introvert or Extravert hypothesis, but before doing so were asked to list whatever information they could recall about Jane that they considered relevant to their judgment. In this study, after these participants listed the relevant evidence, they were given the opportunity to list any additional information they could retrieve from memory. Finally, they were asked to rate Jane's suitability for the hypothesized job on a six point scale anchored at the extremes by "not at all suited" and "very well suited."

The second group of 90 participants (Critical Orientation-Single Hypothesis) also tested one of the two hypotheses, but did so under instructions to be more critical in their recall and evaluation of the

evidence. Their instructions asked them to first list all the information they could recall from the narrative. After listing that information they were asked to rate each bit of evidence for its implications for the hypothesis. These relevance ratings ranged from -3 to +3 with the negative ratings indicating that the evidence recalled connoted Jane's unsuitedness for the hypothesized job, while the positive ratings connoted her job suitability.

In addition, these participants were further randomly assigned to one of three levels of the prior probability manipulation. In the Low Prior Probability level, participants learned that, in the opinion of Jane's friends, "she might not be good" in the hypothesized job. Those in the High Prior Probability level learned that Jane's friends "thought she might be good" at the job under consideration. A third level (No Prior Probability) was denied information about the opinion of Jane's friends.

The final group of 15 participants (Critical Orientation-Both Hypotheses) tested both the Introvert and Extravert hypotheses. Their task was to judge how well-suited Jane was for the position of real-estate salesperson and how well-suited she would be as a research librarian. Note that these instructions did not ask for a relative judgment of Jane's suitability for one job in comparison to her suitability for the alternative. Instead, they instructed these participants to give equal and independent attention to both hypotheses. The order in which the hypotheses were evaluated was counter-balanced across subjects.

Before their final judgment of job suitability, participants in this group, like those who tested a single hypothesis under the critical orientation, listed all the information they could recall from the narrative and rated each part of that evidence on the six point relevance scale for its implication for the hypotheses. Thus, each item recalled by these subjects carried two relevance ratings: one for its implications for the Introvert hypothesis and one for what it implied about the Extravert hypothesis.

CHAPTER V

RESULTS

Equivalence of the Prior Information

The test of factual information was included at the end of the first experimental session to ensure that all participants had equivalent knowledge potentially available during the second session when they put that knowledge to use. An analysis of the number of errors made on that test showed no differences across the three groups, $F = 1.45$, $p > .24$. Those who tested both hypotheses made, on the average, 4.60 errors. In the two remaining groups, the average number of errors was 3.86 and 3.90. Thus, there is no evidence of systematic differences in archival information in advance of hypothesis-testing.

Manipulation Check: Ambivalence of the Prior Information

The narrative about Jane was designed to be equivalently supportive of both the Introvert and Extravert hypotheses. Analysis of the job suitability judgments made by subjects who tested both hypotheses suggests that the narrative was so perceived. These subjects judged that, on the basis of the prior information they recalled, Jane was as suited for the introverted position of research librarian ($\bar{X} = 3.33$) as she was for the more extraverted occupation of real-estate salesperson ($\bar{X} = 2.87$), $t = .94$, $p > .35$.

Three additional analyses were undertaken to further ensure the ambivalence of the archival information. Two independent judges

counted the number of introvert, extravert, irrelevant and ambivalent items of information recalled by participants in this group, using the list of trait descriptors provided and tested by Cantor and Mischell (1977) as guides. If the narrative were imbalanced and favored one hypothesis over the other, there would be a tendency to recall more of the over-represented evidence. An examination of the number of introvert ($\bar{X} = 1.00$) and extravert items ($\bar{X} = 1.47$) showed that there was no such differential recall by these subjects, $t = 1.10$, $p > .25$.

The foregoing results pertain to the "objective" connotation of the descriptors in the narrative and show that those in the group who tested both hypotheses recalled an equal number of items which, by themselves, are normatively associated with introversion as they did items normatively associated with extraversion. But the subjective interpretation of these items may change when they are embedded within a larger body of information. To examine the way in which the items were interpreted, the number of items assigned positive relevance ratings for the two hypotheses was counted. As expected, this analysis revealed that some items which, by themselves, do not distinguish between introversion and extraversion came to be perceived as supportive of the hypotheses. Whereas according to the ratings of the independent judges, subjects listed, on the average, only about two items which are "objectively" or normatively associated with introversion or extraversion, the participants themselves judged that some seven items were relevant to the introversion/extraversion distinction. However, the number of pro-hypothesis interpretations did not differ

between the two hypotheses. These subjects who tested both hypotheses interpreted some three items ($\bar{X} = 3.13$) as implying Jane's suitability for the introverted job and about three ($\bar{X} = 3.53$) as connoting her suitability for the alternative occupation $t = .51, p > .60$.

Thus, the same number of recalled items were perceived as providing support for the introvert and the extravert hypotheses. A final comparison was made to ensure that the two hypotheses were supported to the same extent. The positive (pro-hypotheses) relevance ratings were summed for each of the two hypotheses and compared. No differences were found. The sum of the pro-Introvert ratings ($\bar{X} = 10.93$) was the same as the sum of the pro-Extravert ratings ($\bar{X} = 8.20$), $f(1,14) = 1.09, p > .40$.

All of these results support the assumption that the prior information about Jane was balanced and could confirm either hypothesis; participants who tested them both recalled the same number of items which are normatively associated with introversion and extraversion. Furthermore, apart from the objective connotations of the evidence, the participants themselves interpreted an equal number of items to be supportive of the two hypotheses. Finally, those items which implied Jane's suitability for one occupation did so to the same extent as the items which implied her suitability for the alternative occupation. In short, these findings suggest that evidence favoring both hypotheses was equally available and both hypotheses were equally supported by that evidence. The means associated with these findings are shown in Table 1.

TABLE 1

AMBIVALENCE OF THE PRIOR INFORMATION

| | <u>HYPOTHESIS</u> | |
|--|-------------------|------------------|
| | <u>INTROVERT</u> | <u>EXTRAVERT</u> |
| Job suitability judgment | 3.33 | 2.87 |
| Number of recalled items "objectively" supporting the hypothesis | 1.00 | 1.47 |
| Number of recalled items subjectively interpreted as supporting the hypothesis | 3.13 | 3.53 |
| Sum of pro-hypothesis relevance ratings | 10.93 | 8.20 |

All judgments were made by subjects who tested both hypotheses.
None of the differences between the two hypotheses is significant.

It may be argued that although the hypotheses were presented simultaneously, the decision to accept or reject them was necessarily made in sequence and that the confirmatory error could have operated such that the first hypothesis was favored over the second (as happened under somewhat different procedures in Snyder and Cantor's (1979) second experiment). In the present study, the order of the judgments was counter-balanced. This would ensure that, across all subjects in this group, no systematic differences between the two hypotheses would emerge. But it would also mask any order effects of the two judgments within individual participants. To make certain that these participants did not follow a biased procedure which led them to favor the first over the second hypothesis, the sequence in which the two judgments were made was analyzed. No order effects were obtained. The first job suitability assessment was equivalent to the second ($\bar{X} = 3.2$ and 3.0), indicating that these subjects were quite able to independently evaluate the two hypotheses, that they were not more likely to accept the first at the expense of the second, and that their judgment of Jane's suitability for one job did not necessarily imply her unsuitability for the alternative.

No Effect of the Prior Probability Manipulation

One interest of this research was the potential effect of the hypothesis' prior probability on the processes employed to test it. Here, prior likelihoods were varied by telling subjects in one group (Critical Orientation-Single Hypothesis) of the opinions of Jane's friends with respect to her suitability for the hypothesized occupation. Some participants (High Prior Probability) were told that Jane's friends thought she might be suited, others (Low Prior Probability) learned that Jane's friends did not think she would be suited. Still others (No Prior Probability) were denied any information of this kind. They made their decision about the hypothesis in the absence of any indication of its prior likelihood.

Apparently, this manipulation of prior odds was insufficient to affect the hypothesis-testing process. It had no impact on job-suitability judgments, on the type of evidence retrieved from memory, or on the interpretation of that evidence.

Because of the lack of any effects of this manipulation, the following analyses were pooled over the three levels of prior probability.

Judgments of Job Suitability

Analysis of the job suitability judgments made by those who tested one hypothesis revealed an unexpected main effect of the Introvert over the Extravert hypothesis. In both the Critical and Uncritical

Orientation groups, subjects who evaluated Jane's suitability for the Introverted job judged that she was more suitable than did those who tested her qualifications for the Extravert job.

However, this finding alone does not suggest that these subjects followed a flawed strategy of hypothesis-testing. The presence of a confirmatory error would only be indicated if they were led to a conclusion different from that warranted by the normative or objective strength of the evidence. The best measure of the "correct" job-suitability judgment comes from those who tested both hypotheses and who found both to be equally supported. If subjects who tested one hypothesis differed in their judgments from those who had the benefit of both hypotheses, a confirmatory error would be indicated.

Analyses of variance comparing the job suitability judgments of those who tested both hypotheses against the same judgment made by participants in the other two groups who tested either the Introvert or Extravert hypothesis revealed no main effects nor interactions. That is, participants who tested only one of the hypotheses were not influenced to more strongly accept it than were those who tested both hypotheses. The means are shown in Table 2.

TABLE 2

JOB SUITABILITY JUDGMENTS

| | <u>HYPOTHESIS</u> | |
|--|-------------------|------------------|
| | <u>INTROVERT</u> | <u>EXTRAVERT</u> |
| Critical Orientation-Both Hypotheses | 3.33 | 2.87 |
| Critical Orientation-Single Hypothesis | 3.44 | 2.49 |
| Uncritical Orientation | 3.93 | 2.47 |

Effect of the Hypothesis on Information Recalled

The results of the previous analysis of job suitability judgments suggest that participants in this study did not fall prey to any confirmatory error, at least insofar as that error is evidenced by a tendency to unjustifiably accept a stated hypothesis. The next series of analyses examined the effect of the hypothesis on the kind of information they recalled enroute to their final decision about that hypothesis. For each of the participants who tested one of the two hypotheses (i.e., those in the Critical Orientation-Single hypothesis group and those in the Uncritical Orientation group), the proportions of objectively confirming and disconfirming evidence recalled were counted by dividing the number of pro-hypothesis items by the total number of items recalled. A 2 (Introvert-Extravert hypothesis) x 2 (Critical-Uncritical Orientation) x 2 (Confirming-Disconfirming Evidence) analysis of variance, with type of evidence as a repeated measures factor, revealed a marginal interaction between the hypothesis and the type of evidence reported. Those, under either orientation, who tested the Introvert hypothesis recalled more confirming ($\bar{X} = 25\%$) than disconfirming ($\bar{X} = 18.75\%$) evidence. However, this tendency was not replicated for those who tested the Extravert hypothesis. This interaction between the direction of the hypothesis and the type of evidence recalled about it was marginally significant, $F(1,116) = 2.71$, $p < .10$. No significant differences distinguished between the two Critical and Uncritical Orientation groups. Regardless of their instructions, all subjects recalled essentially the same proportions

of confirming and disconfirming evidence. Table 3 summarizes the types and proportions of evidence recalled by members of both groups.

The results of this analysis suggest that any confirmatory error is not principally due to the biasing effect of the hypothesis on retrieval processes. There did seem to be some selective recall favoring the Introvert hypothesis, but none favoring the Extravert hypothesis (and, in fact, the opposite trend was more apparent).

Effect of the Hypothesis on the Interpretation of Evidence

An alternative explanation of the mechanism underlying the confirmatory error proposes that the hypothesis under consideration colors the interpretation of whatever evidence is available: more evidence is viewed as confirming and less is seen as disconfirming or irrelevant. In order to determine whether, and to what extent, this mechanism is responsible for the apparent confirmatory error, participants in two of the groups were asked to provide relevance ratings of the items of evidence they recalled. Those who tested either a single or both hypotheses under Critical instructions rated each item recalled for its implications for the hypothesis. Positive values were assigned to items that, in the opinion of the subject, implied Jane's suitability for the hypothesized job. If they judged the item to indicate Jane's unsuitability, a negative value was assigned. Values of zero were assigned to items recalled that were subjectively void of implications for the hypothesis.

TABLE 3

PROPORTIONS OF CONFIRMING AND DISCONFIRMING EVIDENCE RECALLED

| | <u>Critical Orientation- Single Hypothesis</u> | | <u>Uncritical Orientation</u> | |
|------------------------|--|------------------|-----------------------------------|------------------|
| | <u>INTROVERT</u> | <u>EXTRAVERT</u> | <u>INTROVERT</u> | <u>EXTRAVERT</u> |
| Confirming evidence | 24% | 13% | 28% | 22% |
| Disconfirming evidence | 18% | 30% | 21% | 18% |

If the interpretation of the evidence is veridical, the subjective connotation of recalled items will be equivalent to their "objective" or normative implications for the hypotheses. In other words, of the items of evidence retrieved from memory, those that are normatively confirming, disconfirming or irrelevant will be perceived that way despite the hypothesis. The first analysis examined the proportions of correct interpretations made by these subjects. If no interpretive errors were made, all (100%) of the evidence recalled would have been interpreted correctly. This was clearly not the case. Across all participants in both groups, less than half (46.5%) the items were correctly interpreted.

An analysis of variance comparing the two groups and the two hypotheses revealed that more interpretive errors were made in evaluating the Introvert hypothesis than in testing the Extravert hypothesis, $F(1,41)=6.68$, $p < .05$. When testing the former, only 37.6% of the items were correctly interpreted, whereas 55.6% of the items were correctly interpreted when the Extravert hypothesis was being considered. The interaction between the number of hypotheses tested (one or both) and the direction of the hypothesis was marginally significant, $F(1,61) = 3.61$, $p < .10$. Subjects who tested both hypotheses were about as correct (or as incorrect) in evaluating an item's relevance for the Introvert hypothesis (41% correct) as they were in assessing an item's implications for the Extravert hypothesis (48% correct). In contrast, those who tested only the Extravert hypothesis were correct in their interpretation of the evidence more

often (i.e., on more items, 58%) than were those who tested only the Introvert hypothesis (36% correct).

The fact that these subjects did not often agree with the normative interpretation of the evidence they recalled does not in itself demonstrate that their appraisal of the information was biased by the hypothesis. The next analyses examined the kind of errors made to uncover the influence of the hypothesis. If the confirmatory error lies in a tendency to interpret evidence in its most favorable light, i.e., as offering support for the hypothesis, three kinds of errors will account for most of the discrepancy between the objective and the subjective interpretations of the evidence: the tendency to view normatively disconfirming evidence as irrelevant, or even as confirming, and the tendency to interpret ambiguous or irrelevant information as supportive.

The proportions of items misinterpreted in each of these three ways were calculated for both groups and are shown in Table 4. The most striking feature of these data is the way disconfirming evidence was evaluated. Those who tested the Introvert hypothesis (but not those testing the Extravert hypothesis) interpreted many objectively extraverted items as irrelevant to the hypothesis. The main effect of the direction of the hypothesis was highly significant, $F = 8.65$, $p < .01$. Even more striking is the proportion of disconfirming items which came to be viewed as supportive of the hypothesis. Averaged across both groups, those testing the Introvert hypothesis judged some 65.1% of the extravert items to be consistent with the Introvert hypothesis.

TABLE 4
SUBJECTIVE INTERPRETATIONS OF RECALLED EVIDENCE

| | CRITICAL ORIENTATION- BOTH HYPOTHESES | | CRITICAL ORIENTATION- SINGLE HYPOTHESIS | |
|---|--|------------------|--|------------------|
| | <u>INTROVERT</u> | <u>EXTRAVERT</u> | <u>INTROVERT</u> | <u>EXTRAVERT</u> |
| <u>CORRECT INTERPRETATIONS</u> | | | | |
| Confirming evidence interpreted as confirming | 80% | 78% | 49% | 78% |
| Disconfirming evidence interpreted as disconfirming | 9% | 93% | 12% | 85% |
| Irrelevant evidence interpreted as irrelevant | 43% | 32% | 39% | 44% |
| Total correct interpretations ^a | 41% | 48% | 37% | 58% |
| <u>CONFIRMATORY MISINTERPRETATIONS</u> | | | | |
| Irrelevant evidence interpreted as confirming | 55% | 65% | 52% | 51% |
| Disconfirming evidence interpreted as irrelevant | 17% | 0% | 26% | 10% |
| Disconfirming evidence interpreted as confirming | 74% | 7% | 62% | 4% |
| Total confirmatory misinterpretations ^a | 55% | 45% | 46% | 36% |

^aExpressed as a percentage of the total number of items recalled.

The salience of the misinterpretation of disconfirming evidence may obscure the fact that this type of error accounts for relatively little of the total amount of all confirmatory errors. While it is true that over 60% of all extravert items recalled by those testing the Introvert hypothesis were misjudged as supportive, most subjects recalled few such disconfirming items (refer to Table 3). Although misclassification of these items occurred frequently, this type of error contributed to less than 25% of all confirmatory errors. Table 5 shows the relative contribution of each type of error to the total amount of confirmatory errors. It can be seen that the largest single source of error comes from the tendency to interpret irrelevant or ambiguous evidence as favorable to the hypothesis. The data displayed in Table 5 also show more clearly the pattern of misinterpretations by those testing the two hypotheses. Subjects evaluating Jane's suitability for the introvert occupation were more likely to view negative evidence as supportive while those testing the extravert hypothesis interpreted more ambiguous information as confirmatory.

Taken together, and across both groups and both hypotheses, some 43% of all items recalled were misinterpreted in favor of the hypothesis, while only 10% were otherwise misjudged. Cast another way, the three types of confirmatory errors account for fully 81.8% of all the discrepancies between the objective and subjective interpretations of the evidence available in memory.

These results demonstrate the clear influence of the hypothesis on the interpretation of evidence. The confirmatory error may also arise

TABLE 5
CONTRIBUTION OF EACH TYPE OF CONFIRMATORY ERROR
TO THE TOTAL AMOUNT OF CONFIRMATORY ERRORS

| | CRITICAL ORIENTATION- BOTH HYPOTHESES | | CRITICAL ORIENTATION- SINGLE HYPOTHESES | |
|---|--|------------------|--|------------------|
| | <u>INTROVERT</u> | <u>EXTRAVERT</u> | <u>INTROVERT</u> | <u>EXTRAVERT</u> |
| Irrelevant evidence interpreted as confirming | 68% | 98% | 65% | 90% |
| Disconfirming evidence interpreted as confirming | 26% | 2% | 25% | 3% |
| Disconfirming evidence interpreted as irrelevant | 6% | 0% | 10% | 7% |

from more subtle effects of the hypothesis on the assessment of evidence by altering the subjective weight of items of evidence. That is, evidence which offers only weak support may come to be viewed as substantially confirming and evidence which argues strongly against the hypothesis may be perceived as less damaging. This study cannot determine the extent to which the diagnostic value of evidence was affected by the hypothesis since there is no way to calculate the true and unbiased weight of each item of evidence. However, since members of the two Critical Orientation groups provided relevance ratings for each item they recalled, it is possible to check for differences across the two groups. This comparison was made by computing, for each member of these two groups, the average value assigned to pro-hypothesis items (those judged to indicate Jane's job suitability). The same average was computed for anti-hypothesis items. There were no differences distinguishing subjects who tested both hypotheses from those who tested a single hypothesis on either measure. In the Both Hypotheses group, the average weight assigned to pro-hypothesis items was 2.17, and the average value of anti-hypothesis items was -1.11 (on the -3 to +3 scale). Those who tested one of the two hypotheses assigned an average value of 2.15 to items they judged as supportive of the hypothesis and an average value of -1.30 to items they considered as disconfirming of the hypothesis.⁸

C H A P T E R V I

DISCUSSION

The goal of this investigation was to gain a better understanding of the way in which people answer questions about others. More specifically, this study sought to determine if, and the extent to which, the question itself may predispose its answer. By measuring both the information retrieved from memory and the assessment of that information, an attempt was made to work back from the final judgment to trace the cognitive processes involved when individuals use information from their own memory to test hypotheses about other people.

The explanation of intuitive hypothesis-testing derived from these data is clouded by the unexpected differences between the Introvert and Extravert hypotheses. In principle, and by design, there should have been no differences between those who tested the Introvert hypothesis and those who tested its alternative. Instead, significant discrepancies occurred at each stage of the process--in the type of information recalled, in the assessment of that evidence, and in the final judgment of job suitability. The narrative contained equivalent amounts of pro-introvert and pro-extravert evidence. From an objective standpoint, the story thus supported the two hypotheses equally. And, according to subjects who had both hypotheses before them, the narrative was balanced: they recalled equivalent amounts of introvert and extravert items, they rated the pro-introvert and the pro-extravert evidence as equivalently supportive of the two hypotheses, and, on the basis of the narrative, they judged Jane to be equally suitable for both of the

hypothesized occupations. Other results from the Critical Orientation-Single Hypothesis group suggest, however, that the initial appearance of the narrative's ambivalence may have been misleading. Subjects in this group, the largest in the study, recalled more introverted than extraverted items regardless of the hypothesis they were testing. Despite the fact that the introvert items were not more numerous, they may have been more available in memory or more salient after the two-day interval between their acquisition and their subsequent use in hypothesis-testing.

While this makes the interpretation of the findings less straightforward, it does not render the data uninformative of the cognitive processes employed enroute to a judgment about a hypothesis. Differences in processing due to the two different orientations can still be detected, and other generalizations from the data can be made if the two hypotheses are considered separately.

If it can then be best assumed that the recall data of those in the Critical Orientation-Single Hypothesis group most accurately characterize the distribution of introvert and extravert evidence actually available to all subjects, it appears that the Uncritical Orientation promotes the confirmatory error. Subjects given that orientation (a replication of the instructions used by Snyder and Cantor) reported as relevant more confirming than disconfirming evidence. If they were evaluating the introvert hypothesis' they listed as relevant more evidence confirming that hypothesis. On the other hand, if they were evaluating Jane's suitability for the extravert occupation,

they reported more evidence consistent with that hypothesis, even though more of its disconfirming data were actually available. These findings replicate the pattern found by Snyder and Cantor (1979) using the same instructions to subjects. If anything, the data from the present study show even more clearly the effect of the hypotheses. Even though the narrative apparently made more introvert than extravert evidence available, subjects in the Uncritical Orientation group testing the extravert hypothesis listed as their evidence more extravert than introvert items.

The difference between the Critical and Uncritical Orientations shows the effect of different instructions or modes of gathering evidence in hypothesis-testing. When individuals, having been given a hypothesis, are asked only to review what they consider relevant to that hypothesis, there is a marked tendency to focus upon more evidence that confirms the hypothesis and less that contradicts it. On the other hand, when individuals are asked to make a more deliberate and thorough review of all the evidence before making any decision about the hypothesis, the confirmatory tendency is suppressed. They are not so influenced by the hypothesis that they recall more of its confirming than disconfirming data.

The results obtained from the Uncritical Orientation group in this study, as well as those found in Snyder and Cantor's experiment, are consistent with findings obtained in related research. Wason (1960, 1968, 1969) and Wason and Johnson-Laird (1972) have shown that people tend to test propositions by searching for instances, exemplars, or

other evidence which would affirm their truth (i.e., "positive hits"), while ignoring (or at least making much less effort to assemble) evidence that would disprove them. Recall, too, that Tversky and Kahneman's (1974) representativeness heuristic is based on the same notion: that intuitive questions are answered by counting the amount of supportive data. This tendency, replicated as it has been in various judgmental tasks, seems to reflect a fundamental feature of intuitive reasoning.

The question which prompted this present investigation was why this tendency occurs. Does the hypothesis itself influence the organization of information held in memory such that the confirming evidence is made easily accessible while disconfirming data is masked or made less retrievable? The answer on the basis of the data gathered here is a qualified no. If the hypothesis operated as a selective schematizing mechanism organizing only, or predominately, its confirming instances, subjects in all groups, and especially those in the Uncritical Orientation group, would have been expected to have reported proportionately less disconfirming evidence. The results show that all participants were quite able to recall substantial amounts of disconfirming evidence.

It therefore appears that the confirmatory error is not principally due to the unavailability of disconfirming data.⁹ Instead, it is traceable to the way in which the question is construed and, more accurately, to the standard used as the basis of comparison. The data support the view that subjects interpret Jane's job suitability to

mean that she possesses the traits and behaviors of the prototypic salesperson or librarian. The prototype, however, is defined principally by traits and behaviors known or assumed to be possessed by the majority of members of those occupations. That is, the prototype is defined mainly by confirming traits or behaviors. Thus, when subjects in the Uncritical Orientation group were asked to list what they considered relevant, they reported the traits and behaviors Jane shared with the prototype. Since the prototype contained mostly confirming exemplars, subjects reported more of Jane's confirming characteristics. This confirmatory error therefore seems to have as its source the tendency to use a comparative standard which is incomplete and one which does not sufficiently allow the falsification of the hypothesis.

The review article by Einhorn and Hogarth (1978) is pertinent here. Although they focus on the way in which antecedents and consequents are learned, their discussion can be as well applied to the way in which prototypes are developed and maintained. Einhorn and Hogarth suggest that confirming evidence is over-represented in memory and is used as the test of intuitive hypotheses simply because in most situations disconfirming evidence rarely ever becomes available. Consider, for example, how individuals in this study learn about "suitable" real estate salespersons. Their conception comes from successful salespeople they have known. And those already in the profession are, for the most part, suitable (if not, they would be asked to seek employment in other fields). The criterion against which Jane is judged is therefore heavily weighted in terms of positive

attributes.

While the development of a confirmatory prototype may be a natural and unavoidable consequence of the way real life correlations are learned, the tendency to use that prototype as the principal standard of comparison can be avoided. When, as those given the Critical Orientation were, individuals are first asked to deliberately recall all that they know about the person involved, and are then asked to examine each bit of that evidence for its implications for the hypothesis, the confirmatory error can be suppressed. These instructions prompt individuals to be less reliant on a comparison-to-prototype judgmental strategy. Instead of evaluating Jane against some cognitive representation of confirming characteristics, subjects in the Critical Orientation groups were able to attend to both the confirming and disconfirming attributes they had recalled from the narrative.

Additional findings from the Critical Orientation groups have also shown how the hypothesis affects the interpretation of evidence held in memory. All subjects in these two groups interpreted most of the evidence recalled to be supportive of the hypothesis. Even those who tested the extravert hypothesis and who recalled more introvert than extravert items interpreted much more of those items as confirming of the hypothesis. The largest source of the difference between the amount of evidence that objectively supported the hypothesis and the amount subjects themselves viewed as confirmatory came from the tendency to view ambiguous or irrelevant items as favorable. Over half of the items which normatively do not distinguish between introversion

and extraversion came to be seen as consistent with the hypothesis. In hindsight, it appears that the magnitude of this kind of misinterpretation of irrelevant evidence is overstated. Part of this finding is due to the way in which items were defined. The classification of evidence as introverted, extraverted, or irrelevant was taken from a previous study by Cantor and Mischell (1977) and was also used by Snyder and Cantor. The classification system treats as equivalently irrelevant items which do not at all refer to Jane's personality ("Jane drove a small car") and items which may refer to her traits but which are not identified with introversion or extraversion ("Jane was punctual"). While items in the latter category may not distinguish between introverts and extraverts, they can provide useful information about Jane's potential job suitability. Knowing that Jane was punctual, for example, may not aid in determining whether she is an introvert or an extravert, but it can be helpful in assessing her suitability for a job. Thus, while some of the items recalled by subjects were irrelevant to the introversion-extraversion dimension, they were pertinent to a judgment about her job suitability. In this light, it is not surprising that so many "irrelevant" items were interpreted as relevant by subjects in this study.

Not all of the irrelevant items were useful in testing the hypothesis. Many were, from any point of view, quite void of implications for Jane's occupational choice. An examination of items of this type showed that their interpretation was influenced by the hypothesis. To take one example, many subjects remembered that Jane

drove a small car (Datsun) to work. A subject testing the introvert hypothesis interpreted this as evidence that she was unflashy and reserved, typical introverted characteristics. Another subject considering Jane's suitability as a real estate salesperson commented that her possession of an economical automobile would be a benefit in that occupation since extensive driving could be expected. As another example, subjects often recalled that Jane intended to learn about the taxation issues involved in the coming town election. A subject evaluating Jane's qualifications as a librarian mentioned that this item of evidence indicated that Jane enjoyed and would be good at doing research. A counterpart testing Jane's suitability for the alternative job interpreted this as a positive sign of Jane's interest in land values and taxation rates.

The finding that ambiguous or ambivalent evidence is interpreted in light of an hypothesis is consistent with results found in other investigations. Ajzen, Dalto and Blyth (1979) found that irrelevant information in the form of an ambiguous personality description came to be viewed as consistent with a pre-existing impression. The results obtained here extend that finding to instances in which no prior impression was formed. Here, subjects were using information they recalled to gain an impression and they tended to interpret the ambiguous information as supportive of the hypothesis under consideration.

The results of this study, summarized briefly, suggest that confirmatory errors were made by all subjects, but perhaps at different

stages in the hypothesis-testing process. All those in the Uncritical Orientation group reported as relevant more confirming than disconfirming evidence and subjects in the Critical Orientation groups tended to interpret most of the evidence they recalled as supportive of the hypothesis, even though, from an objective standpoint, much of that evidence was uninformative about the hypothesis. A comparison of the results between the two orientation groups further suggests that one source of the confirmatory error--the tendency to assemble as evidence mostly confirming data--can be diminished by instructions which prompt individuals to recall all that they know about the hypothesized person and to carefully examine that evidence for its implications for the hypothesis.

This demonstration of unintentional flaws in intuitive judgmental processes is in keeping with virtually all of the contemporary research on the way in which people come to understand others and the world about them. With the possible exception of Ajzen (1977), all of the published investigations of human reasoning reviewed in Chapters One and Two or cited elsewhere in this paper have shown individuals to be in error in the way they make inferences or judgments. The impression created by this body of literature is that the intuitive scientist is all shortcomings and no strengths. It is hard to imagine that any individual who commits the kind of flaws noted in the literature as frequently as he or she is said to commit them would go through a single day without causing serious harm to him/herself or others. It is even harder to imagine that a species comprised of such error-

prone individuals would survive. Since the species and most individuals seem to be doing at least passably well, the impression created by the literature must be misleading. The research, including the present study, has either succeeded in illuminating only trivial judgmental errors, or its methods and procedures have promoted flawed judgments to a degree that is atypical of ordinary circumstances. This study may be guilty of both. The methodological artifacts which overstate the magnitude of the confirmatory error have been described above. But the importance of the confirmatory error itself also deserves closer scrutiny. In this study, as in Snyder and Cantor's original experiment, subjects made confirmatory errors in gathering evidence about Jane. But neither in this study nor in Snyder and Cantor's were they misled by those errors to enthusiastically endorse Jane's job suitability. In this study the average estimate of Jane's suitability for the real estate job was just above the mid-point ($\bar{X} = 3.52$) and the subjects' assessment of her qualifications for the introverted occupation was at the mid-point ($\bar{X} = 2.56$). Similarly, Snyder and Cantor did not find their subjects to be very impressed with Jane's job potential even though they used as evidence more confirming than disconfirming data. On a one-to-six scale, the average job suitability judgments were about mid-way between ill-suited and well-suited ($\bar{X} = 3.86$ for the introvert job and 3.88 for its alternative). The paradox that emerges then is that processes antecedent to a final judgment are shown to be flawed, but the final judgment itself is in line with the normatively correct expectations. Perhaps a new line of research should be devoted to resolving this kind of paradox to illuminate how people succeed so

well in their intuitive reasoning while at the same time making what appear to be pervasive errors in that reasoning process.

Footnotes

¹In most of the studies of subjective judgment, subjects are asked to estimate the likelihood or probability of any outcome or multiple outcomes. Probabilistic dependent measures are used because they allow the comparison of intuitive inferences against those prescribed by probability or statistical theory. However, adherence to the dictates of statistical principles has subtly become the standard of "rationality" in decision making. It is not the best measure of the soundness of all judgments.

In statistics courses students estimate the probability of drawing a red marble from an urn. When they come to social psychology experiments they may be asked to estimate the likelihood that any individual, or the individual described in a brief character sketch, is an engineer or a lawyer. If they ever serve on a jury they are asked to decide if the evidence points, beyond a reasonable doubt, to the defendant's guilt. In all these instances, the judgment is one of probability. However, these examples seem to imply different meanings of the term.

In one sense, and the one implied in judgments of the likelihood of drawing a marble of a certain color from an urn, or of the likelihood that any unspecified individual is an engineer, probability refers to statistical regularity. In this sense, the probability of an event is an expression of the relative frequency with which similar events occur in the long run (cf. von Mises, 1951). Probability here refers to a collective of events, not to a particular member of that collective. To say that the probability of heads on the next toss is

one half is to express some expectation about the set of coin tosses in an unlimited sequence of tosses, each of which is independent. It says nothing about one particular coin toss. Statistical probabilities rest upon certain assumptions (independence and randomness) and are calculated and combined according to specified arithmetic manipulations.

There is, however, another meaning of probability, one which refers to individual events rather than to long run sequences of similar events. It can be argued, furthermore, that most judgmental tasks, whether posed in an experiment or suggested by everyday experience, involve this alternative connotation of probability. When subjects are asked whether Jack is a lawyer or an engineer they are not asked about a collective of people like Jack, they are asked to make a decision about one particular person.

The counter-argument is that this other meaning of probability is not in any fundamental way a different kind of judgment. This position would hold that probabilities of specific events can and should be regarded within the statistical framework. This contention is implicit in the research on intuitive reasoning. By comparing the probability of single events to that prescribed by statistical models, researchers imply that idiographic probabilities should conform to those describing nomothetic expectations.

This argument seems to have little support outside the social sciences. Von Mises (1951), for one, admits that the probability of single particular events is not a subject for the probability calculus. Speaking for the community of classic probability

statisticians, he writes, "The probability of death, when it refers to a single person, has no meaning for us" (p. 11). Various other authors seem to agree and incommensurate connotation of "probability." For Kneale (1949), probability meant justifiability. In his interpretation, to say that a conclusion or hypothesis was probable was to say that the evidence justifies ("probabilifies") the conclusion. Popper (1959) also takes probability to mean evidential support. Similarly, Carnap (1950) discusses probability in terms of degrees of confirmation of conclusions by information. Cohen (1977) describes it as a gradation of inferential support. Despite differences in synonyms, probability is, for all these writers, clearly relative to case-specific judgments and clearly different from expectations about collectives of similar events. Rather than expressing an expectation about relative frequencies in the long run, probability refers to an expectation about a particular event at a particular time. Historical regularities in sequences of events may be included as part of the evidence upon which a probabilistic (in this alternative sense of the word) estimate is made, but they are not the only or the best sources of evidence. For some, even intuition was regarded as valid evidential support (Carnap, 1950; Keynes, 1921).

There seems to be substantial agreement that the classic or mathematicist interpretation of probability is insufficient to account for the meaning implied by its usage in many cases. There is, however, far less agreement about the defining features of the alternative interpretation. Kneale (1949) argued that his concept

of probabilification was not even a quantifiable concept and could not in any meaningful way be incorporated into any formal calculus. It could be ordered in terms of the degree to which the evidence supported the hypothesis, such that it is possible to distinguish instances in which the conclusion is strongly supported from those in which it is barely probabilified. But there are also instances in which it is difficult to determine whether or not, or to what degree, a hypothesis is supported. For Kneale, probability was a logical, not a metrical, concept. Keynes (1921) also embraced the notion of probability as a logical relation between evidence and a conclusion, but argued that probabilistic relations could be represented in a formal mathematical theory. However, in the theory he proposed, probabilities of different events were not comparable in magnitude. So it was not possible to say that the probability of one hypothesis based on some evidence was greater or less than the probability of another hypothesis based on different evidence. Popper's probability-as-corroboration does not obey the formal calculus of classic probability. Carnap takes the term to mean the measure of one's rational degree of belief in a conclusion or hypothesis. That degree of belief could be calculated as a precise value between zero and one. If held by rational people, degrees of belief were said to obey certain function-rules ("credence functions") describing the relationship between evidence and the hypothesis. While Carnap's is then a quantitative theory, it is not the same as the statistical theory of probability (if it were, there would have been little need for him to spend his life making up a new one). Carnap has,

nonetheless, failed to convince his critics that the logical concept of probability can, like the classic concept, be mapped onto a formal calculus and calculated as a precise value. Strawson (1952) argued explicitly against a formal and quantitative expression of logical probability: "We can never describe the strength of evidence more exactly than by the use of such words as 'slender', 'good', 'conclusive'. . ." (p. 247). Similarly, Cohen's "inductive probability" is "rough, indeterminate, and a matter for judgment" (p. 40). Instead of being precisely quantifiable, it allows only comparative or ordinal gradations.

The point of this note is not to offer a resolution to the long-standing difference of opinion about the quantifiability of logical probability. The purpose is to draw attention to the subjective interpretation of probabilistic dependent measures. This discussion should suggest that probability, when it refers to picking a marble out of an urn, may not mean the same thing, nor should it be calculated in the same manner, as the probability that Jack is an engineer. In many cases, an individual's estimate of probability represents his or her belief in the (psycho)logical inferability of the hypothesis, given the evidence. It does not refer to their computations of empirical frequencies according to a formal calculus.

²D is the cause of C if it is not a spurious cause, i.e., if no event antecedent to D accounts for the conditional probability of the effect just as completely. D is a spurious cause of C if there exists a prior event A such that:

$$p \{ C / D \ \& \ A \} = p \{ C / A \}$$

³See Chapter 8 of Wason and Johnson-Laird for a discussion of some differences between logical material implication and psychological implication.

⁴The distinction here between errors and biases is somewhat different from that of Kruglanski and Ajzen. Here, errors are cognitive in origin and biases are motivational.

⁵I would like to thank Dr. Mark Snyder for making his stimulus materials available to me.

⁶Snyder and Cantor (1979) had the independent judges divide the recalled evidence into only two categories--introvert and extravert. Since the narrative used in this study was modified to include neutral or irrelevant descriptors, the independent judges included the irrelevant category as one of the classifications. Ambivalent items are those listed by subjects which make reference to both introvert and extravert characteristics (e.g., "Jane was outgoing when jogging but shy at the office").

⁷Analyses involving the subjects who tested both hypotheses were conducted by following the procedure outlined by Erlebacher (1977). In that paper Erlebacher describes a technique for deriving the appropriate sums of squares, mean squares, degrees of freedom, and quasi-F ratios for an analysis of variance that includes the design type as a factor in the analysis. Design type refers to the distinction between a Within-subject and a Between-subject manipulation. Erlebacher's procedure is well-suited for this design in which the subjects who tested one of the two alternative hypotheses (Between-subject manipulation) were measured against those who tested both of the hypotheses (Within-subject manipulation). I would like to thank Dr. Jerry Myers for recommending Erlebacher's article and for patiently explaining the generalization of that procedure to multi-factor designs.

⁸The effect of a hypothesis on the type of information recalled from memory and on the interpretation of that information may be strongest on the initial items of evidence recalled. This assumption is based on the view that the first items may be more spontaneously retrieved while latter items may be the product of a more deliberate memory search and their retrieval may therefore be less dependent upon the influence of the hypothesis. To examine this possibility, separate analyses of each of the dependent measures described here were conducted comparing the first five items recalled against all other items. No differences were found on any measure.

⁹This is not to deny that cognitive sets induced prior to the acquisition of knowledge do not operate as organizing schemas for subsequent information stored in memory. A considerable amount of recent research has shown that pre-existing trait dimensions (Cantor & Mischell, 1977), self concepts (Markus, 1977), and attitudes (Judd & Kulik, 1980) can direct the way in which relevant knowledge is organized in memory. The "induced set" research essentially deals with the influence of prior expectations or wishes on processes used to encode later information. In contrast, the present study deals with the effect of a hypothesis on previously encoded information.

REFERENCES

- Abelson, R. P. Script processing in attitude formation and decision making. In J. Carroll and J. Payne (Eds.), Cognition and Social Behavior. Hillsdale, NJ: Erlbaum, 1976, 33-45.
- Ajzen, I. Attributions of dispositions to an actor: Effects of perceived decision freedom and behavioral utilities. Journal of Personality and Social Psychology, 1971, 18, 144-156.
- Ajzen, I. Intuitive theories of events and the effects of base rate information on prediction. Journal of Personality and Social Psychology, 1977, 35, 303-314.
- Ajzen, I., Dalto, C., & Blyth, D. Consistency and bias in the attribution of attitudes. Journal of Personality and Social Psychology, 1979, 37, 1871-1876.
- Anderson, C., & Ross, L. The survival of theories in the absence of evidence. Paper presented at the Annual Convention of the Western Psychological Association, San Francisco, April, 1978.
- Anderson, J. R. Language, memory and thought. Hillsdale, NJ: Erlbaum, 1976.
- Bar-Hillel, M. On the subjective probability of compound events. Organizational Behavior and Human Performance, 1973, 9, 396-406.
- Bar-Hillel, M. The base rate fallacy in subjective judgments of probability. Unpublished doctoral dissertation, Hebrew University, 1975.
- Barresi, J., Robbins, D., & Shain, K. Role of distinctive features in the abstraction of related concepts. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 360-368.
- Beckman, L. Effects of students' performance on teacher's and observer's attributions of causality. Journal of Educational Psychology, 1970, 61, 75-82.
- Berman, J. S. & Kenny, D. A. Correlational bias in observer ratings. Journal of Personality and Social Psychology, 1976, 34, 263-273.
- Brooks, L. Nonanalytic concept formation and memory for instances. In E. Rosch and B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum, 1978.
- Brown, R. V., Kahr, A. S., & Peterson, C. Decision analysis for the manager. New York: Holt, Rinehart, & Winston, 1974.

- Bruner, J. S., & Tagiuri, R. The perception of people. In G. Lindzey (Ed.), Handbook of social psychology. Cambridge, MA: Addison-Wesley, 1954.
- Cantor, N., & Mischell, W. Traits as prototypes: Effects on recognition memory. Journal of Personality and Social Psychology, 1977, 35, 38-48.
- Carnap, R. The logical foundations of probability. Chicago: University of Chicago Press, 1950.
- Chapman, L. J., & Chapman, J. L. Illusory correlation as an obstacle to the use of valid diagnostic signs. Journal of Abnormal Psychology, 1969, 74, 271-280.
- Cohen, J. L. A Bayesian approach to impression formation. Unpublished doctoral dissertation, University of Illinois, 1973.
- Cohen, L. J. The probable and the provable. Oxford: Clarendon Press, 1977.
- Collins, A., & Loftus, E. A spreading activation theory of semantic processing. Psychological Review, 1975, 82, 407-429.
- Einhorn, H. J., & Hogarth, R. M. Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 1978, 85, 395-416.
- Franks, J. L., & Bransford, J. D. Abstraction of visual patterns. Journal of Experimental Psychology, 1971, 90, 65-74.
- Freize, I., & Weiner, B. Cue utilization and attributional judgments for the success and failure. Journal of Personality, 1971, 39, 591-606.
- Goldberg, L. R. Simple models or simple processes: Some research on clinical judgments. American Psychologist, 1968, 23, 483-496.
- Hammerton, M. A case of radical probability estimation. Journal of Experimental Psychology, 1973, 101, 252-254.
- Hammond, K. R. Toward increasing competence of thought in public policy formation. In K. R. Hammond (Ed.), Judgment and decision in public policy formation. Boulder, Colorado: Westview Press, 1978.
- Hammond, K. R. Human judgment and social policy. Progress in research on human judgment in social interaction, Report 170. Boulder, Colorado: Institute of Behavioral Science, University of Colorado, 1974.

- Hastie, R., & Kumar, P. Person memory: Personality traits as organizing principles in memory for behaviors. Journal of Personality and Social Psychology, 1979, 37, 25-38.
- Heider, F. The psychology of interpersonal relations. New York: John Wiley & Sons, 1958.
- Homa, D., & Chambliss, D. The relative contributions of common and distinctive information on the abstraction from ill-defined categories. Journal of Experimental Psychology: Human Learning and Memory, 1975, 1, 351-359.
- Homa, D., Cross, J., Cornell, D., Goldman, D., & Schwartz, S. Prototype abstraction of new instances as a function of number of instances defining the prototype. Journal of Experimental Psychology, 1973, 101, 116-132.
- Homa, D., & Vosburgh, R. Category breadth and the abstraction of prototypical information. Journal of Experimental Psychology: Human Learning and Memory, 1976, 2, 322-330.
- Hayes-Roth, B., & Hayes-Roth, F. Concept learning and the recognition and classification of exemplars. Journal of Verbal Learning and Verbal Behavior, 1977, 16, 321-338.
- Janoff-Bulman, R. Causality: Psychological and philosophical perspectives. Unpublished doctoral dissertation. Northwestern University, 1977.
- Jenkins, H., & Ward, W. Judgment of contingency between responses and outcomes. Psychological Monographs, 1965, 79, (Whole No. 594).
- Jennings, D., Amabile, T., & Ross, L. The intuitive scientist's assessment of covariation: Data-based vs. theory-based judgments. In press.
- Johnson, T., J., Feigenbaum, R., & Weiby, M. Some determinants and consequences of the teacher's perception of causality. Journal of Experimental Psychology, 1964, 55, 237-246.
- Jones, E. E., & Davis, K. E. From acts to dispositions: The attribution process in person perception. In L. Berkowitz (Ed.), Advances in experimental social psychology (Vol. 2). New York: Academic Press, 1965.
- Jones, M. R., & Myers, J. L. A comparison of two methods of event randomization in probability learning. Journal of Experimental Psychology, 1966, 72, 909-911.

- Judd, C., & Kulik, J. Schematic effects of social attitudes on information processing and recall. Journal of Personality and Social Psychology, 1980, 4, 569-578.
- Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.
- Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 251-273.
- Kelley, H. H. Attribution theory in social psychology. In D. Levine (Ed.), Nebraska symposium on motivation. Lincoln, Nebraska: University of Nebraska Press, 1967.
- Kelley, H. H. The processes of causal attribution. American Psychologist, 1973, 28, 107-128.
- Keynes, J. M. Treatise on Probability. London: Macmillan Co., 1921.
- Kidd, J. B. The utilization of subjective probabilities in production planning. Acta Psychologica, 1970, 34, 338-347.
- Kneale, W. Probability and induction. Oxford: Oxford University Press, 1949.
- Kruglanski, A., Hamel, I., Maides, S., & Schwartz, J. Attribution theory as a special case of lay epistemology. In J. Harvey, W. Ickes, and R. Kidd (Eds.), New directions in attribution research (Vol. 2). New Jersey: Erlbaum, 1978.
- Langer, E. J., & Abelson, R. P. The semantics of asking a favor: How to succeed in getting help without really dying. Journal of Personality and Social Psychology, 1972, 24, 26-32.
- LaPlace, P. S. A philosophical essay on probabilities. New York: Dover Publications, 1951.
- Lau, R., Lepper, M., & Ross, L. Persistence of inaccurate and discredited impressions: A field demonstration of attributional perseverance. Unpublished manuscript, Stanford University, 1976.
- Lyon, D., & Slovic, P. Dominance of accuracy information and neglect of base rates in probability estimation. Acta Psychologica, 1976, 40, 287-298.
- Markus, H. Self-schemata and processing information about the self. Journal of Personality and Social Psychology, 1977, 35, 63-78.

- McNeill, S., & Messick, D. A Bayesian analysis of subjective probabilities. Acta Psychologica, 1970, 34, 311-321.
- Medin, D. L., & Schaffer, M. M. Context theory of classification learning. Psychological Review, 1978, 85, 207-238.
- Michotte, A. The perception of causality. New York: Basic Books, 1963.
- Miller, D. T., & Ross, M. Self serving biases in the attribution of causality: Fact or fiction. Psychological Bulletin, 1975, 82, 213-225.
- Myers, J. L. Probability learning and sequence learning. In W. K. Estes (Ed.), Handbook of learning and cognitive processes (Vol. 3). Hillsdale, NJ: Erlbaum, 1976.
- Nisbett, R. E., Borgida, E., Crandall, R., & Reed, H. Popular induction: Information is not necessarily informative. In J. S. Carroll and J. W. Payne (Eds.), Cognition and social behavior. Hillsdale, NJ: Erlbaum, 1976.
- Oskamp, S. Overconfidence in case-study judgments. Journal of Consulting Psychology, 1965, 29, 261-265.
- Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.
- Peterson, C. R., Ulelha, Z., Miller, A., Bourne, L., & Stilson, D. Internal consistency of subjective probabilities. Journal of Experimental Psychology, 1965, 70, 526-533.
- Piaget, J., & Inhelder, B. The psychology of the child. New York: Basic Books, 1969.
- Piaget, J. The child's conception of physical causality. Totawa, NJ: Littlefield, Adams, & Co., 1972.
- Popper, K. The logic of scientific discovery. New York: Harper, 1965.
- Posner, M. I., & Kneale, S. W. On the genesis of abstract ideas. Journal of Experimental Psychology, 1968, 77, 353-363.
- Reed, S. K. Pattern recognition and categorization. Cognitive Psychology, 1972, 3, 382-407.
- Reed, S. K. Category vs. item learning. Implications for categorization models. Memory and Cognition, 1978, 6, 612-621.

- Reitman, J. S., & Bower, G. B. Storage and later recognition of exemplars of concepts. Cognitive Psychology, 1973, 4, 194-206.
- Rosch, E. H. Natural categories. Cognitive Psychology, 1973, 4, 328-350.
- Rosch, E. H. Principles of categorization. In E. Rosch and B. B. Lloyd (Eds.), Cognition and categorization. Hillsdale, NJ: Erlbaum, 1978.
- Rosch, E. H., Mervis, C. B., Gray, W., Johnson, D., & Boyes-Braem, P. Basic object in natural categories. Cognitive Psychology, 1976, 8, 382-440.
- Ross, L. The intuitive psychologist and his shortcomings: Distortions in the attribution process. In L. Berkowitz (Ed.), Advances in Experimental Social Psychology (Vol. 10). New York: Academic Press, 1977.
- Ross, L., Amabile, T., & Steinmetz, J. Social roles, social control, and biases in the social perception process. Journal of Personality and Social Psychology, 1977, 35, 485-494.
- Ross, L., Greene, D., & House, P. The false consensus phenomenon: An attributional bias in self perception and social perception processes. Journal of Experimental Social Psychology, 1977.
- Ross, L., Lepper, M., & Hubbard, M. Perseverance in self perception and social perception: Biased attributional processes in the debriefing paradigm. Journal of Personality and Social Psychology, 1975, 32, 880-892.
- Schank, R., & Abelson, R. P. Scripts, plans, goals and understandings: An inquiry into human knowledge structures. Hillsdale, NJ: Erlbaum, 1977.
- Simon, H. A. Models of man: Social and rational. New York: John Wiley & Sons, 1975.
- Slovic, P. From Shakespeare to Simon: Speculations--and some evidence--about man's ability to process information. O. R. I. Research Bulletin, Vol. 12, No. 2. Eugene, Oregon: Oregon Research Institute.
- Slovic, P., & Lichtenstein, S. C. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

- Slovic, P., Fischhoff, B., & Lichtenstein, S. Cognitive processes and societal risk taking. In J. S. Carroll and J. W. Payne (Eds.), Cognition and social behavior. Hillsdale, NJ: Erlbaum, 1976.
- Slovic, P., Fischhoff, B., & Lichtenstein, S. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.
- Smedslund, J. The concept of correlation in adults. Scandinavian Journal of Psychology, 1963, 4, 165-173.
- Snyder, M., & Cantor, N. Testing hypotheses about other people: The use of historical knowledge. Journal of Experimental Social Psychology, 1979.
- Snyder, M., & Swann, W. B., Jr. Hypothesis testing in social interaction, Journal of Personality and Social Psychology, 1978, 36, 1202-1212.
- Snyder, M., & Uranowitz, S. W. Reconstructing the past: Some cognitive consequences of person perception. Journal of Personality and Social Psychology, 1978, 36, 941-950.
- Spiro, R. J. Remembering information from text: The "state of schema" approach. In R. C. Anderson, R. J. Spiro and W. E. Montague (Eds.), Schooling and the acquisition of knowledge. Hillsdale, NJ: Erlbaum, 1977.
- Stael von Holstein, C. A. S. Probabilistic forecasting: An experiment relating to the stock market. Organizational Behavior and Human Performance, 1972, 8, 139-158.
- Strawson, P. F. Introduction to logical theory. New York: John Wiley & Sons, 1952.
- Suppes, P. A probabilistic theory of causality. Amsterdam: North Holland Publishing Co., 1970.
- Tsujimoto, R. N. Memory bias toward normative and novel trait prototypes. Journal of Personality and Social Psychology, 1978, 36, 1391-1401.
- Turoff, M. An alternative approach to cross-impact forecasting. Technological Forecasting and Social Change, 1972, 3, 309-339.
- Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

- Tversky, A., & Kahneman, D. The belief in the law of small numbers. Psychological Bulletin, 1971, 76, 105-110.
- Tversky, A., & Kahneman, D. Judgment under uncertainty: Heuristics and biases. Science, 1974, 185, 1124-1131.
- Tversky, A., & Kahneman, D. Causal schemata in judgments under uncertainty. In M. Fishbein (Ed.), Progress in social psychology. Hillsdale, NJ: Erlbaum, in press.
- von Mises, R. Probability, statistics, and truth (3rd Edition). New York: Macmillan Co., 1951.
- Ward, W., & Jenkins, H. The display of information and the judgment of contingency. Canadian Journal of Psychology, 1965, 19, 231-241.
- Wason, P. C. On the failure to eliminate hypotheses in a conceptual task. Quarterly Journal of Experimental Psychology, 1960, 12, 129-140.
- Wason, P. C. Reasoning about a rule. Quarterly Journal of Experimental Psychology, 1968, 20, 273-281.
- Wason, P. C. Regression in reasoning? British Journal of Psychology, 1969, 60, 471-480.
- Wason, P. C., & Johnson-Laird, P. N. Psychology of reasoning. Cambridge, MA: Harvard University Press, 1972.

Appendix A

Instructions and Narrative

Instructions

In this study we are interested in people's ability to remember the factual details of a story that describes a week in the life of another person. There has recently been a great deal of concern with people's ability to notice factual details and to remember these details when they observe or hear about real-life events. Although people seem to be quite good at recalling the gist of a story, there is less support for their ability to remember the factual details of one. The ability to remember factual details can be quite important in certain aspects of our lives. For instance, eyewitness testimony in a court trial depends on the ability to remember often small details of a visual scene or another person's physical appearance.

In our study you will read a story about a week in the life of an actual person--Jane. The names and places have been changed in order to conceal this person's real identity. We would like you to simply read the story and concentrate as you read on the factual details about people, places, and events in the story. Consider that you are reading a story in order to be able to tell someone else about the details contained in that story. After you read the story, we will ask you to tell us about some of the details from the story.

You will have 7 minutes to read the story. It is 4 pages long so you should have plenty of time to read it through at a comfortable pace.

At 7:15 on Monday morning Jane woke up, turned off the alarm clock, and crawled out of bed. She dressed quickly in a tweed skirt and red blouse, ate her usual cornflakes breakfast, and got ready to leave. This morning Jane had a doctor's appointment before work. The mist had gathered on her new Datsun and she stopped just long enough to clean the windshield before starting off. Jane was cautious and drove slowly in the downtown traffic. She pulled up to the large Jackson building on Hyde Street in time to find a parking spot. In the elevator on the way up to the doctor's office, Jane kept aloof from the other passengers and didn't join in the joking about how the elevators in these old buildings take forever. She got off, along with six other people, at the 5th floor.

Jane greeted the doctor's nurse with a smile. While waiting for the doctor she had a conversation with another patient about running and other sports. Finally, the doctor was able to see her. Jane had a quick check-up, consisting of a blood sample, an eye test, and an EKG. She seemed to be in perfect health and left quickly after the check-up. Doctor's offices always seemed to run behind schedule.

Jane arrived at work a little late since her appointment took longer than she expected. Nevertheless, Jane was both efficient and productive and managed to get several important tasks done before the morning coffee break. The people in Jane's office took their 20 minute coffee break together and spent most of the time talking about their weekends. Jane didn't seem to want to take part in this social

activity. She remained reserved and discrete about her personal life. When one of the men asked her for a date, Jane looked to the ground and seemed to be self-conscious. Everyone around the office was talking about the high price of coffee. It was much too expensive and everyone was trying to find an alternative drink. Tea seemed almost as expensive and didn't quite take the place of a good cup of coffee. After a brief discussion, everyone went back to work.

The rest of the day was fairly hectic--as most Mondays tend to be. Jane's job often required close cooperation with other people, which meant that she had to be patient and courteous, but at the same time, deliberate and effective. Jane was known as a very able worker who had the right blend of interpersonal skills and task dedication.

After work, Jane got ready to do her hour's worth of running for the day. She started down Humboldt street near her house, turned right at James Street, and then ran through the whole Brentwood district. Most of the streets around her house were beautifully lined with trees and had relatively traffic-free paths for running. Jane would run vigorously for at least an hour a day. She kept a busy schedule, going from activity to activity, and gave the impression of having unlimited energy. Somehow she managed to remain friendly and outgoing despite her long day at work. Jane loved to run and seemed enthusiastic and cheerful even when the weather wasn't exactly ideal. Today she was running with a neighbor named Mary Stuart and the two of them spend the entire time in spirited talk--cracking jokes and boasting about their running skills and endurance. During

her run, Jane came across people she didn't know but said hello to each of them anyhow. Sometimes she was uncomfortable with strangers, but not today. She surprised herself in how forcefully she exercised after such a long day at work. Sports had lately become a fad in her neighborhood and it was amazing to see the streets lined with joggers at the end of the day. The sportswear industry was certainly doing a booming business--everybody had on their new pair of Adidas and their fancy new sweatpants. Jane was saving up to buy a new pair of running shoes. She had always been pretty thrifty and had little trouble saving for something she really wanted.

Two days later, Jane went to the local Safeway supermarket to do her weekly shopping. She walked down the aisles, amazed at the high price of food. Even the price of soap had gone up in the last week. As a rule, Jane tried to remain unnoticed at the market and avoided interaction with other shoppers from her neighborhood. She was timid and shy, buying her food and waiting patiently in line. Today, Ethel, the checkout clerk, tried to engage Jane in conversation, but Jane remained bashful and hesitant. The big supermarkets were certainly a lot different than the typical local stores Jane remembered from just five years ago. They now had complicated labelling and computerized check-out systems. As far as Jane was concerned, the move to modern supermarkets was a step backwards. Not only did one have to pay outrageous prices for the food these days, but also had to put up with the noise and crowds.

Jane took classes at Redding College three nights a week. The adult education program in the area was superbly run. One of her classes was Communications and the students were practicing debating skills. This was particularly appropriate for Jane since she was about to run for a position in the local government and would be exchanging views with several other candidates. Jane was always tolerant of other people's views, and treated everyone with courtesy, but she was also serious about her own point of view. The local politicians considered that Jane had a good chance of winning in her campaign. She was ambitious and seemed to be confident in her opinions. They also recognized, however, that she was new to politics and hadn't developed many of the more practical skills. She tended, for example, to be a little too soft-spoken. It was never really clear how these local races would turn out. The public got fairly involved in local politics and were genuinely interested in the issues. As in most areas of the country, the current problem was how to get enough money for local programs without taxing the citizens beyond their ability to pay. The taxes in this area had gone up 15% in the last two years. Jane didn't know a great deal about taxation and municipal finances, but she had reasonably good mathematical skills and figured she could learn about the issues quickly.

Jane's college course schedule was heavy this semester--three night classes. She was generally quite anxious over her work. In most of her classes she felt somewhat inhibited and tried to think through a question before she dared ask it. Professor Osborne assigned

a lot of reading in her Literature class, but Jane did not mind spending hours alone in the library. It had recently been remodeled and the booths for studying were comfortable and quiet. The only problem was that the booths in the back part of most rooms were too warm. Jane had meant to complain to the staff about this, but never got around to actually making the complaint.

That weekend, Jane went to McAllister park and sat and daydreamed for hours. Although the park was full of people, Jane appeared deep in thought and withdrawn from the activity around her. She felt a little sad and spent most of the day by herself. Later that evening she went to a pot-luck party at a neighbor's house. Jane had cooked enough chili to feed practically the whole neighborhood. Since everyone brought something different to eat, the dinner was a lot of fun. After dinner they put on some records and Jane danced for much of the rest of the evening. The party went on for quite a while and late that night they all pitched in to help clean up. It was certainly an improvement in the spirit of the neighborhood and everyone discussed what a difference these get-togethers had made in the atmosphere on the block. The only way to make a city liveable was to keep the neighborhoods beautiful and the people in close contact.

The next morning the alarm clock went off as usual at 7:15. Jane turned over in a groggy state of half-sleep, realized it was Sunday, and cursed the alarm clock. She curled up to try to recover that pleasant state of drowsy unconsciousness, but she was no better than most people at falling back to sleep, so she got out of bed and went to make a pot of coffee.

Appendix B

Test of Factual Knowledge

1. What kind of car did Jane own?
2. What street was the Jackson Building on?
3. On what floor was the doctor's office?
4. How long was the coffee break where Jane worked?
5. What was Jane saving her money for?
6. What supermarket did Jane shop at?
7. What was the name of the college Jane attended?
8. How many courses did she take during the semester?
9. What was the percent increase in local taxes?
10. What was the name of Jane's neighborhood?

