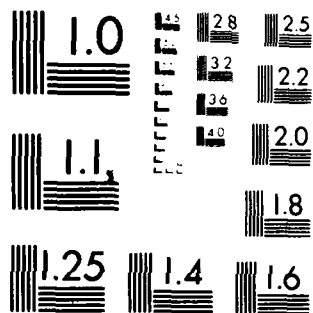


1/4

NL

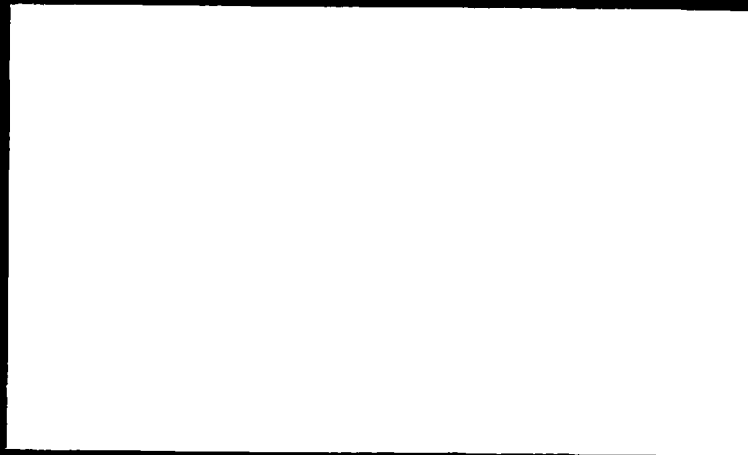
[illegible]



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

AD-
A133 172

CENTER FOR DECISION RESEARCH



DTIC FILE COPY

Graduate School of Business
The University of Chicago

**A THEORY OF DIAGNOSTIC INFERENCE:
JUDGING CAUSALITY**

Hillel J. Einhorn and Robin M. Hogarth
University of Chicago
Graduate School of Business
Center for Decision Research

August 1983

Sponsored by:
Office of Naval Research
Contract Number, N00014-81-K-0314
Work Unit Number, NR 197-071

Approved for public release; distribution unlimited.
Reproduction in whole or in part is permitted for
any purpose of the United States Government.

S E

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 4	2. GOVT ACCESSION NO. A133172	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A THEORY OF DIAGNOSTIC INFERENCE: JUDGING CAUSALITY		5. TYPE OF REPORT & PERIOD COVERED Technical report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Hillel J. Einhorn and Robin M. Hogarth		8. CONTRACT OR GRANT NUMBER(s) N00014-81-K-0314
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Decision Research, University of Chicago, Graduate School of Business, 1101 East 58th Street, Chicago, Illinois 60637		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 197-071
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE August 1983
		13. NUMBER OF PAGES 61
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) Causal judgment, diagnosis, inference, decision making		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Diagnostic inference is concerned with determining the causal process that produced a set of outcomes/results/symptoms. A model of causal reasoning within diagnosis is presented that attempts to answer the following questions: (1) How, and how much, do alternative explanations reduce the causal strength of a hypothesis?; (2) What factors determine the plausibility of a causal explanation?; (3) How does context affect the causal		

DD FORM 1473
1 JAN 73

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

"relevance" of an explanation? We first propose that people use a sequential anchor-and-adjust strategy in discounting an explanation by alternatives. The amount of discounting depends on three factors: the plausibility of alternatives, the initial strength of the hypothesis, and a parameter reflecting the weight given to disconfirmatory evidence. It is then shown that the strength of a causal explanation is highly dependent on an implicit causal background (as in figure/ground relations), and on probabilistic factors called cues-to-causality. The cues considered are temporal order, contiguity, covariation, and similarity of cause and effect. A model for weighting and combining the cues is shown to account for much research in a wide range of fields. The three components of the theory are then tested in a series of experiments and the results are discussed with respect to: (a) the factors that affect the discounting of explanations; (b) issues in combining the cues-to-causality; (c) problems in defining the causal background; and, (d) normative questions in assessing the quality of causal judgments.

Accession For	
NTIS GRA&I	☒
DTIC TAB	☐
Unannounced	☐
Justification	
By _____	
Distribution/	
Availability Codes	
Dist _____	
Special	



A Theory of Diagnostic Inference: Judging Causality

It was late in middle age that Molière's character, Monsieur Jourdain, made the surprising discovery that he had been speaking prose all his life. Similarly, people may be equally surprised to learn that they have been engaged in diagnostic inference all their lives. By "diagnostic inference" we mean the following: given the occurrence of a set of outcomes/results/symptoms, people infer what causal process could have produced the observed effects. The essential aspects of such inferences are that they are causal rather than correlational, backward rather than forward (one goes from effects to prior causes), concerned with a specific rather than the general case, and constructive (one can synthesize, enlarge, or otherwise develop new hypotheses). The importance of diagnostic inference goes beyond its obvious role in making sense of experience; it impacts on choosing between courses of action; and it is crucial for prediction and the defining of "relevant" variables since both depend on some inferred model of the process that generates outcomes (Einhorn & Hogarth, 1982). Furthermore, since the evidence that one has for making diagnoses is fallible and/or conflicting, the process takes place under uncertainty. Thus, the essential nature of inference, "going beyond the information given" (Bruner, 1957), is as true for diagnosis as it is for prediction. However, while much attention in the literature on judgment and decision making has been devoted to prediction (e.g., Kahneman & Tversky, 1973), far less has been paid to diagnosis (for exceptions see, Eddy & Clanton, 1982; Elstein, Shulman, & Sprafka, 1978).

Our approach is to focus on the role that causal judgments play in the diagnostic process. The topic of causal judgment has received considerable attention from a variety of perspectives, e.g., child development (Piaget,

1974; Shultz, 1982); social psychology (Kelley, 1973; Jones, 1979; Nisbett & Ross, 1980); law (Hart & Honoré, 1959; Cohen, 1977; Fincham & Jaspars, 1980); probabilistic inference (Suppes, 1970; Tversky & Kahneman, 1980); medicine (Susser, 1973); methodology (Cook & Campbell, 1979); economics (Zellner, 1979); and, of course, philosophy. However, in recent years, a growing body of research has investigated how people make judgments and decisions under conditions of uncertainty. This approach, called "behavioral decision theory" (for reviews see Slovic, Fischhoff, & Lichtenstein, 1977; Einhorn & Hogarth, 1981) takes as its focus the description, via quantitative models, of the rules and strategies people use in forming judgments and/or choices. Thus, while implicitly and explicitly recognizing the contributions of many of the perspectives enumerated above, we wish to explore how principles from judgment research can illuminate the role of causal thinking in diagnosis.

We begin by drawing an analogy between the processes of diagnosis and perception. In particular: (1) The importance or strength of information in perception depends on the background or field against which it is perceived. For example, object salience involves a figure/ground relation that can be changed by appropriate shifts in the ground as well as in the figure. Similarly, we view the strength of evidence in diagnosis as being highly dependent on an assumed causal background or field and we use the concept of a causal background to determine causal "relevance"; (2) We view the diagnostic process as similar to detecting the appropriate signal in a field of competing signals. However, as in perception, the probabilistic nature of informational cues adds noise to each signal such that a particular pattern of cues is diagnostic of more than a single cause (cf. Campbell, 1966). The importance of this is that the strength of evidence for a particular diagnosis is seen as its net strength; i.e., how well the evidence supports a particular

hypothesis as opposed to its competitors; (3) Diagnosis is a constructive process in that people bring prior expectations to bear in interpreting information and in enlarging hypotheses to account for complex outcomes. In analogous fashion, the importance of expectations and the constructive nature of "achieving" the object are well established in perception (cf. Garner, 1966). Moreover, the introduction of expectations as central to diagnosis (and perception) highlights the role of content knowledge in the assessment of evidence and raises questions of how such knowledge is used; (4) Diagnostic inference often occurs with great speed and a corresponding lack of awareness of the underlying processes. This is also true of perception.

Underlying much diagnostic inference are questions of the following form: outcome Y has occurred, how likely was X the cause? As a specific illustration, imagine that a watch face has been struck sharply by a hammer and the glass breaks. You are then asked to assess how likely the breakage was caused by the force of the hammer. We argue that answers to this question will be mediated by three types of information: (1) The number and strength of specific alternative explanations. Part of the reason that the force of the hammer is a strong causal candidate is due to the fact that it is difficult to imagine specific alternatives that could reduce one's belief in that explanation. (2) The assumed causal background against which the judgment is made (Mackie, 1974). For example, reconsider your response to the above question if the context was changed to a watch factory where a hammer strikes watch faces as part of a testing procedure. As we will demonstrate later, in this context, a defect in the glass is judged as the most likely cause; (3) The judged causal strength of the explanation. We maintain that people use certain cues-to-causality in assessing the quality of an explanation; namely, temporal order, contiguity, covariation, and similarity of cause and

effect. In our example, note that the glass broke immediately after being struck by the hammer; there is a high correlation between the breaking (or not) of glass with the force of solid objects; and there is similarity between the length and strength of cause and effect.

Plan of the Paper

We organize our discussion around the three aspects of causal judgments just noted: (1) How, and how much, do alternative explanations affect the strength of a causal hypothesis?; (2) How does the causal background affect the "relevance" of explanations?; and (3) How are the cues-to-causality combined in assessing the plausibility of a hypothesis and/or its alternatives? Following the development of a theory to answer these questions, we present a series of experiments to test the various components of the theory. Thereafter, we discuss the theory and experimental evidence in relation to: (1) the factors that affect the discounting of an explanation; (2) issues in combining the cues-to-causality; (3) problems in defining the causal background; and, (4) some normative questions in assessing the quality of causal judgments.

The Diagnostic Process

The Effects of Alternatives

How do alternatives affect the judged likelihood of a causal explanation? In this section, we propose a model that rests on the notion that people employ an anchor-and-adjust strategy when assessing the strength of an explanation/hypothesis. To illustrate, consider an outcome Y , an initial explanation X , and alternative explanation Z_1 . Furthermore, denote the "gross strength" of an explanation as being its plausibility or strength before competing alternatives are considered. Thus, the gross strengths of X and Z_1

refer to their plausibility when each is considered the sole explanation of Y. We propose that people anchor on the gross strength of the initial explanation X, and then adjust downward for the gross strength of Z_1 . Moreover, the amount of the adjustment will depend on the strength of the anchor as well as the strength of the alternative. In particular, we assume that alternatives of equal strength discount strong explanations more than weaker ones. For example, imagine that one anchors on a weak hypothesis and is then confronted with a strong alternative. Since the anchor is already low, the size of the adjustment cannot be too large (indeed, if the anchor were worthless, there would be no adjustment). On the other hand, if the anchor was strong, we argue that the same alternative would discount the anchor substantially. Therefore, the basic idea is that the stronger the anchor, the larger the adjustment (holding the strength of alternatives equal). We call the strength of an explanation after it is reduced by an alternative, its "net strength."

The above process can be formally represented as follows:

$$S_1(Y, X|B) = s_0(Y, X|B) - w_0 s(Y, Z_1|B) \quad (1)$$

where,

$S_1(Y, X|B)$ = net strength of the causal link of Y
with X, conditional on background B,
after adjusting for Z_1

$s_0(Y, X|B)$ = gross strength of the causal link of
Y with X, conditional on background B

$s(Y, Z_1|B)$ = gross strength of the causal link of Y
with Z_1 , conditional on background B

w_0 = adjustment weight applied to the gross
strength of Z_1 ($0 < w < 1$)

In equation (1) (and throughout the paper), we adopt the convention that capital "S" stands for net strength and small "s" denotes gross strength. Of course, before any alternative is considered, $S_0 = s_0$. Note that the adjustment weight, w , has the same subscript as the anchor since it is a function of the latter (see below). Now consider what happens when a second alternative, Z_2 , is introduced. We assume that the anchor-and-adjust strategy proceeds sequentially so that the net strength of X becomes the new anchor for the next adjustment. Thus,

$$S_2(Y, X|B) = S_1(Y, X|B) - w_1 s(Y, Z_2|B) \quad (2)$$

Equation (2) can now be generalized to account for the net strength of X after the k th alternative ($k = 1, 2, \dots, K$); thus,

$$S_k(Y, X|B) = S_{k-1}(Y, X|B) - w_{k-1} s(Y, Z_k|B) \quad (3)$$

Furthermore, since $S_k(Y, X|B)$ is a judged likelihood, it is bounded between 0 and 1.

We now consider the functional relation between the strength of the anchor and the adjustment weight, w (called the "adjustment weight function"). It was assumed above that stronger anchors have larger adjustments. This implies that the adjustment weight is a monotonically increasing function of the strength of the anchor. To see this, consider equation (3) when the gross strength of Z_k is constant and the anchor varies in strength. It is clear that as $S_{k-1}(Y, X|B)$ increases, w_{k-1} must also increase to give larger adjustments. In order to model this monotonic relation, we posit a simple and convenient form, although others might serve as well; thus,

$$w_{k-1} = [S_{k-1}(Y, X|B)]^\alpha \quad (\alpha > 0) \quad (4)$$

To discuss (4) and the substantive meaning of α , consider Figure 1. First,

Insert Figure 1 about here

note that the adjustment weight is monotonically increasing with the strength of the anchor, regardless of the value of α . Second, when the anchor is 0 the adjustment weight is 0, for all α . Thus, when a hypothesis is worthless, it cannot be adjusted below 0. Moreover, when the anchor is 1, the adjustment weight is 1. This means that a "certain" explanation will be adjusted solely by the gross strength of an alternative. Third, α affects the amount by which explanations are discounted. For example, $\alpha > 1$ implies that the adjustment weights are less than the anchor; $\alpha = 1$ implies that adjustment weights equal the anchor; $0 < \alpha < 1$ implies that adjustment weights are larger than the anchor. The importance of this for the final net strength of X can be seen by first substituting (4) into (3). This yields;

$$S_k(Y, X|B) = S_{k-1}(Y, X|B) - [S_{k-1}(Y, X|B)]^\alpha s(Y, Z_k|B) \quad (5)$$

Using equation (5) as the computational form of the anchor-and-adjust model, we now illustrate the effect that α can have on net strength. Imagine that the gross strengths of X and Z_1 are .60 and .50, respectively. If $\alpha = .5$, $S_1(Y, X|B) = .21$; if $\alpha = 1$, $S_1(Y, X|B) = .30$; if $\alpha = 2$, $S_1(Y, X|B) = .42$. Therefore, as α increases, the adjustment weight decreases, as does the amount of the adjustment. In accord with this, we interpret α as reflecting the "weight" or importance given to disconfirmatory evidence. Thus, when α is large, alternative hypotheses are weighted less and adjustments are small. Indeed, note that as $\alpha \rightarrow \infty$, the adjustment

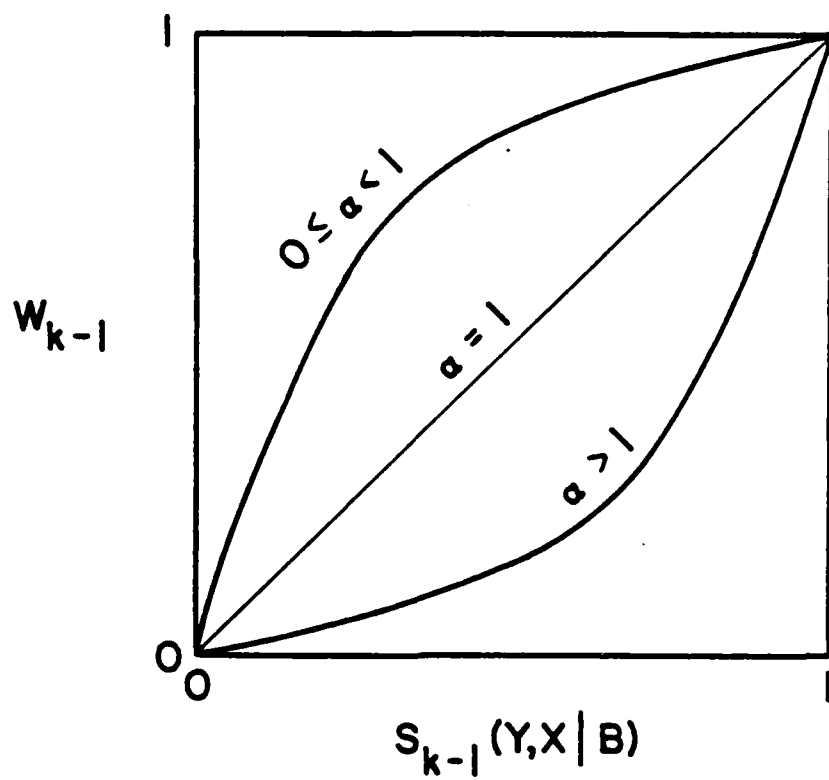


Figure 1. The Adjustment Weight Function

weights go to zero so that alternative explanations have no effect on the initial strength of a hypothesis. In contrast, when $0 < \alpha < 1$, adjustment weights are large and initial hypotheses are strongly discounted by alternatives. When $\alpha = 1$, hypotheses are discounted by alternatives in a "neutral" manner. In the experimental work to be presented later, we both estimate α empirically and predict the net strength of an explanation after the presentation of one and two alternatives. Thus, equation (5) provides a simple and interpretable one-parameter model that is easily subjected to empirical testing.¹

We now discuss how the model specified in (3) and (5) captures important aspects of the causal judgment process. In order to do so, we consider the model in its non-sequential form;

$$S_K(Y, X|B) = s_0(Y, X|B) - \sum_{k=1}^K w_{k-1} s(Y, Z_k|B) \quad (6)$$

Therefore, the net strength of an explanation is equal to its gross strength minus the sum of the adjusted alternative explanations.

There are several important aspects of equation (6): (a) All terms are conditioned on some assumed or implicit causal background. Thus, the strength of any factor as a cause of Y depends on the context being considered (this is considered in detail later); (b) While the gross strength of an explanation can be viewed as analogous to the absolute strength of a signal perceived against a noiseless background, its net strength can be seen as resulting from two conflicting forces; the strength of the signal vs. the strength of competing signals that comprise specific alternative explanations. Note that equation (6) is consistent with the view of causal strength as stressed by Campbell and colleagues (Campbell & Stanley, 1963; Campbell, 1969;

Cook & Campbell, 1979); that is, causal strength should be evaluated by the ruling out of alternatives. Indeed, the assessment of "internal validity," whereby one asks what other factors than X could have produced Y, seems to be important in all causal judgments. In fact, Mackie (1974) states that the primitive notion of a cause involves asking oneself the question: "Would Y have occurred if X had not?" The greater the number of alternative explanations underlying a "yes" answer, the lower the causal strength of X for Y. Note that the posing and answering of the above question (the "counterfactual conditional") may involve doing a real or "thought" experiment. In the former, one compares the effect of X on Y with that of $\bar{X}$ on Y (the control group condition). In this way, the counterfactual question is easily answered. In the latter, one can imagine the world before X, go forward to where X would occur, and then delete it from the scenario. If the scenario is now run forward from that point, one can imagine if Y happens or not. Clearly, in such thought experiments, the construction of "possible worlds" and imaginary scenarios is crucial for judging causal significance.

The idea that counterfactual reasoning and thought experiments are a crucial component of causal inference helps to explain the power of certain explanations in non-experimental situations. As a case in point, consider the following one-shot case study with a single datum: The occurrence of a huge explosion near Los Alamos, New Mexico, in July 1945. No one doubted this to be the effect of detonating an atomic bomb. Clearly, inferring causality in this poorly designed experiment was not difficult whereas assessing causality in the most meticulously designed experiments in social science is often problematic at best. When one considers why the causal inference is so strong in the bomb example, ask yourself the following question: "Would an explosion of such magnitude have occurred if an atomic bomb had not gone off?" While it

is possible to think of alternative explanations for the explosion, they are so unlikely as to be virtually non-existent. Therefore, even in one-shot case studies with no control group, the causal strength of an explanation can be substantial (see Campbell, 1975, for an illuminating discussion of this issue); (c) While the role of alternatives is important in assessing causal strength, equation (6) posits that net strength follows a difference rather than a ratio model. This has important implications for the case where few or no alternatives are imagined. For example, a ratio model (such as probability theory) would treat the strength of evidence for a hypothesis as certain if there were no alternatives. However, in equation (6), net strength can be low when there are no alternatives if the gross strength of X is itself low. (Note that this does not contradict the atomic bomb example given above since we would argue that the gross strength of this explanation is high; i.e., the cues of temporal order, contiguity in time and place, constant conjunction, and similarity of cause and effect, all point to a causal relation.) Moreover, net strength can also be low when gross strength is high if there are many strong alternatives. Indeed, net strength can only be high if gross strength is high and the strength of specific alternatives is low. To illustrate, reconsider the initial watch-hammer scenario and contrast the net strength of the "force of the hammer" explanation with the net strength of any single explanation for the following questions:

1. Why are the outer rings of Saturn braided?
2. Why was Ronald Reagan elected President in 1980?

For the first question, it is difficult to generate a single explanation, thus suggesting its gross strength is low. However, although there are no competing explanations, net strength remains low in accord with equation (6). For the second question, there are many strong explanations (e.g., the

situation of the economy; the rise of the moral majority; the unresolved Iranian hostage problem; etc.). Therefore, while the gross strength of these are high, the net strength for any single one is low precisely because the others are plausible alternatives. On the other hand, the watch-hammer question leads to high net strength since the explanation is strong and there are few plausible alternatives. In short, it is argued that like good patterns, good explanations have few alternatives (Garner, 1970); or, to be more precise, whereas good explanations imply few alternatives, the lack of alternatives does not imply good explanations; (d) While we only consider the causal strength of a single factor in producing Y, equation (6) can be generalized to the assessment of scenarios based on multiple causes. For example, imagine that Bob has been fired (Y), and you know that he was often late to work (X_1), didn't get along with his co-workers (X_2), and was a mediocre performer (X_3). In order to judge the strength of the explanation that Bob was fired because of all three factors, define a complex factor Ω such that, $\Omega = (X_1 \cap X_2 \cap X_3)$. The causal strength of Ω can now be assessed via equation (7). Thus,

$$S_K(Y, \Omega | B) = s_O(Y, \Omega | B) - \sum_{k=1}^K w_{k-1} s(Y, \theta_k | B) \quad (7)$$

where, $\theta_k = k$ th alternative scenario ($k = 1, 2, \dots, K$)

Note that the causal background and alternative explanations (both simple and complex) are still crucial. Thus, the gross strength of Ω would be reduced if, for example, most workers were late, didn't get along with co-workers, etc.; or, a good alternative explanation existed (e.g., the company was going broke and had to let people go). The idea of complex factors or scenarios as

explanations is an important topic that requires separate treatment (cf. Mackie, 1974). However, since scenarios are comprised of individual links, it is first necessary to understand the factors that affect these basic components before studying the more complex case.

Relevance and the Causal Background

We were careful in the preceding section to condition all terms on the causal background, B. The reason for doing this will be discussed in this section. To begin, we ask why some variables seem more causally relevant than others. To answer this question, we first need to consider what events/outcomes trigger diagnostic curiosity. We propose that events of diagnostic interest are those that are unusual, abnormal, or unlikely. Thus, one rarely seeks the cause of why one feels "average," why traffic flowed normally, or why some accident is typical. To be sure, diagnostic curiosity can be aroused vis-à-vis normal events. However, this is most likely to happen when those events violate expectations and are therefore seen as unusual after all. For example, we might want to know why traffic flowed normally if major highway improvements were just completed, or why we feel "average" after hearing about a death in the family. Therefore, diagnostic inference is invoked to make sense of deviations via causal explanation. However, it is important to note that the meaning of a deviation is itself crucially dependent on some assumed background or field. Indeed, even averages can be made unusual with the appropriate shift of background--consider Oscar Wilde's statement that, "moderation shouldn't be taken to extremes."

In searching for a cause of some outcome which is a deviation from the normal or average, we propose that attention is directed toward prior deviations or abnormal events. Thus, unusual effects are seen as the result of unusual causal circumstances. In fact, one can consider this belief a

special case of the "representativeness" heuristic (Kahneman & Tversky, 1972) in that causes and effects are similarly discrepant from some assumed causal background. However, the manner in which the causal background affects the strength of causal links needs to be considered in more detail. Specifically, it is argued that causal relevance is generally related to the degree that a variable is a difference-in-a-background (Mackie, 1974). By this is meant that factors that are part of some presumed background are judged to be of little or no causal relevance. For example, does birth cause death? While the former is both necessary and sufficient for the latter (and thus covaries perfectly with it), it seems odd to consider one the cause of the other. The reason is that death presumes that one has been born. Therefore, "birth" is part of the background and its causal relevance is low.

The importance of the background is not limited to situations in which there is perfect correlation between causes and effects. Consider why oxygen is irrelevant as the cause of a house fire, but relevant in a fire on a spaceship. Since oxygen is equally necessary for fires in both places, some notion of a difference-in-a-background is needed to distinguish these cases. For example, in accord with our model let, Y = fire, X = oxygen, B = causal background for the house fire, and, C = causal background for space travel. In the house fire, the gross strength of oxygen as a causal agent is essentially zero since the causal background B already contains the presumption that oxygen was present. Thus, $s(Y, X|B) = 0$ since X is part of B and cannot be a difference in that background. Moreover, recall that if gross strength is zero, net strength will be zero since the adjustment weight, w , will be zero. Now consider the spaceship fire; note that $s(Y, X|C)$ is not zero since oxygen is not part of the causal background of space flights. Indeed, leaking oxygen would be an important difference-in-the-background.

However, this is not to say that oxygen would necessarily be a strong causal candidate since its net strength would depend on alternative explanations. On the other hand, it would not be immediately dismissed as irrelevant, as in the case of a house fire.

There are several implications to be drawn from considering causal relevance in relation to some assumed background: (a) Shifts in background - imagine the following scenario: Joe is a chemical worker who contracts lung cancer and sues the company for causing his disease. His lawyer argues that the cancer rate of workers in this factory is nine times the national average for workers in comparable industries. Note that the background in this argument is industries of a certain type and the causal argument rests on there being a difference (higher cancer rates) in this background. However, the lawyers for the chemical company may shift the background by arguing that Joe has smoked cigarettes for years, comes from a family with respiratory problems, and so on. Note that the background is *now changed* to people with certain personal habits and characteristics, and in this background, lung cancer may not be unusual. This argument reduces the strength of the chemical factory explanation in two ways - it introduces a strong alternative explanation and, the background shift changes Y from an unusual event that requires a special causal explanation, to a usual event that requires no such explanation. It is expected that the conflict that arises in evaluating evidence that is highly sensitive to background shifts is particularly difficult to resolve; (b) Narrowing/widening the background - equation (6) suggests that net strength can also be altered by narrowing or widening the same background. This occurs because alternative hypotheses are either ruled out by narrowing the context or expanded by widening it. In terms of equation (6), changes in the number of imagined alternatives is represented by a

smaller limit of summation (K) in the adjustment term. Therefore, the width of the background can also affect the causal strength of an explanation. In the above scenario, for example, note how Joe's case would be strengthened if it could be shown that the cancer rate in his factory was nine times the rate of other chemical factories making exactly the same product. The reason is that by narrowing the field to chemical plants making the same product, the number of alternative explanations is reduced, thereby making the difference in the narrowed field more causally relevant. A similar idea has been advanced by Bar-Hillel (1980). In considering the research showing that people ignore base rates in making probability judgments, she demonstrates that base rates will be used if they are made more specific or if they can be given a causal interpretation. She suggests that both specificity and a causal interpretation increase the "relevance" of information, and thus its use. From our perspective, Bar-Hillel's treatment of relevance is consistent with our concept of net strength; both are increased by a causal interpretation of evidence and a narrowing of the background (specificity) which reduces alternatives; (c) Depth of the background - consider the issue of reductionism in causal explanations, where causes at a molar level are different from those at a molecular level. If one thinks of the background B as analogous to the field of vision under a microscope, then shifts in magnification of the lens define different fields. Moreover, since causal relevance is a difference-in-the-field, it is obvious that a cause at one level will not necessarily be relevant at another. This microscope analogy makes clear that the "appropriate" level of magnification depends on one's purposes and the extent of one's knowledge of the phenomenon in question. Thus, a biochemist may see the causal link between smoking and lung cancer as due to chemical effects of tar, nicotine, and the like, on cell structure, while an immunologist might see the

causal link as due to the suppression of the immune system in controlling diseases in general. However, it should be noted that the level of the field is not totally arbitrary in everyday inferences. Indeed, there is remarkable consensus among individuals as to the appropriate level of the assumed background. On the other hand, where large discrepancies exist in knowledge about a particular topic, as in comparing experts to non-experts, such consensus is often lacking.

Components of Gross Strength: Cues-to-Causality

The factors that comprise the gross strength of a causal hypothesis are now considered in detail. Specifically, it is hypothesized that gross strength (conditional on an assumed background), is a function of various "cues-to-causality" such as temporal order, contiguity, covariation, and similarity of cause and effect. However, before discussing these, we note that the term "cues" has a specific meaning that corresponds with its use in Brunswik's psychology (1952; also see Hammond, 1955; Campbell, 1966). Thus:

- (1) The relation between each cue and causality is probabilistic. That is, each cue is only a fallible sign of a causal relation;
- (2) People learn to use multiple cues in making inferences in order to mitigate against the potential errors arising from the use of single cues;
- (3) The use of multiple cues is facilitated by the intercorrelation (redundancy) between them in the environment. This both reduces the negative effects of omitting cues, and aids in directing attention to the presence of others;
- (4) Although multiple cues reduce uncertainty in inference, they do not entirely eliminate it.

The concept of cues-to-causality also contains the following aspects:

- (a) While each cue can be viewed as a unitary concept, it is more useful to consider them as containing several elements. For example, contiguity can be decomposed into temporal and spatial components and temporal contiguity can be

further divided into the time interval between cause and effect and the regularity of the interval. The importance of considering the elements of each cue will become apparent as we proceed, especially with regard to covariation and similarity; (b) The cues are considered to be primitives in the construction of causal theories or scripts/schemas (Abelson, 1981). By this is meant that they serve as basic building blocks in the development of schemas and, conditional on such schemas, they are used to modify and expand on prior theories. This implies that the relations between cues and causal judgments will be affected by prior knowledge and expectations. For example, imagine that one has advertised a product and sales go up dramatically the next day. If one believes that advertising works by a gradual diffusion process, the sales increase may not be attributed to the ad precisely because the events occurred too close in time. On the other hand, contiguity could be seen as monotonic with causal strength by others with different theories, or by the same person in another context. Therefore, the relation between cues and causal judgments is conditional on prior theory. In terms of the basic model represented in equation (6), the conditioning of gross strength on a background B suggests that the particular context engages prior knowledge which, in turn, conditions the cues.

We now consider the individual cues. The first cue to be considered is temporal order (denoted Q_1). The importance of temporal order seems obvious since it labels which of two variables in a relation is cause and which is effect. Furthermore, temporal order is often necessary in learning, as in classical conditioning. Indeed, when the order of presenting the conditioned and unconditioned stimuli is reversed, learning is difficult and attempts at backward conditioning have generally been unsuccessful.

An interesting feature of temporal order is the speed and facility with

which it is used--often without explicit awareness. This is particularly the case in the interpretation of language and can be illustrated by contrasting ordinary discourse with a system that is both acausal and atemporal; e.g., probability theory (cf., Tversky & Kahneman, 1980). To illustrate, consider the conjunction "and," which frequently implies temporal order in everyday English (Strawson, 1952); e.g., he went into the supermarket and bought some coffee. If "going into the supermarket" and "buying some coffee" are represented by S and K, respectively, how should one understand the question, "What is the probability of S and K?" Whereas a statistician would represent the question as $p(S \cap K)$ and ignore the temporal meaning of "and," others may well perceive the question as formally requiring $p(K|S)$. Indeed, to direct attention to the conjunction of the events, it might be helpful to reverse S and K in order to break the implied time order, i.e., "What is the probability of buying some coffee (K) and going into the supermarket (S)?" An experiment to test this assertion was performed using graduate students with at least one statistics course. One group ($n_1 = 24$) was asked to choose how they would represent "S and K" probabilistically, while a second group ($n_2 = 24$) was asked to represent "K and S". Subjects chose from either $p(S \cap K)$, $p(K|S)$, $p(S|K)$, or "none of the above." The results showed an increase for $p(S \cap K)$ when the time order was reversed (58% to 75%). Of further interest was the finding that 38% of the subjects chose $p(K|S)$ in the first group (in accord with the natural order of the events) while no subjects chose $p(K|S)$ in the second group. Clearly, temporal order is an important cue that is difficult to ignore, even when it may be appropriate to do so.

The second cue to be considered is contiguity (denoted Q_2). Contiguity is important because it aids in focusing attention on what variables occurred close in time to, and/or in the vicinity of, some effect Y (cf. Michotte,

1946). Indeed, Siegler has shown that for young children (5-6 years old), temporal contiguity is a very strong cue for inferring causality (Siegler & Liebert, 1974; Siegler, 1976). Moreover, these studies show that older children are less dependent on contiguity alone, being able to make use of multiple cues. In the absence of high contiguity, variables may still be seen as causally related when they can be linked together via prior theory. For instance, the temporal gap between intercourse and birth requires some knowledge of human biology and chemistry to maintain the links between those events. Similarly, to connect the raising of oil prices in the mid-East with increases in the U.S. inflation rate necessitates an economic model to bridge the spatial gap.

Our third cue-to-causality, perceived covariation (denoted Q_3), has been the subject of much research in the judgment literature (see e.g., Crocker, 1981). Whereas variables that covary may be continuous, dichotomous, or a mixture of both, the literature has typically considered judgments between dichotomous variables (X and Y) that can be represented by a 2×2 contingency table. We conceive of covariation judgments based on the cell frequencies in such tables to result from a weighted linear combination process. That is,

$$Q_3 = \sum_{i=1}^4 \beta_i q_i \quad (8)$$

where, $q_1 = (X \cap Y)$; $q_2 = (X \cap \bar{Y})$; $q_3 = (\bar{X} \cap Y)$; $q_4 = (\bar{X} \cap \bar{Y})$; and the β_i are weighting parameters.

Equation (8) provides a simple and convenient way of summarizing much of the research on covariation judgments. For example, Smedslund (1963) and Jenkins and Ward (1965) showed that their subjects' judgments were based almost exclusively on $X \cap Y$ (i.e., $\beta_1 > 0$, $\beta_2 = \beta_3 = \beta_4 = 0$); Ward and

Jenkins (1965), however, changed the way information was presented to subjects (from sequential to intact displays), and found different patterns of use (many subjects ignored disconfirming evidence, i.e., $\beta_2 = \beta_3 = 0$; many other subjects weighted all cells); Einhorn and Hogarth (1978) noted that information is frequently absent from real-world tasks such that $\beta_1, \beta_2 > 0$ but $\beta_3 = \beta_4 = 0$. Furthermore, a recent meta-analysis by Lipe (1982), has shown β_1, β_2 and β_3 to be significant and in the expected direction ($\beta_1 > 0$; $\beta_2, \beta_3 < 0$), when subjects' judgments were regressed onto the four data cells. She also found that β_1 was the largest weight, thus confirming the finding that positive constant conjunction plays an especially large role in judgments of covariation.

On the other hand, numerous studies have also shown that people can and do make use of all the q_i 's that are available (see Alloy & Abramson, 1979; Crocker, 1981). Indeed, Crocker (1982) demonstrated that the type of question subjects are asked can greatly affect attention, and thus the weight particular information is given (also see Arkes & Harkness, 1983). For example, in a study by Schustack and Sternberg (1981), people were given information in the form of the four cues above but were asked for causal rather than covariation judgments. Their results showed significant positive coefficients for both types of constant conjunction (i.e., $\beta_1 > 0, \beta_4 > 0$), and significant negative coefficients for disconfirming data (i.e., $\beta_2 < 0, \beta_3 < 0$). In another study (Waller & Felix, 1982), subjects were asked to judge the same information by answering both a causal and a correlational question. In accord with our view that covariation is a fallible cue to causality, they found a moderate but significant correlation between the two types of judgments ($r = .57$).

In addition to the type of question asked, the purpose for which covari-

ation judgments are made can affect the amount and kind of information used. For example, Seggie and Endersby (1972) have demonstrated that people are sensitive to the strength and direction of all four components of covariation when these are linked to taking actions, as opposed to making judgments *per se*. Moreover, this occurred when data were presented in both sequential and intact displays, and this latter result has also been replicated by Lipe (1982). In a related vein, a classic paper from industrial psychology speaks to the issue of how good people need to be at detecting covariation. Taylor and Russell (1939) examined the sensitivity of success rates, in terms of dichotomous performance measures, to differences in the levels of test-performance correlations. They showed that high success rates can be achieved with low correlations under a variety of conditions. In other words, the contexts of many decisions may not require people to be sensitive to more than rough levels of covariation.

The explicit use of covariation data as a basis for judgments of causality has also been the focus of much research in attribution theory (cf. Kelley & Michela, 1980). Indeed, Kelley (1973, p. 108) speaks of the "covariation principle," i.e., "An effect is attributed to the one of its possible causes with which, over time, it covaries" (italics in original). Kelley's insight was to note that various patterns of information concerning distinctiveness, consensus, and consistency, corresponded to covariation with given alternative causes (i.e., person, stimulus, circumstances), and this view has received empirical support (see e.g., McArthur, 1972). While such a position is compatible with our own (see Lipe, 1983), our emphasis on multiple cues, the causal field, and a detailed combining rule, distinguishes the two approaches.

Finally, in accordance with our framework, we emphasize that perceived

covariation is conditioned on a specific causal background. However, we also argue that people are sensitive to the extent to which a statistical relation holds up across several backgrounds. In short, confidence that one has identified a causal relation may be bolstered to the extent that it is robust against changes in conditions (Toda, 1977). To illustrate, if researchers detected a statistical relation between a particular diet and a form of cancer in the U.S., the causal significance of this finding might be changed depending on the degree to which the relation was found to hold in other countries.

The cue of similarity (denoted Q_4) is fundamental to causal judgments. Like covariation, similarity can be modeled as a function of its elements, some of which add, and some subtract, from its strength. That is, following Tversky (1977), similarity judgments can be defined as a weighted linear function of the common elements of two objects (cf. constant conjunction) minus the distinctive elements of each (cf. disconfirming data). However, to extend this conception of similarity from objects to causes and effects, it is necessary to specify the common and distinctive elements of the latter. These can be considered at several levels. First, there is a long-standing notion that cause and effect should exhibit some degree of physical resemblance. Mill noted that this is a deeply rooted belief that, "not only reigned supreme in the ancient world, but still possesses almost undisputed dominion over many of the most cultivated minds" (cited in Nisbett & Ross, 1980, p. 115). Mill thought that such a belief was erroneous and many cases exist in which physical resemblance has been misleading. For example, Nisbett and Ross (1980) point out that physical resemblance was the cornerstone of a medical theory called the "doctrine of signatures" whereby cures for diseases were thought to be marked by their resemblance to the symptoms of the disease.

Thus, the curing of jaundice was attributed to a substance that had a brilliant yellow color (see also Shapiro, 1960; Shweder, 1977). However, whereas physical resemblance may be a cue of low validity, it does not mean it has no validity. Indeed, there are many examples of where it is useful.

At a second level, one can consider similarity based on such elements as the length and strength of cause and effect. That is, if the effect of interest is large (i.e., is of substantial duration and/or magnitude), people will expect the cause(s) to be of comparable size. For example, the germ theory of disease advanced by Pasteur must have seemed incredible to his contemporaries in that people were asked to believe that invisible creatures caused death, plagues, and so on. In the same way, it is equally difficult for many to believe that billions of dollars spent on social programs in the '60s and '70s could have had little or no effect, or that long term and complex effects like poverty can have short term and simple causes.

It seems clear that of all the cues, similarity is most dependent on prior knowledge and context. Indeed, when similarity is considered at higher levels of abstraction, as in metaphor, the line between similarity as a cue and similarity as encompassing prior theory is ill-defined. Furthermore, since similarity involves particular causes and effects, it could be argued that content knowledge, and thus prior theory, is always engaged in assessing the strength of this cue. While we agree with such a position, we nevertheless feel that it is useful to delineate the various types of similarity used in determining causal strength.

Combining Cues-to-Causality

Given that multiple cues are used in making inferences, it is important to determine how they are combined. Fortunately, there is an extensive literature concerned with modeling the cue combination process (see Hammond,

McClelland & Mumpower, 1981 for an overview). From our perspective, a major issue in this literature concerns the type of combining rule people use; i.e., is the rule compensatory (thereby implying trade-offs between the cues), non-compensatory (implying no trade-offs), or some mixture (allowing some cues to trade-off but restricting others)? For the cues considered here, the last alternative seems most attractive. The reason is that contiguity and covariation seem likely to trade-off to some degree; similarity to a lesser degree; and temporal order least of all. Therefore, the following partially compensatory model for combining cues-to-causality in determining gross strength, is proposed;

$$s(Y, X | B) = Q_1 \gamma (\lambda_2 Q_2 + \lambda_3 Q_3 + \lambda_4 Q_4) \quad (9)$$

where, Q_1 = temporal order = (0,1)

Q_2 = contiguity

Q_3 = covariation

Q_4 = similarity

$$\gamma = \begin{cases} 0 & \text{if } Q_4 < \text{threshold} \\ 1 & \text{if otherwise} \end{cases}$$

λ_i = importance weight for the i th cue ($i = 1, \dots, 4$)

Note that if either temporal order is inappropriate or similarity is below threshold, gross strength is zero.² Otherwise, the cues of contiguity, covariation, and similarity will trade-off. The evidence in favor of (9) comes from several sources. First, consider Michotte's (1946) demonstrations of the perception of causality induced by moving objects. In particular, Michotte's subjects perceived causal relations when the movement of objects after contact was congruent with prior trajectories and/or positions. On the

other hand, when contiguity was high but similarity low, no causal relation was perceived. For example, there was no causal impression when one object touched the other and the latter changed color, got larger, or rose. Indeed, Michotte noted that to obtain a causal effect, "requires a certain degree of similarity between the movement of the agent and the change in the patient, without which the change would not appear as an 'extension' of the first" (Michotte, 1946, p. 210). Further evidence concerning the threshold nature of similarity is provided by the literature demonstrating the limits of classical conditioning. For example, whereas Watson and his colleagues were able to condition little Albert to fear rabbits by pairing the appearance of a rabbit with that of a large noise, they could not produce the same effect when the rabbit was replaced by a block of wood or a cloth curtain (Nisbett & Ross, 1980, p. 104).

Garcia and his colleagues (Garcia, et al., 1968; Garcia, et al., 1972; Garcia, 1981) have also shown both the necessity of similarity and the fact that it will trade-off with other cues. For example, they have demonstrated that rats can learn to associate, after one trial, distinctive tasting food and a gastro-intestinal illness (induced by x-rays) several hours later. Thus, the similarity of food taste and intestinal illness compensated for the lack of temporal contiguity. However, the threshold nature of similarity was shown by the fact that rats do not learn to associate a different shape of food to the illness, when the taste is familiar. In a related vein, Seligman (1970) has reviewed many learning studies and concluded that organisms are differentially prepared to learn different types of relations. The extent to which such is biological and could be overcome by relevant environmental contingencies is controversial. However, the fact remains that some level of similarity between cause and effect, in terms of congruity of length

and strength and/or physical resemblance, is a crucial cue and may often be necessary.

The idea that similarity can be traded-off has also been demonstrated in studies of children's causal judgments (see Sedlak & Kurtz, 1981 for a review). For example, Shultz and Ravinsky (1977), in contrasting covariation with similarity, found that 6-year olds were unwilling to label dissimilar factors as causes, even in the presence of systematic covariation. On the other hand, older children (10-12 years old) favored covariation over similarity. They also found that the relative weights given to similarity vs. temporal contiguity varied according to age; similarity outweighed temporal contiguity for 6-year olds, but older children favored contiguity over similarity. It is interesting to note that whereas animals and young children may resolve conflicts between similarity and contiguity by compromise judgments, adults can resolve the conflict in a more sophisticated way; viz., by distinguishing between "precipitating" and "underlying" causes. The former is generally some action or event that is high in temporal and spatial contiguity but low in similarity of length or strength with the event. The latter is generally based on high similarity of length and strength, with contiguity being less important. For example, the precipitating cause of World War I was an assassination in Sarajevo, but the underlying cause(s) were economic upheaval, German nationalism, and so on.

Whereas conflicts exist between pairs of cues, conflict can also exist between all the cues. This issue can be highlighted by considering the concept of spurious correlation (Einhorn & Hogarth, 1982). The existence of this concept suggests that some correlations are more (or less) causally related than others, and thereby raises the issue of how to tell the difference (cf. Simon, 1954). For example, consider the correlation between the

number of pigs and the amount of pig-iron (Ehrenberg, 1975). Such a correlation seems spurious when the common causal factor, "economic activity," is considered. On the other hand, consider the correlation between amount of rain and number of auto traffic accidents in a city, over the course of a year. Such a correlation does not seem spurious (or at least, less spurious). What is the difference between these two cases?

If one makes use of the cues-to-causality, the spuriousness of the correlation between pigs and pig-iron becomes apparent. That is, although the covariation and temporal contiguity cues point to a causal relation, the other cues do not. Specifically, temporal order cannot be used to specify which variable is cause or effect; there is low spatial contiguity (it being unlikely that farms and factories are in close physical proximity); and the similarity of the variables is only with respect to their names. Indeed, the judgment that the relation is spurious is made easily and is in full accord with equation (9). That is, the two most important cues in the equation, temporal order and similarity, point away from a causal relation. Thus, the judgment of spuriousness can be made with much confidence. Now consider the second case: assuming that there is a statistical relation, note how the other cues reinforce that link. The temporal order of rain and accidents is clear; contiguity is high both for time and space; and similarity, via the use of prior knowledge about the effects of slippery roads, is high. There seems less doubt that the correlation is "real."

When cues-to-causality conflict, spurious correlation is not the only outcome; e.g., a low or zero statistical correlation could mask a true causal relation. To illustrate, imagine that we were ignorant as to the cause of birth. However, it has been suggested that sexual intercourse is related to pregnancy and the following experiment was designed to test this hypothesis:

TABLE 1

Data Matrix for Hypothetical
Intercourse-Pregnancy Experiment

		Pregnancy		
		Yes	No	
Intercourse	Yes	20	80	100
	No	5	95	100
		25	175	200

100 couples were allocated at random to an intercourse condition, and 100 to a non-intercourse condition. As indicated in Table 1, 25 females became

Insert Table 1 about here

pregnant, and 175 did not. In light of our current knowledge (but unknown to our hypothetical selves), we can state that the 5 people in the no-intercourse/yes-pregnancy cell represent "measurement error," i.e., faulty memory in reporting, lying, etc. Since the statistical correlation is small ($r = 0.34$), we might question whether the hypothesis is worth pursuing. Indeed, if the sample size were smaller, the correlation might not even be "significant." Moreover, even with a significant correlation, $r^2 = 0.12$, which is hardly a compelling percentage of the Y variance accounted for by X.

There are two important implications of this example. First, whereas statistics texts correctly remind us that correlation does not necessarily imply causation, the imperfect nature of this cue to causality is also reflected in the statement: causation does not necessarily imply correlation. We have somewhat facetiously labeled examples of the latter as "causalations," giving them equal standing with the better-known and opposite concept of spurious correlation. Second, causalation demonstrates that sole reliance on a single cue, such as covariation, is inadequate for understanding causal relations. Indeed, the use of multiple cues highlights the role of judgment in such interpretations (see also Simon, 1954) and the cues-to-causality provides a basis for understanding how these are formed.

Empirical Evidence

Our theory deals with both the variables and combining rules used in making causal judgments. In addition, we have organized our discussion around three major components; the discounting of explanations by alternatives, the

role of the causal background, and the importance of cues-to-causality. Our experimental strategy is to test each of these components separately although their interdependence leads to joint testing in some experiments. The experimental work is organized as follows: (a) we first provide a test of the anchor-and-adjust model by fitting its parameter to experimental data and then comparing its predictions for various combinations of explanations and alternatives; (b) we next consider how the cues-to-causality are combined in determining the gross strength of an explanation; (c) we demonstrate how a shift in the causal background affects judgments of causal relevance.

Experiment 1

The purpose of this experiment was to test the anchor-and-adjust model by estimating the parameter α (equation (5)) from empirical data and then subjecting the model to a predictive test. To do this, subjects were first given a short paragraph to read in which the four cues to covariation (recall equation (8)) were presented for a dichotomous outcome (Y) and a suspected cause (X). Subjects were then asked to rate how likely they thought the outcome was caused by X. After doing this, another paragraph was presented in which an alternative explanation (Z_1) was given. The evidence for Z_1 also consisted of the same type of covariation data. Subjects were then asked to rate the likelihood of the original hypothesis in light of the additional evidence. A second alternative (Z_2) was then presented in the same way and the subject again rated the likelihood of the original explanation. Therefore, in our terms, each subject made a judgment of the gross strength of X (no net strength judgments (after alternatives Z_1 and Z_2)).

Subjects. A total of 197 subjects participated in this experiment; 119 were University of Chicago students and staff recruited through ads placed on

campus; 31 were University of Illinois students; and 47 were members of a church group that agreed to participate. All subjects were paid \$2.00 (a donation was made to the church for those in the latter group).

Stimuli. Two different content scenarios were used.³ The first involved the cause of birth defects. The three explanations were: (a) whether the mothers had drunk at least one alcoholic drink per day during pregnancy; (b) whether the mothers drank coffee daily; and, (c) whether the parents had a history of mental illness. The second scenario concerned a marathon race in which the participants ran faster than, or equal to, their own average in previous races. The three explanations given for differential performance were: (a) whether the participants had run in the same event before; (b) whether the runners engaged in sexual activity the day before the race; and (c) whether they had participated in a special one-week diet before the race. In order to induce different gross strengths of the explanations in both scenarios, the statistical correlation between the possible cause and the effect was varied. In the birth defects scenario, the alcohol explanation had an $r = .34$; coffee had an $r = .20$; mental illness, $r = .19$. In the marathon scenario, the diet explanation had an $r = .34$; previous race had $r = .25$; sexual activity, $r = .03$.

Design and procedure. For each scenario, there are 6 permutations of the 3 explanations ($3! = 6$). Thus, each explanation can appear twice as the initial hypothesis, although with a different order of the alternatives. Accordingly, there were 6 experimental conditions representing each of the possible orders and subjects were randomly assigned to one of the conditions. After subjects completed one scenario, they were assigned to one of the six conditions in the other scenario and completed the three ratings as before. Order of presentation of the scenarios was randomized across subjects.

Estimating the model. To fit equation (5) to the experimental data, we used the average judgments of the subjects in each of the six orders. Thus, the anchor-and-adjust model can be re-written as,

$$\bar{S}_k(Y, X|B) = \bar{S}_{k-1}(Y, X|B) - [\bar{S}_{k-1}(Y, X|B)]^\alpha \bar{s}(Y, Z_k|B) + \epsilon \quad (10)$$

where, ϵ = error due to judgmental inconsistency

To illustrate how we estimated α , consider the birth defects scenario with the explanations given in the order, "alcohol-mental illness-coffee." When the alcohol explanation was given first, the average judged likelihood (gross strength) was .58. After the first alternative, net strength was .51; after the second alternative, net strength was .46. Now consider the first net strength as a function of the gross strengths of X and Z₁ (mental illness);

$$.51 = .58 - (.58)^\alpha \bar{s}(\text{mental illness}) + \epsilon \quad (11)$$

Since the mental illness explanation appears as the initial hypothesis in two of the other orders, its average rating was used as the gross strength of the alternative (.40) in (11). When equation (11) is solved for α , $\alpha = 3.28$.

In a similar manner, we computed α for the first net strengths in the five other orders and took the average as our best estimate ($\bar{\alpha}$). This value was then substituted into equation (10) to predict the net strengths of X after both one and two alternatives. Therefore, the basic test of the model concerns how well it predicts the discounted causal strength of an explanation.

Results. The first results concern whether the covariation data affected the average gross strengths of the three explanations in both scenarios. In the birth defects scenario, the average gross strengths (with the corresponding r's in parentheses) were: alcohol, .52 (r = .34); coffee, .49 (r = .20); and mental illness, .40 (r = .19). In the marathon scenario, the results

were: diet, .58 ($r = .34$); previous race, .46 ($r = .25$); and sexual activity, .22 ($r = .03$). Therefore, the average gross strengths of the explanations were monotonically increasing with the covariation cue.

The major results for predicting the net strength of X in the two scenarios are shown in Table 2. The table shows the six orders for both

Insert Table 2 about here

scenarios, the average gross strength of X for the specific order ($s_o(X)$), the average net strengths after alternatives Z_1 and Z_2 (S_1, S_2), and the predicted net strengths ($\hat{S}_1, \hat{S}_2$) using the model in (10). The actual values in the table are averages based on approximately 30 subjects per order. First, note that the $\bar{\alpha}$ values are greater than 1 and quite similar in both scenarios ($\bar{\alpha} = 2.61$ and 2.42 for the birth defects and marathon scenarios, respectively). Recall that when $\alpha > 1$, adjustment weights are less than the anchor and disconfirmatory evidence receives a small weight. Therefore, for our subjects in these scenarios, a single explanation is not greatly discounted by alternatives. We emphasize the conditional nature of our results by stressing that other scenarios may induce a different weighting of alternative explanations.

The basic test of our model involves the accuracy of the predictions of net strength. Consider the birth defects scenario and note how closely the model's predictions match the actual data. Indeed, the mean absolute deviation (MAD) of actual versus predicted is .017. The results for the marathon scenario were not quite as good (MAD = .030). Moreover, in this scenario, two orders (diet-race-sex; race-sex-diet) had $S_2 > S_1$, contrary to the model. However, over both scenarios and both net strengths, we consider these results as strongly supporting the anchor-and-adjust model (the astute reader will no doubt infer that our own $\alpha > 1$ for our theory).

TABLE 2

Actual and Predicted Net Strengths for
Two Scenarios in Experiment 1

<u>O r d e r s</u>			<u>Strength of X</u>				
X	Z ₁	Z ₂	s ₀ (X)	S ₁	$\hat{S}_1$	S ₂	$\hat{S}_2$
Birth defects ($\bar{\alpha} = 2.61$)							
Alcohol	Ment. Ill.	Coffee	.58	.52	.49	.45	.43
Alcohol	Coffee	Ment. Ill.	.49	.42	.41	.38	.38
Coffee	Ment. Ill.	Alcohol	.50	.45	.44	.41	.39
Coffee	Alcohol	Ment. Ill.	.48	.40	.40	.36	.36
Ment. Ill.	Coffee	Alcohol	.43	.37	.37	.35	.33
Ment. Ill.	Alcohol	Coffee	.36	.26	.33	.27	.25
Marathon ($\bar{\alpha} = 2.42$)							
Diet	Sex	Race	.57	.51	.51	.46	.42
Diet	Race	Sex	.59	.45	.46	.49	.42
Race	Sex	Diet	.41	.40	.38	.41	.34
Race	Diet	Sex	.52	.42	.40	.41	.39
Sex	Race	Diet	.16	.13	.16	.10	.12
Sex	Diet	Race	.27	.26	.25	.19	.24

Experiment 2

The major emphasis of Experiment 1 was to put the anchor-and-adjust model to a predictive test. To do so, we manipulated gross strength via the covariation cue. The purpose of the experiments reported here is to demonstrate the effects of varying several cues-to-causality simultaneously. In so doing, we wish to test the form of equation (9); i.e., the rule that describes how the cues are combined in assessing gross strength. In addition, we provide further predictive tests of the anchor-and-adjust model.

Experiment 2A

Subjects were first required to read scenarios and then to rate the likelihood that two variables were causally related. The cues manipulated in the experimental design were contiguity, similarity (defined operationally below), and the four data cues (q_1) that comprise covariation in the dichotomous case. After the rating, subjects were provided with a specific alternative and then asked to re-assess causal strength.

Subjects. There were 32 subjects recruited through an advertisement in the University student newspaper. They were offered \$5 an hour to participate in an experiment on judgment. Their median age was 24, their educational level was high (mean of 4.4 years of post high school education), and there were 16 males and 16 females.

Stimuli. The stimuli consisted of eight scenarios varying in length from 100 to 200 words. These concerned: (1) The efficacy of accounting reports in a chain of supermarkets; (2) the study habits of a graduate student; (3) food-poisoning following a church picnic; (4) weight-loss after attending a health program; (5) the effects of environmental factors on the health of high school

students; (6) the playing schedule of a tournament tennis player; (7) the effects of new school textbooks on academic performance; and (8) the relation of diet to the performance of marathon runners.

Operational Definitions

Two levels, high and low, of each of the causal cues were made operational in the following manner. For similarity, we first created cause-effect pairs that we deemed to vary in similarity. This was independently verified by having subjects rate the similarity of the pairs on a 0-10 scale. The mean judgments for the high similarity pairs was 6.7 while the mean for the low similarity pairs was 3.1. Note that since equation (9) implies that similarity cannot be traded-off below a threshold, low similarity in this experiment had to be above some minimum level. Independent ratings were also collected for judgments of similarity for specific alternatives (mean of 7.5). For contiguity, high and low levels were simply defined by their physical values (e.g., time in days). Several studies reviewed above have shown that perceived covariation is sensitive to the difference between confirming and disconfirming data. Thus, to operationalize this variable in the high condition, the ratio of confirming to disconfirming data was set at approximately 2 to 1; in the low condition, the scenarios contained equivalent amounts of confirming and disconfirming data. In fact, to avoid giving subjects identical patterns of data across scenarios, the distribution of data in the four dichotomous cells was slightly varied. Statistically, the high covariation condition can be characterized by correlations between .33 and .40, the low condition by coefficients between .00 and .10.

Procedure and design. Subjects were presented with a booklet containing the 8 scenarios as well as several other experimental tasks. They were instructed to work at their own pace and the average completion time was 1

hour. The 8 scenarios were interspersed with other material to minimize carry-over effects and to provide variety for the subjects. After reading each scenario, subjects were required to mark their response on a 0-100 rating scale. Furthermore, they were permitted to make notes or calculations. After completing the rating, they were presented with a specific alternative explanation on the following page. They then re-rated the strength of the original causal variable and proceeded to the next task.

The experiment followed a 4-factor within-subjects design where the first three factors were the causal cues, and the fourth factor contained the 8 scenarios arranged in a Latin-square. Specifically, each subject rated 8 different scenarios, where each scenario contained one of the $2 \times 2 \times 2 = 8$ combinations of the cues. In order to form an 8×8 Latin-square, 4 subjects were randomly assigned to each of 8 groups.⁴

Results. Tables 3 and 4 report the main effects and interactions of

Insert Tables 3 and 4 about here

the cues. Note that the main effects for similarity and covariation are significant and in the expected direction (i.e., higher values lead to greater causal strength). Note also that there is a small but significant interaction between similarity and covariation (the interaction shows that similarity has a larger effect when combined with high rather than low covariation; i.e., it follows a "fan" shape). However, there is no effect for contiguity.

That the specific content of scenarios is important is evidenced by the scenario main effect as well as two weaker interactions. The scenario $\times$ similarity interaction arises because there was no effect for similarity in two scenarios (involving the weight-loss, and tennis player). The scenario $\times$ covariation interaction occurred because differences in the two levels of covariation had quite different effects in specific scenarios; e.g., almost no

TABLE 3
Analysis of Variance for Experiment 2A*

Source of variation	df	MS	F	p <
<u>Between subjects</u>	31			
Groups	7	820.78	< 1	n.s.
Subjects within groups	24	1,687.50		
<u>Within subjects</u>	224			
Similarity (A)	1	12,762.53	29.45	.01
Contiguity (B)	1	1,410.94	3.26	n.s.
Covariation (C)	1	41,692.53	96.21	.01
A x C	1	2,697.50	6.22	.05
B x C	1	561.09	1.30	n.s.
Scenarios (D)	7	2,145.56	4.95	.05
D x A	7	1,397.28	3.22	.01
D x C	7	1,040.65	2.40	.05
D x A x C	7	620.10	1.43	n.s.
Error (within)	191	433.37		

*The experiment was also analyzed by a regression model using a dummy variable coding scheme. The pattern of significance of main effects and interactions was similar to that shown in this table. In addition, the overall fit of the regression analysis was characterized by an R^2 (adjusted for degrees of freedom) of .34.

TABLE 4

Mean Ratings of Causal Strength by
Covariation and Similarity:
Experiment 2A

		Covariation		
		High	Low	
Similarity	High	.58 (.37)	.26 (.19)	.42
	Low	.37 (.25)	.19 (.14)	.27
		.47	.22	.34

effect in the accounting report scenario, but large effects in the scenarios concerning study habits, the tennis player, and the marathon runners. Whereas we had no theory to predict these specific interactions, we consider them important in that they support the notion that the cues are perceived conditionally on the context or causal field in which they are embedded.

A surprising result was the lack of an effect for contiguity. However, we reasoned that since the cues-to-causality are partially redundant in the natural environment, subjects may not pay full attention to all potential cues in a scenario. We consequently re-wrote four of the eight scenarios to emphasize contiguity. Thirty-two subjects were then recruited and a further experiment conducted. Results showed main effects for similarity, covariation, and contiguity, and all in the expected directions. Thus, the results for the first two cues replicate our earlier findings and the significant result for contiguity illustrates that this cue can be used if it is made sufficiently salient.

A further test of the anchor-and-adjust model. Since each subject re-assessed causal strength after being presented with a specific alternative, we were further able to estimate equation (5) and put it to a predictive test. However, unlike Experiment 1, we had no independent estimate of the gross strength of the alternative. Rather, this had to be inferred from the data. To demonstrate our procedure, consider Table 4 which not only shows the main effects for covariation and similarity, but the mean ratings after seeing the specific alternatives (in parentheses). Since the alternatives were identical in all four conditions, equation (5) implies that the relation between mean judgments before and after the alternatives can be represented by four equations in two unknowns, viz:

$$.37 = .58 - .58^{\alpha} C \quad (\text{high covariation/high similarity}) \quad (12a)$$

$$.25 = .37 - .37^{\alpha} C \quad (\text{high covariation/low similarity}) \quad (12b)$$

$$.19 = .26 - .26^{\alpha} C \quad (\text{low covariation/high similarity}) \quad (12c)$$

$$.14 = .19 - .19^{\alpha} C \quad (\text{low covariation/low similarity}) \quad (12d)$$

where, C = mean gross strength of the specific alternatives.

The above equations suggest the following test: estimate α and C using two equations, and then predict the net strength judgments in the other two conditions by using the estimated quantities. Accordingly, we estimated $\alpha = 1.25$ and $C = .41$ from equations (12a) and (12b). These parameter estimates were then used to predict the net strengths of equations (12c) and (12d). The predictions were .18 (observed net strength = .19) and .14 (observed net strength = .14), respectively.⁵ Apart from the accuracy of these predictions, we believe these results to be significant in that they demonstrate that our model can be used to make specific predictions even when independent estimates of the gross strength of alternatives are not available.

Experiment 2B

Equation (9) states that if the similarity cue is below a threshold value, gross strength will be zero. We tested this by varying the similarity of X and Y so that some alternative explanations would have zero gross strength and therefore, should not discount the initial hypothesis.

Subjects. Eighty subjects participated in Experiment 2B. They were all MBA students at the University of Chicago, enrolled in the basic graduate level statistics course.

Stimuli. The stimuli consisted of two of the eight scenarios drawn from those used in Experiment 2A. For each scenario, there were two possible alternatives, one high in gross strength and the other low. Moreover, for each scenario, half of the stimuli were paired with the strong alternative and

half with the weak. Both initial stimuli were characterized as having high values of all three of the causal cues and these were operationalized in the same way as Experiment 2A. The similarity ratings for the alternative explanations averaged 8.9 in the strong vs. 1.0 in the weak condition. Note, in particular, that the low similarity rating for the alternatives is below that used in Experiment 2A (3.1 on the same 0-10 scale) since we wished to design alternatives for which the similarity cue was below the postulated threshold in equation (9).

Procedure and design. Subjects were given booklets containing the two scenarios and they were asked to rate the causal strength of a given factor on the same 100 point scale used in Experiment 2A. Following this, subjects were given an alternative explanation and then re-rated the causal strength of the original hypothesis. After making these two ratings, the second scenario was considered in the same way. Subjects were randomly assigned to one of two conditions: half the subjects received scenarios paired with strong alternatives, and the other half received scenarios paired with weak alternatives. In addition, the order of scenario presentation was randomized across subjects.

Results. For the judgments paired with poor alternatives, our assumption concerning the similarity threshold in equation (9) implies that the net strength of X should equal its initial gross strength. Furthermore, since the judgments made after receiving the "good" alternatives are based on a subset of the alternatives used in Experiment 2A, we can also predict these net strength judgments by using the estimates for α and the gross strength of alternatives (C) from that experiment (recall that these values are $\alpha = 1.25$, $C = .41$). The resulting predictions and observations are presented in Table 5. This table shows that, in accordance with our theory, the weak

Insert Table 5 about here

alternatives have virtually no effect. In addition, the mean absolute prediction error for the net strengths after the strong alternative is .03. The fact that the estimates used to make these latter predictions come from an independent set of data (i.e., different subjects in a different experiment), further attests to the predictive power of our model.

Experiment 3

We noted earlier that shifts in the causal background can change the relevance of a causal explanation. The following experiment was designed to test this.

Method and Results

Sixty-seven subjects were recruited from the University of Chicago community and asked to respond to various experimental stimuli as part of a study on decision making. Each subject was asked to respond to two scenarios, with a gap of some 40 minutes between them (during which time other experimental tasks were administered). Subjects were randomly assigned to one of two groups that received the scenarios in different orders. The two scenarios were as follows:

- (1) A watch is placed on a table, face upwards. A hammer is then brought down sharply on the face of the watch. The glass of the watch face breaks and shatters.
- (2) In a watch factory, procedures exist for testing various aspects of the end product. One procedure is the following: A watch is placed on a table, face upwards. A hammer is then brought down sharply on the face of the watch. Imagine that on one occasion the glass of the watch face breaks and shatters.

Both scenarios were followed by identical questions:

Question: What caused the glass to break?

- a. The force of the hammer?
- b. The defect in the glass?
- c. Some other explanation (please specify).

Please circle the most likely cause.

Results of the experiment are presented in Table 6, and are shown for the

Insert Table 6 about here

combined groups since the order of presentation had no significant effect. The table shows both the marginal and joint distributions of responses to the different versions of the scenario. In the first scenario, 60 subjects (91%) judged the force of the hammer as the most likely cause; however, in the factory setting the defect in the glass is seen as the most likely causal agent (36 subjects or 55%). Moreover, of the 60 subjects who said the force of the hammer was the most likely cause in the first scenario, 32 subjects reversed the order in the second. The experimental evidence clearly demonstrates that the relevance of a causal hypothesis can be changed by varying the background.

Discussion

The implications of our theory are now discussed with respect to:

(1) the factors that affect the discounting of an explanation; (2) issues in combining the cues-to-causality; (3) problems in defining the causal background; and, (4) some normative questions in assessing the quality of causal judgments.

Discounting Explanations

The idea that alternatives reduce the causal strength of a hypothesis has been amply demonstrated by many (cf. Kelley's "discounting principle," 1973; Schustack & Sternberg, 1981). Furthermore, some researchers (e.g., Jones, 1979) have proposed a discounting process in which one anchors on a hypothesis

TABLE 5

Mean Ratings of Causal Strength Before and
After Weak vs. Strong Alternatives
in Experiment 2B

	Before $s_0(X)$	After s_1	Prediction* $\hat{s}_1$
Weak Alternative			
Scenario 1	.45	.43	.45
Scenario 2	.64	.61	.64
Strong Alternative			
Scenario 1	.59	.43	.38
Scenario 2	.65	.42	.41

*For strong alternatives, $\alpha = 1.25$, $C = .41$. For weak alternatives, $C = 0$ by assumption.

TABLE 6

Effects of Shifts in the Causal Background

		Scenario 1			
		Force of hammer	Defect in glass	Other explanations	
Scenario 2	Force of hammer	23	0	0	23
	Defect in glass	32	2	2	36
	Other explanations	5	0	2	7
		60	2	4	66

Note: One subject responded to only one version of the stimulus and is therefore excluded from the analysis.

and adjusts for the plausibility of alternatives. While we are in obvious agreement with this position, without further elaboration, it begs the questions of how, and how much, the plausibility of explanations affects the adjustment process. We believe that our theory takes a first step toward answering these questions. Indeed, our model states that the effects of plausibility are complex since the size of the adjustment depends on three factors: the gross strengths of alternatives, the gross strength of the hypothesis, and the weight given to disconfirmatory evidence (α). Furthermore, by specifying the dynamics of the adjustment process and incorporating them in a simple quantitative model, we were able to make testable predictions of the amount of discounting.

We now consider some implications and extensions of the anchor-and-adjust model. First, we interpreted the parameter α as reflecting the weight given to alternatives in the adjustment process. Moreover, in our experiments, $\alpha > 1$, thereby implying that the initial gross strengths of the hypotheses were not greatly discounted by alternatives. While we are tempted to explain this as being consistent with much psychological research on the underweighting of disconfirmatory data (e.g., Ross & Lepper, 1981) and the lack of search for disconfirming hypotheses (e.g., Mynatt, et al. 1977, 1978; Tweney, et al., 1980), we stress that a systematic research program is needed to examine the determinants of α . At the very least, our approach suggests that α can serve as a quantitative and interpretable dependent variable for studying such factors as individual differences, expertise about the substantive content of the scenario, set, and so on. Second, there are a number of "procedural" variables that can be studied via our model (cf. Lopes, 1983). For example, order of hypothesis presentation, simultaneous vs. sequential display of information, and the like, may affect final net strength. While a discussion

of these is beyond the scope of this paper, we note that our model predicts order and presentation mode effects under certain conditions. Third, we have assumed that alternatives always discount an explanation (or leave it unchanged), although it is possible for an alternative to increase net strength. For example, imagine that you believe strongly in a scientific theory for which there are few competitors. You are then presented with an absurd alternative explanation which leads to the following inference: if this is the best alternative that people can generate, your belief in the original theory should be increased. Our model might be extended to handle such effects by allowing w_{k-1} to be negative in equation (3). In any event, this type of complex inference, as well as the procedural effects discussed above, have not yet been put to experimental tests. However, they illustrate the richness of anchor-and-adjust strategies in inference (cf. Lopes, 1982b), and the importance of a dynamic perspective in building descriptive models of the judgment process (Hogarth, 1981).

Although we have concentrated on the role of alternative explanations in discounting a hypothesis, it is important to note that there is a constructive aspect to diagnostic inference. That is, the ultimate purpose of such inference is to generate some causal explanation for observed effects. Thus, while a particular explanation may be judged as inadequate after it is discounted by alternatives, this does not mean that the diagnostic process terminates at this point. Indeed, one is still left with the question, "If it wasn't X, what did cause Y?" Therefore, while the testing of hypotheses via comparison with alternatives is part of diagnostic inference, the latter also involves a continuing search for better explanations. The distinction between testing hypotheses and searching for better ones can be likened to a "disconfirmation" vs. "replacement" model of inference. Indeed, the replace-

ment view is consistent with the Kuhnian notion that theories in science are not discarded, despite evidence to the contrary, if they are not replaced by better alternatives (Kuhn, 1962). We believe that the replacement view is equally strong in everyday inference. A useful analogy might be the following: how many people would read detective stories if the author only revealed who didn't do it?

Combining the Cues-to-Causality

By considering causal judgments as resulting from the weighting and combining of cues-to-causality, our theory directs attention to issues that might otherwise be ignored. We now consider some of these: (a) What are the ecological validities of the cues? - It has been assumed throughout this paper that the cues-to-causality have imperfect but non-zero ecological validities; i.e., each cue is predictive of a true causal relation. How do we know this? Simply put, we don't. The reason is that without some measure of "true" causality, no determination of accurate causal knowledge is strictly possible. However, the fact that the cues we have considered are implicated in a wide variety of studies with both human and animal subjects, leads us to believe that they would not continue to be used if they were useless. Therefore, our argument is a functional and practical one; viz., given the importance of learning and inferring causal relations for survival, we do not believe that the cues on which this depends are totally worthless. On the other hand, we do not advocate the position that if something is used, it must be beneficial for the organism. Such a view is untenable for many reasons (see Einhorn & Hogarth, 1981); (b) What role does cue redundancy (inter-correlation) play in causal judgments? - While we have treated the cues-to-causality as conceptually distinct, it seems likely that they are correlated in the environment. However, the determination of these correlations would

require an elaborate (and problematic) ecological analysis that is beyond the scope of this paper. Nevertheless, the assumption of correlated cues seems warranted since people have strong expectations concerning what cues go together. Indeed, just as in the perception of incomplete figures (where one fills in the missing parts), scenarios are filled in by assuming that cues not explicitly mentioned are in fact present. Thus, the fact that one generally perceives the world as coherent, suggests that the cues-to-causality are redundant to some degree; (c) Reducing inconsistency in causal judgment - Many studies have shown that inconsistency in the execution of judgmental strategies leads to decrements in performance (Hammond, Hirsch, & Todd, 1964; Goldberg, 1970). Thus, to the extent that people are inconsistent in the cognitive strategies used to make causal judgments, it follows that the accuracy of these judgments will be reduced. This observation assumes, of course, that there is some ecological criterion of causality. However, in the absence of a measurable criterion, the mechanical combining of the cues suggests the possibility of improving diagnostic judgment via a "bootstrapping" model in the same way as has been demonstrated in predictive judgment (see e.g., Dawes, 1971).

Role of the Causal Background

The most important implication of the causal background is that causal strength is not a thing-in-itself, but rather a relation between factors. Thus, we believe that the essential role of context in assessing causal relations makes the search for a purely structural definition of cause difficult. To be sure, equations (1) - (9) attempt to provide such a structure, but they are limited by the lack of specificity as to what constitutes the background in any given situation. In one sense, this vagueness can be seen as a positive attribute in that it reflects the corresponding

vagueness that people have about the assumptions that underlie their causal judgments. On the other hand, it highlights the need for a theory of how, when, and why, particular backgrounds are invoked. The components of such a theory will not only need to consider how expectations affect inferences, but also how expectations change with shifts in the background. Furthermore, the role of prior knowledge in both conditioning the cues and in providing a meaningful context for understanding causal connections must be developed. In this regard we find the concept of a "script" to be a useful way to conceptualize an organized set of expectations (Abelson, 1981). However, much remains to be done in linking scripts/schemas to the causal background and the assessment of causal strength.

A second important implication of the causal background relates to surprise. Thus, when expectations that rest on an assumed background are violated, this can be an important cue for reorganizing or re-structuring one's hypotheses. For example, imagine a hit-and-run accident in which all 10 witnesses said the offending car was going 73 miles per hour at the moment of impact. Since we expect much greater variability in such estimates, as well as round numbers, this surprising unanimity might cue one to ask whether the witnesses had colluded in their responses. Similarly, the structure of outcomes can suggest new hypotheses such that the diagnosis contradicts the surface meaning of the evidence. Thus, scientific data that are too perfect can suggest fraud (see, for example, Kamin, 1974, on Burt's twin data; Bishop, et al., 1975, on Mendel's pea experiments), evidence in a trial that is too consistent and obvious can suggest the defendant was "framed," and one can "protesteth too much" in a variety of circumstances. Such examples illustrate that violations of expectations can trigger a re-structuring of alternatives. Of course, specifying the conditions that lead to re-structuring as opposed to other responses remains an important and unanswered question.

Normative Problems

Our theoretical analysis raises the following question: since there is no agreed-on theory of causality, is it possible to say anything about the quality of diagnostic inferences and the causal judgments on which they are based? While we have no definitive answers, we discuss two trade-offs that are germane to this question: (1) acquisition of causal knowledge vs. superstition; (2) achieving "order-out-of-chaos" vs. limiting creativity.

Causal Knowledge vs. Superstition

We have asserted that the cues to causality have ecological validity and that accurate causal knowledge depends partly on their use. However, since the cues are imperfectly valid, the discovery of causal relations can be likened to a complex, multivariate signal detection task where the presence of cause is sought against a background of randomness or noise (cf. Lopes, 1982a). There are several implications of this signal detection analogy. First, people must set a cut-off point to decide whether or not some factor is to be considered a cause. Second, the position of this cut-off will reflect two types of errors and their associated costs. That is, on the one hand, people can infer causes when they do not exist; on the other hand, they can make the error of failing to infer true causal relations. Moreover, whereas several studies have addressed the former and discussed human susceptibility to "illusions of control" (e.g., Langer, 1975), there has been less awareness of illusions of lack of control (however, see Seligman, 1975; Alloy & Abramson, 1979). Nonetheless, given the importance of inferring causal relations for survival, one could argue that the former illusion is less costly than the latter. Indeed, one can consider superstition as the price that one pays for causal knowledge (cf. Skinner, 1966), although it is an open question as to whether the price is worth the benefits in any particular

situation. Finally, a third implication of the signal detection analogy is that people may exhibit differential sensitivity to seeing causal relations through either training (e.g., developing expertise), or ability.

In addition to using the cues to causality, our position also implies that accurate causal judgment involves the elimination of alternative explanations. However, as vividly demonstrated by the concept of spurious correlation, several variables may be highly correlated in the natural ecology so that the determination of causal relations is problematic. Thus, from this viewpoint one can sympathize with the ingenuous teenager who asked Dear Abby: would she get pregnant from holding hands with her boy-friend? Given that this causal candidate and the true cause are both correlated and share many of the same cues-to-causality, only a true experiment could resolve the issue. Indeed, the importance of experiments for disentangling correlated factors has been stressed by Hammond (1978). He points out that much learning through experience often rests on the weakest mode of inference--unaided judgment based on passive observation. From a normative viewpoint, the prevalence of correlated alternatives reinforces the need for experimentation in making valid causal inferences (cf. Einhorn & Hogarth, 1978).

Order Out-of-Chaos vs. Creative Thought

The causal field and the cues to causality both play an important role in limiting the number of interpretations people make in inferential tasks, and thus in creating "order-out-of-chaos." Furthermore, the adoption of a particular background and the use of the cues proceed quickly and are often marked by a lack of awareness that a delimiting process has taken place. The benefits to be gained from such automatized processes are large. However, they come at the cost of reducing the probability that people can achieve more creative representations of inferential tasks. Indeed, Campbell (1960) has

stressed the importance of deliberately introducing random variation to stimulate creative efforts, especially in science. Without such random perturbations, he argues that the forces that maintain a person's particular conception of a problem are too strong. Moreover, the literature on creativity has many examples of techniques that are aimed precisely at making people aware of the delimiting assumptions they bring to tasks (see, e.g., Adams, 1976). In addition, when using such techniques, people are often requested to refrain from counterfactual reasoning and to make specific use of analogies and paradox to enjoin previously disconnected ideas. In short, to restructure problems in creative ways frequently requires attempts to counter the habitual forces of causal reasoning.

Conclusion

This paper has emphasized the fundamental role of causal judgments in diagnostic inference and argued that causal judgments are made in relation to a causal background or field; people use multiple, probabilistic cues-to-causality in forming their judgments; and, an explanation is discounted as a function of its initial strength and the plausibility of alternatives. Moreover, these ideas can be summarized by a perceptual analogy in which figures are seen against ground (causal candidates are differences-in-a-background), good figures are consistent with Gestalt principles (good explanations arise from internally consistent patterns of cues), and, good figures have few alternatives (as do good explanations).

Whereas our model accounts for many findings in the literature as well as our own experimental results, it by no means explicates all aspects of causal reasoning. In particular, inferences made on the basis of complex scenarios, the assessment of causal chains, issues of multiple and redundant causation,

etc., present formidable difficulties and challenges for behavioral research. However, given the complexity of these issues, it seems appropriate to have started with a simple model based on alternatives, background, and cues; i.e., the ABC of causal judgment.

Footnotes

This work was supported by a contract from the Office of Naval Research. We would like to thank Zvi Gilula, Marlys Lipe, John Lyons, Haim Mano, Ann McGill, and Werner Wothke for their assistance on this project. In addition, the following people provided us with many useful comments on earlier versions of the paper: Robert Abelson, Berndt Brehmer, Colin Camerer, Norman Dalkey, Don Fiske, five anonymous referees, William Goldstein, Ken Hammond, Joshua Klayman, Howard Kunreuther, Lola Lopes, Al Madansky, John Payne, Jay Russo, Paul Schoemaker, and Arnold Zellner.

¹Since equation (5) is bounded by 0 and 1 and $\alpha > 0$ from equation (4), it can be shown that,

$$\alpha > \max \left[0, 1 - \left\{ \frac{\log s(Y, Z_k | B)}{\log S_{k-1}(Y, X | B)} \right\} \right]$$

This implies that when low anchors are paired with strong alternatives, α is closer to 1 than 0. Such a constraint makes sense, under these circumstances, since weak anchors cannot be discounted to be less than "worthless" by strong alternatives.

²Equation (9) assumes that the cues are measured without error. However, the cue of temporal order could trade-off with other cues if there were doubts about the order in which X and Y occurred. Equation (9) would then become,

$$s(Y, X | B) = \gamma(\lambda_1 Q_1 + \lambda_2 Q_2 + \lambda_3 Q_3 + \lambda_4 Q_4)$$

³Copies of all scenarios, in all experiments, can be obtained from the authors.

⁴Since a Latin square involves an incomplete design, it should be noted that one cannot test all possible interactions. However, since we have no a priori theory regarding interactions, this is a minor limitation of the design. On the other hand, since some interactions can be tested, we chose to examine those of greatest potential importance.

⁵Similarly, using equations (12c) and (12d) to estimate α and C , we obtain $\alpha = 1.47$ and $C = .57$. When these values are used to predict the net strengths of (12a) and (12b), the results are .32 (observed net strength = .37) and .24 (observed net strength = .25), respectively.

References

- Adams, J. L. Conceptual blockbusting: A pleasurable guide to better problem solving. San Francisco: San Francisco Book Co., 1976.
- Abelson, R. P. Psychological status of the script concept. American Psychologist, 1981, 36, 715-729.
- Alloy, L. B., & Abramson, L. Y. Judgment of contingency in depressed and nondepressed students: Sadder but wiser? Journal of Experimental Psychology: General, 1979, 108, 441-485.
- Arkes, H. R., & Harkness, A. R. Estimates of contingency between two dichotomous variables. Journal of Experimental Psychology: General, 1983, 112, 117-135.
- Bar-Hillel, M. The base-rate fallacy in probability judgments. Acta Psychologica, 1980, 44, 211-233.
- Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. Discrete multivariate analysis: Theory and practice. Cambridge, MA: The MIT Press, 1975.
- Bruner, J. S. Going beyond the information given. In J. S. Bruner et al. (Eds.), Contemporary approaches to cognition. Cambridge, MA: Harvard University Press, 1957.
- Bruner, J. S., & Postman, L. On the perception of incongruity: A paradigm. Journal of Personality, 1949, 18, 206-223.
- Brunswik, E. The conceptual framework of psychology. Chicago: University of Chicago Press, 1952.
- Campbell, D. T. Blind variation and selective retention in creative thought as in other knowledge processes. Psychological Review, 1960, 67, 380-400.
- Campbell, D. T. Pattern matching as an essential in distal knowing. In K. R. Hammond (Ed.), The psychology of Egon Brunswik. New York: Holt, Rinehart and Winston, 1966.

- Campbell, D. T. Reforms as experiments. American Psychologist, 1969, 24, 409-429.
- Campbell, D. T. Degrees of freedom and the case study. Comparative Political Studies, 1975, 8, 178-193.
- Campbell, D. T., & Stanley, J. C. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally, 1963.
- Cohen, L. J. The probable and the provable. Oxford: Clarendon Press, 1977.
- Cook, T. D., & Campbell, D. T. Quasi-experimentation: Design and analysis for field settings. Chicago: Rand-McNally, 1979.
- Crocker, J. Judgment of covariation by social perceivers. Psychological Bulletin, 1981, 90, 272-292.
- Crocker, J. Biased questions in judgment of covariation studies. Personality and Social Psychology Bulletin, 1982, 8, 214-220.
- Dawes, R. M. A case study of graduate admissions: Applications of three principles of human decision making. American Psychologist, 1971, 26, 180-188.
- Eddy, D. M., & Clanton, C. H. The art of diagnosis. New England Journal of Medicine, 1982, 306, 1263-1268.
- Ehrenberg, A. S. C. Data reduction: Analyzing and interpreting statistical data. New York: Wiley, 1975.
- Einhorn, H. J., & Hogarth, R. M. Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 1978, 85, 395-416.
- Einhorn, H. J., & Hogarth, R. M. Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 1981, 32, 53-88.
- Einhorn, H. J., & Hogarth, R. M. Prediction, diagnosis, and causal thinking in forecasting. Journal of Forecasting, 1982, 1, 23-36.

- Elstein, A. S., Shulman, L. E., & Sprafka, S. A. Medical problem solving: An analysis of clinical reasoning. Cambridge, Mass: Harvard University Press, 1978.
- Fincham, F. D., & Jaspars, J. M. Attribution of responsibility: From man the scientist to man as lawyer. Advances in Experimental Social Psychology, 1980, 13, 82-138.
- Garcia, J. Tilting at the paper mills of academe. American Psychologist, 1981, 36, 149-158.
- Garcia, J., McGowan, B., Ervin, F. R., & Koelling, R. Cues: Their relative effectiveness as reinforcers. Science, 1968, 160, 794-795.
- Garcia, J., McGowan, B., & Green, K. Sensory quality and integration: Constraints on conditioning. In A. H. Black & W. F. Prokasy (Eds.), Classical conditioning II: Current research and theory. New York: Appleton-Century-Crofts, 1972.
- Garner, W. R. To perceive is to know. American Psychologist, 1966, 21, 11-19.
- Garner, W. R. Good patterns have few alternatives. American Scientist, 1970, 58, 34-42.
- Goldberg, L. R. Man versus model of man: A rationale, plus some evidence for a method of improving on clinical inferences. Psychological Bulletin, 1970, 73, 422-432.
- Hammond, K. R. Probabilistic functionalism and the clinical method. Psychological Review, 1955, 62, 255-262.
- Hammond, K. R. Toward increasing competence of thought in public policy formation. In K. R. Hammond (Ed.), Judgment and decision in public policy formation. Denver, CO: Westview Press, 1978.

- Hammond, K. R., Hursch, C. J., & Todd, F. J. Analyzing the components of clinical inference. Psychological Review, 1964, 71, 438-456.
- Hammond, K. R., McClelland, G. H., & Mumpower, J. Human judgment and decision making. New York: Praeger, 1980.
- Hart, H. L., & Honore, A. M. Causation in the law. Oxford: Clarendon Press, 1959.
- Hogarth, R. M. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. Psychological Bulletin, 1981, 90, 197-217.
- Jenkins, H. M., & Ward, W. C. Judgment of contingency between responses and outcomes. Psychological Monographs: General and Applied, 1965, 79 (Whole No. 594), 1-17.
- Jones, E. E. The rocky road from acts to dispositions. American Psychologist, 1979, 34, 107-117.
- 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.
- Kamin, L. J. The science and politics of IQ. Potomac, MD: Erlbaum, 1974.
- Kelley, H. H. The processes of causal attribution. American Psychologist, 1973, 28, 107-128.
- Kelley, H. H., & Michela, J. L. Attribution theory and research. Annual Review of Psychology, 1980, 31, 457-501.
- Kuhn, T. S. The structure of scientific revolutions. Chicago: University of Chicago Press, 1962.
- Langer, E. J. The illusion of control. Journal of Personality and Social Psychology, 1975, 32, 311-328.

- Lipe, M. G. A cross-study analysis of covariation judgments. Center for Decision Research, Graduate School of Business, University of Chicago, 1982 (April).
- Lipe, M. G. Attribution theory: A proposed model. Center for Decision Research, Graduate School of Business, University of Chicago, 1983 (June).
- Loftus, E. F. Eyewitness testimony. Cambridge, MA: Harvard University Press, 1979.
- Lopes, L. L. Averaging rules and adjustment processes: The role of averaging in inference. Wisconsin Human Information Processing Program, University of Wisconsin, Report 13, 1981 (December).
- Lopes, L. L. Doing the impossible: A note on induction and the experience of randomness. Journal of Experimental Psychology: Learning, Memory, and Cognition, 1982a, 8, 626-636.
- Lopes, L. L. Toward a procedural theory of judgment. Wisconsin Human Information Processing Program, University of Wisconsin, Report 17, 1982b (December).
- Mackie, J. L. The cement of the universe: A study of causation. Oxford: Clarendon Press, 1974.
- McArthur, L. A. The how and what of why: Some determinants and consequences of causal attribution. Journal of Personality and Social Psychology, 1972, 22, 171-193.
- Michotte, A. La perception de la causalité. Paris: Vrin, 1946.
- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. Confirmation bias in a simulated research environment: An experimental study of scientific inference. Quarterly Journal of Experimental Psychology, 1977, 29, 85-95.

- Mynatt, C. R., Doherty, M. E., & Tweney, R. D. Consequences of confirmation and disconfirmation in a simulated research environment. Quarterly Journal of Experimental Psychology, 1978, 30, 395-406.
- Nisbett, R. E., & Ross, L. D. Human inference: Strategies and shortcomings of social judgment. Englewood Cliffs: Prentice-Hall, 1980.
- Piaget, J. Understanding causality. New York: Norton, 1974.
- Ross, L., & Lepper, M. R. The perseverance of beliefs: Empirical and normative considerations. In R. A. Shweder (Ed.), New directions for methodology of social and behavioral science: Fallible judgment in behavioral research. San Francisco: Jossey-Bass, 1980.
- Schustack, M. W., & Sternberg, R. J. Evaluation of evidence in causal inference. Journal of Experimental Psychology: General, 1981, 110, 101-120.
- Sedlak, A. J., & Kurtz, S. T. A review of children's use of causal inference principles. Child Development, 1981, 52, 759-784.
- Seggie, J. L., & Endersby, H. The empirical implications of Piaget's concept of correlation. Australian Journal of Psychology, 1972, 24, 3-8.
- Seligman, M. E. P. On the generality of the laws of learning. Psychological Review, 1970, 77, 406-418.
- Seligman, M. E. P. Helplessness. San Francisco: Freeman, 1975.
- Shapiro, A. K. A contribution to a history of the placebo effect. Behavioral Science, 1960, 5, 109-135.
- Shultz, T. R. Rules of causal attribution. Monographs of the Society for Research in Child Development, 1982, 47, 1-51.
- Shultz, T. R., & Ravinsky, F. B. Similarity as a principle of causal inference. Child Development, 1977, 48, 1552-1558.
- Shweder, R. A. Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. Current Anthropology, 1977, 18, 637-658.

- Siegler, R. S. The effects of simple necessity and sufficiency relationships on children's causal inferences. Child Development, 1976, 47, 1058-1063.
- Siegler, R. S., & Liebert, R. M. Effects of contiguity, regularity, and age on children's causal inferences. Developmental Psychology, 12, 574-579.
- Simon, G. A. Spurious correlation: A causal interpretation. Journal of the American Statistical Association, 1954, 49, 467-479.
- Skinner, B. F. The phylogeny and ontogeny of behavior. Science, 1966, 153, 1205-1213.
- 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.
- Strawson, P. F. Introduction to logical theory. London: Methuen, 1952.
- Suppes, P. A probabilistic theory of causality. Amsterdam: North-Holland, 1970.
- Susser, M. Causal thinking in the health sciences. New York: Oxford University Press, 1973.
- Taylor, H. C., & Russell, J. T. The relationship of validity coefficients to the practical validity of tests in selection: Discussion and tables. Journal of Applied Psychology, 1939, 23, 565-578.
- Toda, M. Causality, conditional probability and control. In A. Aykac & C. Brumat (Eds.), New developments in the application of Bayesian methods. Amsterdam: North-Holland, 1977.
- Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

Tversky, A., & Kahneman, D. Causal schemas in judgments under uncertainty.

In M. Fishbein (Ed.), Progress in social psychology, Vol. 1. Hillside, NJ: Erlbaum, 1980.

Tweney, R. D., Doherty, M. E., Worner, W. J., Pliske, D. B., Mynatt,

C. R., Gross, K. A., & Arkelin, D. L. Strategies of rule discovery in an inference task. Quarterly Journal of Experimental Psychology, 1980, 32, 109-123.

Waller, W. S., & Felix, W. L. The auditor and learning from experience:

Some empirical evidence. Department of Accounting, University of Arizona, 1982 (November).

Ward, W. C., & Jenkins, H. M. The display of information and the judgment of contingency. Canadian Journal of Psychology, 1965, 19, 231-241.

Zellner, A. Causality and econometrics. In K. Brunner & A. H. Meltzer (Eds.), Three aspects of policy and policymaking, Carnegie-Rochester Conference Series, Vol. 10, North-Holland, 1979.

OFFICE OF NAVAL RESEARCH

Engineering Psychology Group

TECHNICAL REPORTS DISTRIBUTION LISTOSD

CAPT Paul R. Chatelier
Office of the Deputy Under Secretary
of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D. C. 20301

Dr. Dennis Leedom
Office of the Deputy Under Secretary
of Defense (C'I)
Pentagon
Washington, D. C. 20301

Department of the Navy

Engineering Psychology Group
Office of Naval Research
Code 442 EP
Arlington, VA 22217 (2 cys.)

Aviation & Aerospace Technology
Programs
Code 210
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Communication & Computer Technology
Programs
Code 240
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Physiology & Neuro Biology Programs
Code 441NB
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Department of the Navy

Tactical Development & Evaluation
Support Programs
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Manpower, Personnel & Training
Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mathematics Group
Code 411-MA
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics and Probability Group
Code 411-S&P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR K. Hull
Code 230B
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Department of the Navy

Special Assistant for Marine Corps
Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Dr. J. Lester
ONR Detachment
495 Summer Street
Boston, MA 02210

Mr. R. Lawson
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

CDR James Offutt, Officer-in-Charge
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D. C. 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, D. C. 20375

Dr. J. S. Lawson
Naval Electronic Systems Command
NELEX-06T
Washington, D. C. 20360

Dr. Robert E. Conley
Office of Chief of Naval Operations
Command and Control
OP-094H
Washington, D. C. 20350

CDR Thomas Berhage
Naval Health Research Center
San Diego, CA 92152

Department-of the Navy

Dr. Robert G. Smith
Office of the Chief of Naval
Operations, OP987H
Personnel Logistics Plans
Washington, D. C. 20350

Dr. Andrew Rechnitzer
Office of the Chief of Naval
Operations, OP 952F
Naval Oceanography Division
Washington, D. C. 20350

Combat Control Systems Department
Code 35
Naval Underwater Systems Center
Newport, RI 02840

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation
Group
Orlando, FL 32813

CDR Norman E. Lane
Code N-7A
Naval Training Equipment Center
Orlando, FL 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. H. Talkington
Ocean Engineering Department
Naval Ocean Systems Center
San Diego, CA 92152

Department of the Navy

Mr. Paul Heckman
Naval Ocean Systems Center
San Diego, CA 92152

Dr. Ross Pepper
Naval Ocean Systems Center
Hawaii Laboratory
P. O. Box 997
Kailua, HI 96734

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D. C. 20380

Dr. L. Chmura
Naval Research Laboratory
Code 7592
Computer Sciences & Systems
Washington, D. C. 20375

HQS, U. S. Marine Corps
ATTN: CCA40 (Major Pennell)
Washington, D. C. 20380

Commanding Officer
MCTSSA
Marine Corps Base
Camp Pendleton, CA 92055

Chief, C³ Division
Development Center
MCDEC
Quantico, VA 22134

Human Factors Technology Administrator
Office of Naval Technology
Code MAT 0722
800 N. Quincy Street
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Programs
NAVAIR 334A
Washington, D. C. 20361

Department of the Navy

Commander
Naval Air Systems Command
Crew Station Design
NAVAIR 5313
Washington, D. C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 03416
Washington, D. C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 81323
Washington, D. C. 20360

Larry Olmstead
Naval Surface Weapons Center
NSWC/DL
Code N-32
Dahlgren, VA 22448

Mr. Milon Essoglou
Naval Facilities Engineering Command
R&D Plans and Programs
Code 03T
Hoffman Building II
Alexandria, VA 22332

Capt. Robert Biersner
Naval Medical R&D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Department of the Navy

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

Commanding Officer
Naval Health Research Center
San Diego, CA 92152

Commander, Naval Air Force,
U. S. Pacific Fleet
ATTN: Dr. James McGrath
Naval Air Station, North Island
San Diego, CA 92135

Navy Personnel Research and
Development Center
Planning & Appraisal Division
San Diego, CA 92152

Dr. Robert Blanchard
Navy Personnel Research and
Development Center
Command and Support Systems
San Diego, CA 92152

CDR J. Funaro
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Stephen Merriman
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

Department of the Navy

Dean of the Academic Departments
U. S. Naval Academy
Annapolis, MD 21402

Dr. S. Schiflett
Human Factors Section
Systems Engineering Test
Directorate
U. S. Naval Air Test Center
Patuxent River, MD 20670

Human Factor Engineering Branch
Naval Ship Research and Development
Center, Annapolis Division
Annapolis, MD 21402

Mr. Harry Crisp
Code N 51
Combat Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Mr. John Quirk
Naval Coastal Systems Laboratory
Code 712
Panama City, FL 32401

CDR C. Hutchins
Code 55
Naval Postgraduate School
Monterey, CA 93940

Office of the Chief of Naval
Operations (OP-115)
Washington, D. C. 20350

Professor Douglas E. Hunter
Defense Intelligence College
Washington, D. C. 20374

Department of the Army

Mr. J. Barber
HQS, Department of the Army
DAPE-MBR
Washington, D. C. 20310

Department of the Navy

Dr. Edgar M. Johnson
 Technical Director
 U. S. Army Research Institute
 5001 Eisenhower Avenue
 Alexandria, VA 22333

Director, Organizations and
 Systems Research Laboratory
 U. S. Army Research Institute
 5001 Eisenhower Avenue
 Alexandria, VA 22333

Technical Director
 U. S. Army Human Engineering Labs
 Aberdeen Proving Ground, MD 21005

Department of the Air Force

U. S. Air Force Office of Scientific
 Research
 Life Sciences Directorate, NL
 Bolling Air Force Base
 Washington, D. C. 20332

AFHRL/LRS TDC
 Attn: Susan Ewing
 Wright-Patterson AFB, OH 45433

Chief, Systems Engineering Branch
 Human Engineering Division
 USAF AMRL/HES
 Wright-Patterson AFB, OH 45433

Dr. Earl Alluisi
 Chief Scientist
 AFHRL/CCN
 Brooks Air Force Base, TX 78235

Foreign Addressees

Dr. Daniel Kahneman
 University of British Columbia
 Department of Psychology
 Vancouver, BC V6T 1W5
 Canada

Foreign Addressees

Dr. Kenneth Gardner
 Applied Psychology Unit
 Admiralty Marine Technology
 Establishment
 Teddington, Middlesex TW11 0LN
 England

Director, Human Factors Wing
 Defence & Civil Institute of
 Environmental Medicine
 Post Office Box 2000
 Downsview, Ontario M3M 3B9
 Canada

Dr. A. D. Baddeley
 Director, Applied Psychology Unit
 Medical Research Council
 15 Chaucer Road
 Cambridge, CB2 2EF England

Other Government Agencies

Defense Technical Information Center
 Cameron Station, Bldg. 5
 Alexandria, VA 22314 (12 copies)

Dr. Craig Fields
 Director, System Sciences Office
 Defense Advanced Research Projects
 Agency
 1400 Wilson Blvd.
 Arlington, VA 22209

Dr. M. Montemerlo
 Human Factors & Simulation
 Technology, RTE-6
 NASA HQS
 Washington, D. C. 20546

Dr. J. Miller
 Florida Institute of Oceanography
 University of South Florida
 St. Petersburg, FL 33701

Other Organizations

Dr. Robert R. Mackie
Human Factors Research Division
Canyon Research Group
5775 Dawson Avenue
Goleta, CA 93017

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. H. McI. Parsons
Human Resources Research Office
300 N. Washington Street
Alexandria, VA 22314

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22311

Professor Howard Raiffa
Graduate School of Business
Administration
Harvard University
Boston, MA 02163

Dr. T. B. Sheridan
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Arthur I. Siegel
Applied Psychological Services, Inc.
404 East Lancaster Street
Wayne, PA 19087

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Harry Snyder
Department of Industrial Engineering
Virginia Polytechnic Institute and
State University
Blacksburg, VA 24061

Other Organizations

Dr. Ralph Dusek
Administrative Officer
Scientific Affairs Office
American Psychological Association
1200 17th Street, N. W.
Washington, D. C. 20036

Dr. Robert T. Hennessy
NAS - National Research Council (COHF)
2101 Constitution Avenue, N. W.
Washington, D. C. 20418

Dr. Amos Freedy
Perceptrics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert C. Williges
Department of Industrial Engineering
and OR
Virginia Polytechnic Institute and
State University
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 17th Street, N. W.
Washington, D. C. 20036

Dr. Deborah Boehm-Davis
General Electric Company
Information Systems Programs
1755 Jefferson Davis Highway
Arlington, VA 22202

Dr. Ward Edwards
Director, Social Science Research
Institute
University of Southern California
Los Angeles, CA 90007

Dr. Robert Fox
Department of Psychology
Vanderbilt University
Nashville, TN 37240

Other Organizations

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Boulder, CO 80309

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, D. C. 20064

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Christopher Wickens
Department of Psychology
University of Illinois
Urbana, IL 61801

Mr. Edward M. Connelly
Performance Measurement
Associates, Inc.
410 Pine Street, S. E.
Suite 300
Vienna, VA 22180

Professor Michael Athans
Room 35-406
Massachusetts Institute of
Technology
Cambridge, MA 02139

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Other Organizations

Dr. Babur M. Pulat
Department of Industrial Engineering
North Carolina A&T State University
Greensboro, NC 27411

Dr. Lola Lopes
Information Sciences Division
Department of Psychology
University of Wisconsin
Madison, WI 53706

Dr. A. K. Bejczy
Jet Propulsion Laboratory
California Institute of Technology
Pasadena, CA 91125

Dr. Stanley N. Roscoe
New Mexico State University
Box 5095
Las Cruces, NM 88003

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 01803

Dr. Marvin Cohen
Decision Science Consortium
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. Wayne Zachary
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Dr. William R. Uttal
Institute for Social Research
University of Michigan
Ann Arbor, MI 48109

Dr. William B. Rouse
School of Industrial and Systems
Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Other Organizations

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Hillel Einhorn
Graduate School of Business
University of Chicago
1101 E. 58th Street
Chicago, IL 60637

Dr. Douglas Towne
University of Southern California
Behavioral Technology Laboratory
3716 S. Hope Street
Los Angeles, CA 90007

Dr. David J. Getty
Bolt Beranek & Newman, Inc.
50 Moulton street
Cambridge, MA 02238

Dr. John Payne
Graduate School of Business
Administration
Duke University
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
School of Engineering and
Applied Science
University of Virginia
Charlottesville, VA 22901

Denise Benel
Essex Corporation
333 N. Fairfax Street
Alexandria, VA 22314

Psychological Documents (3 copies)
ATTN: Dr. J. G. Darley
N 565 Elliott Hall
University of Minnesota
Minneapolis, MN 55455

ATE
LMED
-8