Behavior and Philosophy, *30*, 101-122 (2002). © 2002 Cambridge Center for Behavioral Studies

RECONSTRUCTION OF PAST EVENTS FROM MEMORY: AN ALTERNATIVE TO THE HYPOTHETICO-DEDUCTIVE (H-D) METHOD

Sam S. Rakover
Department of Psychology, Haifa University, Haifa, Israel

ABSTRACT: According to the demand of the Hypothetico-Deductive (H-D) method, a theory is confirmed when the prediction-observation (p-o) gap is small and disconfirmed when the gap is large. A major goal of this paper is to introduce a research domain for which this demand does not hold. In contrast to the H-D method's demand, this research, called the Catch model for reconstructing a face previously seen from memory, requires an increase, within limits, in the p-o gap. The Catch model research substantiates theoretically and empirically a new proposed method that I call the "Deductive-Reconstruction" (D-R) method. This method provides essential conditions whose fulfillment guarantees successful reconstruction of past events (a face previously seen) from memory. It is argued that the D-R method fits the area of research of reconstructing past events from memory better than the H-D method. Application of the H-D method to the Catch model's research domain leads to an internal contradiction and failure to reconstruct past events (a face previously seen) from memory. Finally, the nature of the D-R method along with the Catch model is discussed from three points of view: confirmation, explanation, and generality.
Key words: Reconstruction of past events, memory, Hypothetico-Deductive method, Deductive-Reconstruction method.

The Prediction-Observation (p-o) Gap and the Hypothetico-Deductive (H-D) Method

A major and natural requirement in scientific methodology is a decrease in the prediction-observation (p-o) gap. According to this, one tests a theory by comparing a prediction with an observation. The higher the correspondence between what is implied from the theory and the actual observation, or the smaller the p-o gap, the greater the confidence in the theory's efficiency. If this gap is wide the theory is disconfirmed, and we change it or replace it with a new and better one. I shall call this the "p-o decrease demand."

The p-o decrease demand is empirically and theoretically an entrenched and fundamental demand of the natural and the social sciences. It appears as an essential component in many scientific methods, such as the Hypothetico-Deductive (H-D) method (see Glymour, 1980; Lipton, 1991; Salmon, 1967),

---

AUTHOR'S NOTE: I am grateful to G. Hon, J.E.R. Staddon (editor), and two anonymous reviewers, who read an earlier version of the manuscript and made many important comments and suggestions.

Please address all correspondence to Sam S. Rakover, Ph.D., Department of Psychology, University of Haifa, Haifa 31905 Israel. Email: rakover@psy.haifa.ac.il.

Abduction, which is also called Inference to the Best Explanation (e.g., Josephson & Josephson, 1994; Lipton, 1991), in the customary statistical methods used by scientists, and in the "error-statistical philosophy of experiment" that is based on these statistical methods (see Mayo, 1996).

Of all these various methods, which are founded on the p-o decrease demand, I will focus on the H-D method for the following two main reasons. First, the method is considered by many scientists as one of the most important procedures for testing empirical hypotheses and theories (see Glymour, 1980; Lipton, 1991; Salmon, 1967). Glymour (1980) writes,

> "Despite the inability of the hypothetico-deductive account to explain evidential relevance, . . . , [it] remains today one of the most popular. The reason, I think, is that it is so obviously the correct account of a great deal of the history of science." (p. 47-48)

Salmon (1967) summarizes the H-D method schematically as follows:

> "From a general hypothesis and particular statements of initial conditions, a particular predictive statement is deduced. The statements of initial conditions, at least for the time, are accepted as true; the hypothesis is the statement whose truth is at issue. By observation we determine whether the predictive statement turned out to be true. If the predictive consequence is false, the hypothesis is disconfirmed. If observation reveals that the predictive statement is true, we say that the hypothesis is confirmed to some extent." (p. 18)

The second reason for focusing on the H-D method is that psychologists have adopted this method, which has become the fashion in psychology. As an example of the use of the H-D method, let us consider two attempts to explain the "facial inversion effect." This effect relates to the finding that recognition of inverted faces (chin above, hair below) is less accurate (by about 25%) than recognition of upright faces. Furthermore, inversion impairs recognition of faces more than of objects such as houses and landscapes (e.g., Diamond & Carey, 1986; Rakover & Teucher, 1997; Valentine, 1988). Diamond and Carey (1986) proposed a memory theory according to which recognition of upright faces is based on both featural and configurational information (where featural information is about individual features and configurational information is about the spatial relations among the features. See Rakover, 2002). Inversion impairs configurational information, which in turn decreases face recognition. Rakover and Teucher (1997) tested Diamond and Carey's theory by deducing a prediction that inversion should not impair recognition of isolated facial features such as eyes, nose, and mouth. The results disconfirmed this theory (the p-o gap was too wide): inversion did impair recognition of isolated features. The results were explained by an alternative theory, which I call the "facial schemas," which does predict the impairment of recognition of inverted isolated facial features.

A Critique of the H-D Method: A Research Area That Requires an Increase (Within Limits) of the P-O Gap

The H-D method is not perfect and requires supplementary knowledge and procedures such as auxiliary hypotheses, background theory, logical developments, and probability theory (Bayesian inference) to determine the acceptance/rejection of a hypothesis, model, or theory (See Grimes, 1990; Kuhn, 1970; Popper, 1972; Rakover, 1990; Salmon, 1967). Furthermore, Glymour (1980) argues that the H-D method has problems determining the relevance between theory and evidence. The method cannot determine to which part of a theory (in conjunction with auxiliary hypotheses) the observation is related: the H-D method confirms the entire theory with all its components, and it cannot distinguish and disconfirm a particular part of the theory (see also Lipton, 1991; Rakover, 1990). Suppe (1998) claims that the H-D method does not correctly explain the actual scientific work—the structure of a scientific paper reporting experimental results.

Although these criticisms are important, I will not dwell on them since my goal is to criticize the H-D method by concentrating on its fundamental component—the p-o decrease demand. I will do this through discussing an empirical research area aimed at reconstructing a face from memory, called the "Catch model" (see Rakover & Cahlon, 1989, 1999, 2001). The success of the Catch model research is based not so much on fulfilling the p-o decrease demand, but rather on allowing, within limits, the increase of the p-o gap, that is, on fulfilling the "p-o increase (within limits) demand." The qualification "within limits" (to be specified below) means that one cannot increase the p-o gap infinitely. Without this restriction, one can reach the absurdity of predicting an event the size of a ping-pong ball and be satisfied by observing an event the size of the sun. The Catch model research belongs to the domain of cognitive psychology that attempts to accurately reconstruct past events (PE) from memory. For example, given an eyewitness' response and an appropriate theory, the police attempt to reconstruct a picture of the suspect's face as seen in the past by the eyewitness.

Hence, for the first time we seem to have a case against the H-D method's p-o decrease demand. This method does not fit all possible areas of research in memory. There is an area of research in memory that requires not the customary p-o decrease, but rather the p-o increase (within limits). This area deals with reconstruction of PE from memory.

The P-O Increase (Within Limits) Demand and the Deductive-Reconstruction (D-R) Method

The above two examples, the facial inversion effect and eyewitness, raise an important question. Do these two examples belong to two different areas of research that require two different research methodologies? My answer is in the affirmative. I distinguish two general areas of research in memory. While the "remembering area" handles the phenomena of remembering PE, the "reconstructing area" handles the phenomena of reconstructing PE from memory. The methodology connected with the former area of research, the "explanation-testing methodology," uses procedures for testing a memory theory, and for predicting, and explaining, a memory phenomenon, that is, the discrepancies between PE and the remembered PE. The methodology connected with the latter area of research, the "reconstructing methodology," uses a procedure for accurately reconstructing PE from memory. While the explanation-testing methodology employs the H-D method (the most popular and successful one), the reconstructing methodology ought to employ a different research method, which I call the "Deductive-Reconstruction" (D-R) method. The H-D method is based on the methodological requirement of p-o decrease, whereas the D-R method is based on the methodological requirement of p-o increase up to a behavioral limit, which I call the "maximum errors allowable" (MEA) (see below). According to the p-o decrease demand, the greater the decrease in the p-o gap, the greater the efficiency of a theory in predicting, and explaining, the events observed. However, according to the p-o increase (within limits) demand, the greater the increase, the greater the strength of a theory and of its robustness in withstanding errors that interfere with reconstruction of PE. In sum, I propose that the methodology relevant to the remembering area is that which emphasizes theory efficiency, whereas the methodology relevant to the reconstruction area is that which emphasizes theory robustness.

The connection between research areas and methodologies can be presented schematically in the following way:

1) The "remembering area" is connected to the "explanation-testing methodology": Remembering = f(PE, memory theory).

2) The "reconstructing area" is connected to the "reconstructing methodology": PE = g(performance, reconstruction theory).

The goals of type 1 methodology, which is the prevalent in experimental psychology, are to explain a phenomenon and to test a theory. We attempt to predict and explain remembering by appeal to past events (PE) and to a memory theory or model. One of the most popular and successful models of explanation is the Deductive-Nomological (D-N) developed by Hempel (1965). This methodology is also used to test a theory. One of the most popular and successful methods of theory testing is the H-D method. The facial inversion effect, described above, exemplifies the employment of type 1 methodology: the effect has been explanatorily approached by two possible hypotheses—the configurational and the facial schemas.

The goal of type 2 methodology is to accurately reconstruct PE. We attempt to reconstruct PE from memory by appeal to a participant's performance in a particular task and a reconstruction theory or model that consists of two rules:

a) the "task-rule" that determines how the particular task should be performed,
b) the "analysis-rule" that determines how to analyze the task-rule's output.

The method to be used by type 2 methodology is the proposed D-R method that describes the conditions for a successful reconstruction of PE from memory.

Rakover and Cahlon (1989, 1999, 2001) reported the Catch model and several experiments that exemplified the employment of type 2 methodology. Their procedure has two stages. In the study stage, a participant is presented with a picture of a face—the target face. In the test stage the participant is presented with a series of pairs of test faces, one pair at a time. The participant's task—a similarity judgment task—is to choose the face that most resembles the remembered target face. Their model sets two rules: a task-rule, called similarity-rule, that determines which test face is most similar to the target face, and an analysis-rule, called reconstruction-rule, that determines how to analyze the similarity choices in a way that leads to the reconstruction of the target face previously seen.

In the present paper I will show that reconstruction of the target face from memory is guaranteed when the following three conditions, which constitute the D-R method, are met:

A) *The Provability Condition*: It has to be proven that the target face is successfully reconstructed given the reconstruction theory that consists of the similarity and reconstruction rules.

B) *The Matching Condition*: If a participant's choices in the similarity judgment task match the choices produced by the similarity-rule, then the target face is actually reconstructed from the participant's memory.

C) *The Error-Robustness Condition*: It is required that the target face will be reconstructed from memory despite many errors, that is, mismatches between predictions (the choices generated by the similarity-rule) and observations (the participant's choices). The greater the p-o increase (within limits), the greater the robustness in reconstructing the target face. Hence, the behavioral limit is based on MEA.

The Error-Robustness condition can be apprehended by the following analogy. Reconstruction of a face is a goal to be achieved, and MEA is the quality required from the tools for achieving that goal: endurance of high interfering pressures, that is, withstanding many errors.

The concept of error described in condition C is different from the customary one. According to the latter concept, errors are conceived of as measurement inaccuracies distributed randomly around an assumed true value, whereas according to the former concept, errors are conceived as a behavioral criterion, a limit, which answers the following question. How many errors (defined in terms of the mismatches between the choices generated by the similarity-rule and the participant's choices) are allowed to be made without harming a successful reconstruction of the target face? Clearly, one does not want the reconstruction to fail because of one or two errors. This would make the reconstruction method entirely inefficient. Hence, it is required that the limit be based on as many errors as possible, that is, on MEA.

Is the D-R Method Really Needed?

While the H-D method's p-o decrease demand is accepted universally almost axiomatically, the D-R method's p-o increase (within limits) demand is counter-intuitive and hard to justify. Hence, the following question arises: Can one apply the H-D method to the reconstruction area and reach the goal of the D-R method? My answer is no. First, to the best of my knowledge, there is no memory theory that assures reconstruction of PE. Second, although there are very few instances where remembering is without errors, in general, the explanation-testing methodology is designed to deal with the phenomena of forgetting, errors, distortions and false memories. Hence, this methodology cannot ensure the reconstruction of accurate PE. Rather its main goal is to explain why memory of PE is impaired and is not perfect. Finally, it will be shown that the application of the H-D method to the domain of reconstruction of PE leads to internal contradictions and failure to reconstruct the target face from memory. If the Catch model research fulfills the p-o decrease demand, then it will fail in reconstruction of PE—the target face previously seen. However, if this research fulfills, within limits, the p-o increase demand, then it will be successful. It will be demonstrated that there are cases in which the H-D method accepts a theory that does not lead to reconstruction of PE, whereas the D-R method rejects this very same theory, and vice versa.

The Plan of the Paper

To ensure clear presentation of the ideas herein, the paper is organized as follows. First, I will present a brief description of the Catch model for identifying a face previously seen, developed by Rakover and Cahlon (1989, 1999, 2001). Second, I will propose the D-R method and show that the Catch model substantiates and supports the D-R method both theoretically and empirically. Third, I will show that in certain cases, application of the D-R and the H-D methods to the reconstruction area can lead to different consequences: while the H-D method can cause an internal contradiction and reconstruction failure, the D-R method does not cause such consequences. Finally, I will briefly discuss the nature of the D-R method and the Catch model from three points of view: confirmation, explanation, and generality.

The Catch Model: A Brief Summary

Research done with the customary Photofit and the Identikit techniques, used by police for identifying faces, has shown that their composites bear little resemblance to the target face (e.g., Bruce, 1988; Davies, 1981, 1986; Kovera, Penrod, Pappas, & Thill, 1997). This can be attributed to two main factors.

First, while a witness perceived and processed a face as a whole, Identikit, as well as Photofit, tries to reconstruct the target face by recalling parts of the face, that is, isolated facial values such as narrow, round eyes, and long, short, or wide noses. I believe that the human brain is able to decompose the memory of a whole face into its parts, but this process is achieved at the considerable cost of a significant rise in memory errors. Second, the use of many samples of facial values to locate the specific facial value most similar to the target face causes serious interference in the process of retrieving information from the witness's memory (e.g., Farah, 1992; Rakover & Cahlon, 1989, 1999, 2001; Tanaka & Farah, 1993). For example, Rakover and Cahlon (2001) have found that as the number of facial values increases, the accuracy in the reconstruction of a face previously seen decreases.

Given the above considerations, we (Rakover and Cahlon) decided to develop a new model and experimental procedure for identifying the target face by using recognition memory of a whole face. The Catch model is a mathematical model designed to identify the target face. It does this by the reconstruction of the target face from its features (e.g., hair and forehead, eyes, nose, mouth, and chin). These features are identified **after** the subject decides which of two whole test faces is more similar to the remembered target face, as part of a two-stage experiment. In the first stage, the study stage, the subject is presented with a picture of a face—the target face—for about twenty seconds. In the test stage the subject is presented with a series of pairs of test faces, one on the left and one on the right. Neither of the test faces is the target. At each trial the subject is presented with one test pair, and s/he is required to choose the face that most resembles the remembered target face. The subject's choices are then analyzed by what I call the "analysis-rule," which leads to the reconstruction of the target face.

A face is defined as a vector of facial values (e.g., small *mouth*, blue *eyes*), where each value belongs to a different facial dimension (e.g., *mouth*, *eyes*). In the examples in the parentheses, the italicized words stand for facial dimensions, which in conjunction with their adjectives represent their values. An illustration of a division of a face into five dimensions and three values per dimension is depicted as follows:

A. Hair and Forehead	B. Eyes	C. Noses	D. Mouths	E. Chins
a_1) completely bald	b_1) slanting	c_1) long	d_1) wide	e_1) pointed
a_2) medium	b_2) narrow	c_2) medium	d_2) medium	e_2) rounded
a_3) with hair	b_3) round	c_3) short	d_3) small	e_3) squared

In this example, there are $3^5 = 243$ possible faces. There are five dimensions, each having three different values: $a_i\ b_j\ c_k d_m e_n$ (i,j,k,m,n = 1,2,3). Only one of the 243 facial composites is the target face (F_t) and the rest are test faces (F). An example of a target face is: F_t: $a_1\ b_1\ c_1\ d_1\ e_1$, or for the sake of brevity, F_t: 1 1 1 1 1, (e.g., F_t: completely bald, slanting eyes, long nose, wide mouth, and pointed chin); and an example of a test pair consisting of two test faces is F_L: 1 1 1 2 3 and F_R: 1 2 3 3 3 (where L stands for the left side of a test pair and R stands for the right side).

We assume that F_t is represented in memory as a gestalt, together with background stimulation and the knowledge of when and where F_t was perceived. This information is processed automatically, and can also be partially processed in various ways, depending on the subject's intentions and goals, and the requirements of the memory task. We further assume that F is identified as F_t in terms of cognitive memory processes involving similarity judgments. The values of F are compared with the values of F_t represented in memory, and the level of similarity between F and F_t is determined in terms of the number of their common values. If the similarity is maximal, then F is identified as F_t. This assumption led to the model's task-rule—the similarity-rule—suitable for the similarity judgment task:

a) Each F is compared with F_t to see if it has the same values as F_t.

b) For each F the number of "matches," that is, the number of values common to both F_t and a test face F, is determined. We denote the number of "matches" by $\mu(W_F)$, where W_F is the vector associated with F as follows:

$$(W_F)_i = \begin{cases} 1, \text{if } (F_t)_i = (F)_i \\ 0, \text{otherwise} \end{cases}, \text{ and } \mu(W_F) = \sum_{i=1}^{k} (W_F)_i \,.$$

($\mu(W_F)$ constitutes an index of similarity between a test face and F_t.)

c) The F with the highest number of "matches" or $\mu(W_F)$ is selected.

For example, if the target face is F_t: 1 1 1 1 1, and a test pair is F_L: 1 1 2 3 1, F_R: 2 3 2 1 3, then the Catch model will select F_L, since $\left(\mu(W_{F_L}) = 3\right) > \left(\mu(W_{F_R}) = 1\right)$.

The task-rule (the similarity-rule) described here is typical of many similarity-rules reported in the literature (e.g., Estes, 1994; Medin & Schaffer, 1978; Nosofsky, 1984; Tversky, 1977. For reviews and discussions see, e.g., Melara, 1992; Nosofsky, 1992). Of all these similarity-rules, the one developed by Tversky (1977) is the most relevant in the present case. Despite many differences between what the Catch model represents and what Tversky's contrast model represents, under certain conditions it can be shown that the present similarity-rule is a particular case of the contrast model. (Note that the development of the present similarity-rule occurred before Rakover & Cahlon became aware of Tversky's important work.)

Given the choice data, F_t is identified by using the analysis-rule—the reconstruction-rule:

a) For each test pair, the values that appear in the chosen F, and do not appear in the nonchosen F, are recorded. We shall call these values the *differentiating values*. For example, if the test pair is F_L: 1 1 1 2 3 and F_R: 1 2 3 3 3

$$\begin{pmatrix} 11123 \\ 12333 \end{pmatrix}$$

and F_L is chosen, then the *differentiating values* are: —1 1 2 - , where a dash signifies values that do not differentiate between F_L and F_R ;

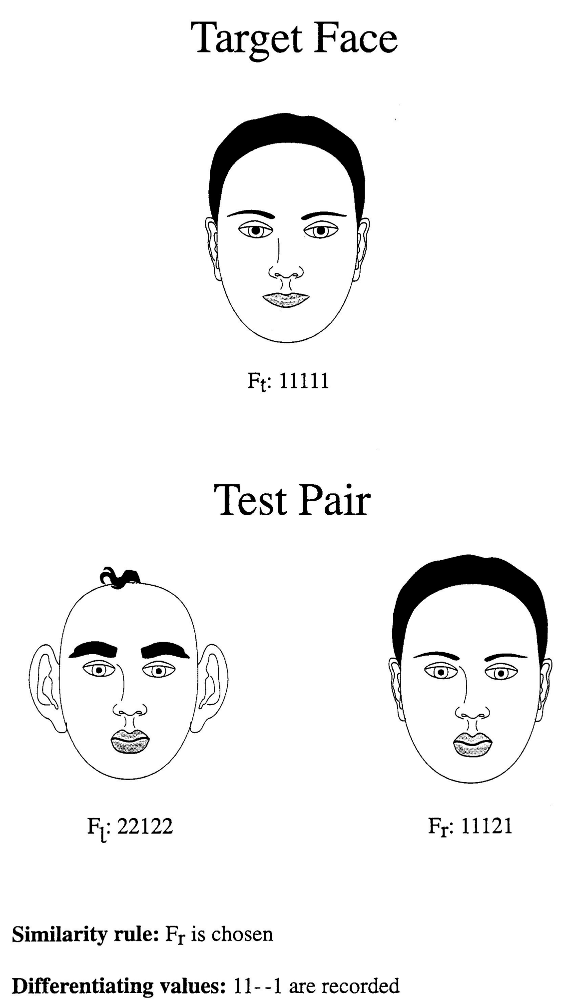
b) For each facial dimension, across several choice trials, the one *differentiating value* chosen most frequently by the subject (i.e., which is associated with the highest *frequency of choice*) is selected. Consider, for example, facial dimension (A). If a_1 is associated with the *frequency of choice* of 100, a_2 with 75, and a_3 with 25, then a_1 is selected as the expected F_t value; and

c) F_t is reconstructed by using the values selected in (b).

In most of our experiments we used Penry's (1971a, 1971b) Photofit Kit to compose various faces. Facial composites were made of five facial dimensions and n facial values per dimension. A special computer program was developed to present these composites on a PC screen.

As an illustration of the Catch model, let us examine Figure 1.

Figure 1: A Target face and a Test pair.



The faces that appear in Figure 1 are 3 of 32 faces, which may be composed of five facial dimensions, each of these having two values:

a) **Hair**	(1) full	(2) bald with a curl
b) **Eyebrows**	(1) normal	(2) thick
c) **Nose**	(1) normal	(2) broad
d) **Mouth**	(1) normal	(2) round
e) **Ears**	(1) normal	(2) big

The figure presents the target face 11111, and below it are two test faces. The test face on the left is 22122 and that on the right is 11121. As may be seen, the Catch model chooses the face on the right because it has four values in common with the target face, while the face on the left has only one value in common with the target face. In keeping with the analysis-rule, the model registers 1 1 - - 1 as the differentiating values, where the dash "-" signifies values common to the two test faces.

The research program of the Catch model deals with many interesting problems, such as the following.

1) *The "Equal pairs" problem*: The model cannot make a similarity decision in cases where the number of matches between features of the left test face and the F_t is equal to the number of matches between features of the right test face and the F_t (e.g., F_t : 11111; test pair: F_L : 11323 and F_R : 21331). We (Rakover and Cahlon) call these pairs "Equal pairs" (note that "Unequal pairs" are those pairs for which the model can make a similarity decision). We solve this problem by showing that the target face can be reconstructed despite the occurrence of equal pairs and by solving the "Saliency problem."

2) *The Saliency problem*: The Catch model assigns the same weight (one point) to each facial dimension—a situation which generates the Equal pairs problem and does not correspond to reality. Rakover and Cahlon (1999) propose a solution to this problem by ascribing different weights to different dimensions of the target face.

3) *The problem of the number of values (n) per dimension*: As mentioned above, we used Penry's (1971a, 1971b) Photofit Kit in our experiments. Composites were made of five facial dimensions and with n = 2,3,6, and 9 facial values per dimension. We found that as n increased, correct reconstruction of the target face tended to decrease. This decrease is due to a retroactive process, since the test faces were composed of a greater number of facial values. We dealt with this retroactive interference effect in a number of ways. For example, we expanded the analysis-rule to include the nonchosen test faces, and we categorized the facial values in terms of their shared visual similarity, using a Multi-Dimensional Scaling technique. As a result, we obtained several similarity categories per dimension. (This categorization was done independently and before the reconstruction of F_t.) Using these similarity categories, we were able to successfully reconstruct a group of faces similar to F_t (i.e., a set of facial vectors similar to the vector of F_t). All this required continuous experimentation and the development of special computer programs for generation of faces, test pairs, and the sampling of a small number of test pairs from all possible pairs (see Rakover & Cahlon, 2001).

The Deductive-Reconstruction (D-R) Method and Its Substantiation by the Catch Model

The D-R Method: Conditions for Reconstruction of Past Events

The purpose of the D-R method is to provide essential conditions for judging whether a given task-rule and analysis-rule will result in reconstruction of past events (such as F_t) from memory. In other words, the D-R method is a procedural schema, in which one inserts task-rule (such as the similarity-rule) and analysis-rule (such as the reconstruction-rule), to determine whether reconstruction of past events from memory can be successful. The D-R method consists of the following three conditions.

First, the provability condition. A mathematical proof that shows that the reconstruction is deduced from the task-rule, which determines how a particular task has to be performed, and a analysis-rule, which determines how the performance should be analyzed. This proof guarantees that past events (e.g., the target face) are indeed reconstructed.

In the Catch model, the task-rule is the similarity-rule which determines for the similarity judgment task which of the two test faces most resembles the remembered target face. The analysis-rule is another specific rule, the reconstruction-rule, which by analyzing the chosen faces into their differentiating values, leads to the reconstruction of the target face.

Second, the matching condition. Given the Catch model, if the subject's choices match the predictions made by the model's task-rule, then past events will be successfully reconstructed from the subject's memory. If the subject's choices are identical to the choices produced by the similarity-rule, then the same target face will also be reconstructed from the subject's memory. Accordingly, replacing the similarity-rule's choices with the subject's choices will render the same final result: reconstruction of the target face.

Third, the "error-robustness" condition. What if the subject's responses do not completely match the task-rule's predictions, that is, if the p-o gap is not zero? If the model requires a full and complete match between the subject's choices and the task-rule's predictions (i.e., choices attained using the similarity-rule), then the reconstruction of the target will fail. Hence, one must ask the following crucial question: to achieve the reconstruction of the target, how many of the subject's choices have to match the choices made by the similarity-rule? How many errors can the subject make and still achieve reconstruction of the target face? (An error is defined in terms of the similarity-rule: e.g., if the similarity-rule determines that the left test face of a test pair is more similar to the target than is the right test face, then an error is defined as the choice of the right test face.) Given this, one would like the reconstruction theory to be Error-Robust, that is, that reconstruction of the target succeed despite the subject's multiple errors. In other words, one would like the theory to withstand a maximum number of errors and still be able to reconstruct the target face. The greater the number of errors sustained, the higher the method's Error-Robustness in achieving reconstruction of the target. I shall call this number of errors the "maximum errors allowable" (MEA), where allowable refers to the maximal number of errors that the model can sustain and still allow successful reconstruction of the target face. Hence, the MEA is the specification of the "within limits" qualification in the p-o increase demand discussed above.

Finally, one has to ask what will happen if a subject's errors are greater than the MEA? The answer is that the reconstruction cannot be achieved. Given the Catch model, the target face cannot be reconstructed from the subject's memory. Hence, one has to look for a new task-rule and a new analysis-rule.

To summarize, the D-R method is based on these three conditions:

1) *The Provability condition*. Past events (the target face) have to be deduced from the task-rule (the similarity-rule), which determines how a particular task (the similarity judgment task) should be performed, and the analysis-rule (the reconstruction-rule), which determines how to analyze the task-rule's output.

2) *The Matching condition*. If a subject's performance in the particular task (the similarity judgment task) matches the output produced by the task-rule, then the target face will actually be reconstructed from the subject's memory.

3) *The Error-Robustness condition*. The reconstruction of the target face is successful if the number of errors is not higher than the MEA.

Given the D-R method's conditions, does the Catch model indeed substantiate the D-R method?

Substantiation of the D-R Method by the Catch Model

Given the Catch model described above, we were able to prove mathematically the following principal conclusions:

1) The chosen test face contains a larger number of differentiating values that belong to F_t than does the rejected face.

2) Given a facial dimension, the differentiating value with the highest frequency of choice is the value belonging to the target face. Namely, the Catch model identifies the target face theoretically. *Hence, the Catch model satisfies the Provability condition.*

3) Even if our sample of test pairs, which is randomly drawn from all possible test pairs, contains only a small number of pairs, it is possible to identify the target face. As the number of test pairs in the sample increases, so does the likelihood of identifying the target face.

4) As the number of subjects (or eyewitnesses) increases, so does the likelihood of identifying the target face.

5) Even if errors are made in the choice of the test faces, namely when the choice of test face is not the choice determined by the model (e.g., if F_t is 11111, and a test pair: F_L is 11231 and F_R is 23312, then an error will be made when F_R is chosen, and not F_L), the target face is identified. (For example, MEA = 31.7% for unequal pairs constructed from four dimensions with two values per dimension.) *Hence, the Catch model satisfies the Error-Robustness condition.*

As conclusions 1-5 indicate, *the Catch model satisfies both the Provability condition and the Error-Robustness condition*.

The Catch model has also proved empirically successful. In several laboratory experiments using a small and random sampling of test pairs, we were able, in many cases, to reconstruct F_t or the group of faces most similar to F_t. Moreover, strong experimental support was found for all the predictions derived from our mathematical propositions and corollaries. In these experiments, the percentage of errors that still allowed identification of F_t reached about 35 percent (for details see Rakover & Cahlon, 1989, 1999, 2001). *Hence, the Catch model has succeeded in satisfying the Matching condition and the Error-Robustness condition.*

In view of this, it is safe to conclude that the Catch model fulfills the above three conditions and therefore substantiates the D-R method.

A Comparison Between the D-R and the H-D Methods

In this section, I will first discuss the differences between these two methods, and then offer responses to two critiques regarding these differences.

Differences

The major difference between the D-R and the H-D methods is as follows. In accordance to the H-D method, a theory is tested in terms of the accuracy of its predictions, that is, the p-o decrease demand; whereas according to the D-R method a theory is tested in terms of its capability to reconstruct past events from memory by fulfilling, within limits, the p-o increase demand. Given this, I shall examine two topics: the issue of error and confirmation, and the relation between a memory theory and a reconstruction theory.

Error and confirmation. Based on the Catch model, consider the hypothetical example depicted in the following table:

Source of choices	10 test pairs	Results
Similarity-rule used by D-R method	R R R R R R R R R R	Target face is reconstructed.
Subject choices	L L L L R R R R R R	Target face is reconstructed, since MEA = 40%.
Similarity-rule used by the H-D method	R R R R R R R R R R	Similarity-rule is rejected, since only 60% of the subject’s choices are predicted.

		Target face is reconstructed, since MEA = 40%.
New similarity-rule used by the H-D method	L L L L L R R R R R	New similarity-rule is accepted, since 90% of the subject’s choices are predicted. Target face cannot be reconstructed, since percent error (50%) > MEA = 40%.

Given ten test pairs, the similarity-rule chooses the right (R) test face in all ten test pairs. Application of the reconstruction-rule to these choices leads to the reconstruction of the target face. A subject has chosen the left (L) test face in the first four test pairs and the right test face in the last six test pairs. In this case the target is also reconstructed, since the MEA is 40%. (This hypothetical number is a rough estimate based on the results of several experiments.)

Would the similarity-rule as applied by the H-D method be considered a satisfactory theory? I think not. The similarity-rule predicts only 60% of the subject’s choices, but, according to the H-D method’s terms, this gap between the predictions and the observations is too large. Hence, the similarity-rule will be refuted and will have to be replaced by a “New similarity-rule.” Let us assume now that this New similarity-rule used by the H-D method predicts 90% of the subject’s choices: the left test face is chosen in the first five test pairs and the right in the last five test pairs. Although the New similarity-rule predicts the subject’s choices much better than did the previous one, and therefore the New rule is preferred by the H-D method over the previous rule, it cannot reconstruct the target face. In fact, the New rule produces five errors (50%), a number that is well above the maximum errors allowable. Clearly, the efficiency of the H-D method in achieving the goal of reconstructing the target face in the present case is nil.

This hypothetical situation occurs because the H-D method uses the smallest p-o gap to decide (accept-reject) a given theory. In contrast, in the D-R method, the decision whether to accept or reject a theory is based on the ability to reconstruct the event or image from the past by increasing, within limits, the p-o gap. The greater the number of errors allowed, the greater is its error-robustness and capability in attaining this goal. For example, if MEA = 0, then even one error is sufficient to obstruct the reconstruction of the target face. Hence, in view of the above, it is proposed that the D-R method fits the reconstruction situation (e.g., the Catch model research program) better than the H-D method.

Memory theory and reconstruction theory. Can one use a memory theory instead of a reconstruction theory and reach an accurate reconstruction of PE? The answer is no. First, to the best of my knowledge, there is no memory theory that assures reconstruction of PE as does the Catch model. Second, although there are very few instances where remembering is without errors, in general a memory theory deals with issues of forgetting, errors, distortions, fabrications and false memories. For many years these matters have been the focus of several lines of research, such as effects of schema, real-life false memories, eyewitness testimony and autobiographical memory (for a critical review see Koriat, Goldsmith, & Pansky, 2000). Third, since in general a memory theory deals with errors of memory, where an error is defined in terms of the difference between the true PE and the participant's memories, it is hard to see how one can validate these memories when PE are not known to the researcher. So far, research attempts to discriminate between true-false memories have failed (e.g., Loftus, 1997; Ross, 1997). Finally, I have shown above that the application of the H-D method to the domain of reconstruction leads to internal contradictions and failure to reconstruct the target face from memory. Hence, a memory theory does not seem capable to ensure the reconstruction of accurate PE. Rather it explains why memory of PE is impaired.

Nevertheless, these two kinds of theories—a memory theory and a reconstruction theory—can approach each other when one of the following two conditions is fulfilled.

A) When the subject's behavior matches the task-rule's predictions, and the memory theory's predictions match the subject's behavior.

B) When according to the D-R method, one proves that a memory theory used in conjunction with the appropriate analysis-rule succeeds in reconstructing past events.

Two Critiques

The "final analysis" critique. Despite these arguments, one may propose that the differences between the H-D method and the D-R method are not particularly significant. The success of the D-R method depends on the production of a good reconstruction, e.g., of a facial image that closely resembles the target face. That is, in the "final analysis" the D-R method has to fulfill the p-o decrease demand too. Hence, there is no difference between the H-D and the D-R methods. My answer to this critique is as follows.

It is true that the D-R method depends on an empirical test: the ability to produce a good reconstruction. In this respect, there is a similarity between the two methods. The main difference between these two methods is as follows: while the H-D method produces an internal contradiction, the D-R method does not. Consider the similarity-rule in the table shown above. Fulfilling the p-o decrease demand, the H-D method is contradictory: on the one hand it recommends rejection of the similarity-rule, since 40% of the subject's choices are errors, that is, the p-o gap is too wide, but on the other hand it recommends acceptance of this very rule, since this similarity-rule leads to a successful reconstruction of the target face. This kind of contradiction does not occur in the D-R method, since the similarity-rule is just a tool, a vehicle to achieve a successful reconstruction of PE. As such, the similarity-rule is evaluated in terms of the quality of Error-

Robustness, in terms of fulfilling the p-o increase (within limits) demand. That is, the reconstruction is successful when the subject's errors are less than MEA.

The statistical critique. One may suggest that the essential difference between the H-D method and the D-R method is the size of the p-o gap relative to the size of the standard error of data distribution. Given a large p-o gap, the H-D method would accept a task-rule (the similarity-rule) if the standard error were very large. The statistical test then would not reject the null hypothesis, since the p-o gap ranges within the customary level of significance, that is, predictions correspond to observations. Hence, in the case of a large standard error, the H-D method does lead to the reconstruction of past events, as does the D-R method (for discussions on these statistical terms see Mayo, 1996; Winer, 1971). My answer to this critique is as follows.

First, given that indeed the size of the p-o gap is large (about 35%, as mentioned above), this critique depends on the assumption that the data distribution is characterized by a very large standard error. Since such statistics usually correspond to a situation in which experimental conditions are not controlled, in my view this critique is very limited. In most cases (when the experiments are properly conducted), the H-D method rejects good task-rules that lead to past event reconstruction.

Second, consider once again the similarity-rule in the previous table. Based on the p-o decrease demand, the H-D method recommends rejection of the similarity-rule, since 40% of the subject's choices are errors, that is, the p-o gap is too wide. It accepts the New similarity-rule since only 10% of the subject's choices are errors, that is, the p-o gap is small. However, the New similarity-rule does not lead to the reconstruction of past events. Hence, the H-D method accepts task-rules that do not lead to past-event reconstruction, and rejects task-rules that do lead to successful reconstruction.

Finally, the statistical critique is based on the assumption that errors are randomly distributed. They are produced by unsystematic processes involved in measurement, sampling, assignment of subjects, etc. In contrast, the errors treated by the D-R method and the Catch model are behavioral. They are part of our behavior that determines whether one can successfully reconstruct PE. Behavioral errors are the data to be explained by scientific psychology, whereas random errors are not the subject matter for explanation. Statistical distributions of errors are used to infer whether a given phenomenon is genuine. If it is, then an explanation is provided. Hence, the statistical concept of error is different from the one used by the D-R method (see also Discussion).

Discussion

In this section I will briefly discuss the nature of the D-R method and the Catch model from three points of view: confirmation, explanation, and generality. From the conformation point of view, I will compare the present approach with Inference to the Best Explanation as a method for confirming hypotheses. With explanation, I will discuss the question of what kind of account is provided by the present approach to a reconstruction of PE. And with generality, I will discuss the scope of the D-R method.

Confirmation

As described above, the H-D method does not fit the reconstruction area (e.g., the Catch model research program). While according to the D-R method a reconstruction theory is confirmed when it leads to a successful reconstruction of past events by increasing, within limits, the p-o gap, according to H-D method, a theory is confirmed when the p-o gap is small. Does the method of Inference to the Best Explanation (IBE or Abduction), which is proposed as an alternative to the H-D method (see Josephson & Josephson, 1994; Lipton, 1991), fit this research area better than does the D-R method? My answer is no. The IBE is a kind of inference that, given a collection of observations (Data), searches for a hypothesis that is confirmed (is probably true) if it explains the Data better than alternative hypotheses do. In addition to the fact that the IBE is based on the p-o decrease demand and the D-R method is not, this method is different from the D-R method in the following major points. First, while IBE is, in essence, an inductive method, the D-R method is essentially deductive (see Rappaport, 1996, for examination of whether IBE is different from Mill's inductive methods). The D-R method rests on the Provability condition, which deductively guarantees that indeed past events will be reconstructed successfully. The method's other two conditions (Matching and Error-Robustness) guarantee that if the subject's behavior corresponds with the task-rule's predictions, then past events will be reconstructed from the subject's memory. According to the D-R method, one does not first examine the data and then look for the best (available) explanation; rather, one uses first a task-rule that fulfills the Provability condition, and then examines whether the subject's behavior fulfills the Matching and Error-Robustness conditions. Second, given that the Matching and Error-Robustness conditions are fulfilled, there will be nothing new, no surprise, in the reconstruction of past events from memory, because this result is expected from the Provability condition. However, with the IBE method, finding a satisfactory explanation is always a new discovery. Finally, while the IBE method proposes an ontological, causal hypothesis that explains the Data, the D-R method is based on a logical procedure that leads to past event reconstruction.

Explanation

A successful reconstruction of past events from memory (e.g., a target face) is explained by the fulfillment of D-R method's three conditions—Provability, Matching, and Error-Robustness. What is the relation between this kind of explanation and the following two models of explanation—the Teleological model and Hempel's (1965) Deductive-Nomological (D-N) model? One may argue that the D-R method provides a teleological kind of explanation, since the task-rule and the analysis-rule can be viewed as means to achieve the goal of past event reconstruction. However, while the explanation provided by the D-R method is similar to the D-N model, since it is based on deduction, the teleological explanation is based on "practical syllogisms" and not on deduction (for a discussion of a teleological explanation see Rakover, 1997).

Schematically, Hempel's model is based on the following components:

Premises: (a) Antecedent or initial conditions
(b) Laws

Conclusion: A description of the phenomenon to be explained.

The description of the phenomenon to be explained is logically deduced from the initial conditions in conjunction with at least one law. Hence, this model provides us with a general schema for explanation, for answering why-questions, by showing that the phenomenon deduced is an example of a law.

Can one formulate the D-R method along with the Catch model into Hempel's model of explanation? The answer is yes:

Stage I
Premises I: (a) Presentation of the target face and the test pairs
(b) Task-rule (similarity-rule)

Conclusion I: Choices of test faces
Stage II
Premises II: (a) Choices fulfill the Matching and the Error-Robustness conditions
(b) Analysis-rule

Conclusion II: Reconstruction of the target face

This "serial explanatory schema" is based on two stages: at Stage I the model determines the choice of a test face, and at Stage II the model determines the reconstruction of the target face. Hence, the explanation for the reconstruction of the target face is made with the help of a deductive process that uses task-rule and analysis-rule. The explanation shows deductively that reconstruction of the target face is a particular case of the D-R method and of the Catch model. Accordingly, one may conceive of the D-R method as a teleological explanation molded into the D-N model.

Generality

To the best of my knowledge, the Catch model is the only research program that substantiates the D-R method theoretically and empirically. Nevertheless, it should be pointed out that: a) Rakover & Cahlon (1999, 2001) have shown that there are several task-rules and analysis-rules that can lead to the target face reconstruction; and b) the domain of reconstructing past events from memory does not include only the reconstruction of a face previously seen. It includes other research topics that involve reconstruction of PE, such as autobiographical memory, the reconstruction of traumatic past events (e.g., post-traumatic stress disorder, childhood cruel sexual abuse), and the reconstruction of repressed traumatic childhood conflicts by classical psychoanalysis.

I shall now briefly present an interpretation of Freud's (1943) early psychoanalytical psychotherapy in terms of the D-R method. Freud proposed that traumatic conflicts that were repressed in childhood are the source of mental disturbance in adults. The goal of psychoanalysis is to reconstruct these conflicts and states that brought about these mental complexes from the unconscious, and to bring them into consciousness in order to treat them under the supervision and care of the psychoanalyst. Schematically, Freud's method for reconstructing traumatic past events consists of the following components. A) The patient performed several tasks (e.g., free associations and dream description); B) Freud treated the patient's performance by employing the psychoanalytic theory that determines what are the signs of a childhood trauma and how to analyze performance in these tasks; and C) Freud reconstructed the traumatic past events on the basis of B). Given this, it becomes clear that Freud's approach fits into the D-R method's domain. Nevertheless, the approach lacks the mathematical or logical basis, which exists in the Catch model. There is no guarantee that indeed the psychoanalytical reconstruction is valid. One reason for this is the "suggestion problem."

According to this problem, the therapeutic process itself affects the patient's thoughts and emotions (e.g., the psychotherapist imposes his or her views on the patient as to the mental trauma the patient has undergone) and markedly distorts and divert the clinical data. The philosopher of science Adolf Grünbaum (1984, 1986) notes that Freud, who was aware of the suggestion problem, put forward an interesting logical argument, which Grünbaum terms the "Tally argument," as a defense against suggestion. Nevertheless, Grünbaum argued convincingly that this argument is false.

Given the above, it can be seen that memory is a necessary condition for the application of the D-R method. This is why I propose that a phenomenon, which does not have a mechanism similar to human memory, cannot be investigated by employing the D-R method, but rather by the H-D method.

The H-D method is applied to both types of situations: explaining-predicting a phenomenon by appealing to a theory and the initial conditions (e.g., the stimuli presented to the participants), and reconstructing the initial conditions by appealing to a theory and a phenomenon. For example, on one hand we can predict the distance of fall from the law of free-falling bodies and time (i.e., the initial conditions); and on the other hand we can use this very same law and the distance to predict time. That is, in the natural sciences, one can empirically test the same theory by comparing the predicted observation with the actual observation, and by comparing the predicted initial conditions with the actual initial conditions.

Given this, the following question arises. Why, in the natural sciences, does the H-D method fits all kinds of research, whereas in memory research the H-D method does not fit all kinds of research (i.e., it does not fit the reconstruction area of research)? My answer is based on the fact that memory generates errors and that errors are treated differently in the natural sciences and in psychology.

In the natural sciences, errors are attributed not to the phenomenon to be explained (e.g., a free-falling body does not make an error), but rather to theories or experimental methods (e.g., errors of sampling and measurement). (For a discussion of types of experimental errors see Hon, 1989.) In contrast to the natural sciences, psychology assumes that subjects in experiments do make mistakes and that these are an important part of the behavior to be explained. For example, the debate on the inversion effect, mentioned in the introduction, is about errors in recognition of upright vs. upside-down faces. Hence, psychology handles two types of errors: methodological (theories or experimental methods) and behavioral. The methodological error is treated in psychology by means of the statistical tools that assume that errors are randomly distributed around a true value, whereas the behavioral error is subject matter to be accounted for by a psychological theory.

Given this distinction, I propose that these two kinds of errors lead to the employment of two types of methods in psychology. When one deals with the remembering area of research, one uses the H-D method, similar to its use in the natural sciences. However, when one deals with the reconstruction area, I recommend using the D-R method. In the former situation, one tests whether or not the p-o gap is within the range of the random error distribution. (If, for example, the gap extends this range, then according to the D-H method the theory from which the predictions were deduced is disconfirmed.) In the latter situation, one uses errors in a different way: as a behavioral criterion for achieving reconstruction of past events from memory. If the number of errors is not higher than the maximum allowable (MEA), then past events can be reconstructed from memory.

References

Bruce, V. (1988). *Recognizing faces.* Mahwah, NJ: LEA.

Davies, G. (1981). Free recall systems. In G. Davies, H. Ellis, & J. Shepherd (Eds.), *Perceiving and remembering faces.* New York: Academic Press.

Davies, G. (1986). The recall and reconstruction of faces: Implications for theory and practice. In H. D. Ellis, M. A. Jeeves, F. Newcombe, & A. Young (Eds.), *Aspects of face processing.* Boston: Martinus Nijhof.

Diamond, R., & Carey, S. (1986). Why faces are and are not special: An effect of expertise. *Journal of Experimental Psychology: General, 115*, 107-117.

Estes, W. K. (1994). *Classification and cognition.* New York: Oxford University Press.

Farah, M. J. (1992). Is an object an object an object? Cognitive and neuropsychological investigation of domain specificity in visual object recognition. *Current Directions in Psychological Science, 1*, 164-169.

Freud, S. (1943). *A general introduction to psychoanalysis.* New York: Garden City Publishing. (Original work published 1917)

Glymour, C. (1980). *Theory and evidence*. Princeton, NJ: Princeton University Press.

Grimes, T. R. (1990). Truth, content, and the Hypothetico-Deductive method. *Philosophy of Science, 57*, 514-522.
Grünbaum, A. (1984). *The foundation of psychoanalysis: A philosophical critique.* Berkeley, CA: University of California Press.
Grünbaum, A. (1986). Précis of the foundations of psychoanalysis: A philosophical critique. *Behavioral and Brain Sciences, 9*, 217-284.
Hempel, C. G. (1965). *Aspects of scientific explanation and other essays in the philosophy of science.* New York: The Free Press.
Hon, G. (1989). Toward a typology of experimental errors: An epistemological view. *Studies in History and Philosophy of Science, 20*, 469-504.
Josephson, J. R., & Josephson, S. G. (Eds.). (1994). *Abductive inference: Computation, philosophy, technology.* Cambridge: Cambridge University Press.
Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology, 51*, 481-537.
Kovera, M. B., Penrod, S. D., Pappas, C., & Thill, D. L. (1997). Identification of computer-generated facial composites. *Journal of Applied Psychology, 82*, 235-246.
Kuhn, T. S. (1970). *The structure of scientific revolutions.* Chicago: University of Chicago Press.
Lipton, P. (1991). *Inference to the best explanation*. London: Routledge.
Loftus, E. F. (1997). Creating false memories. *Scientific American, 277*, 51-55.
Mayo, D. G. (1996). *Error and the growth of experimental knowledge*. Chicago: The University of Chicago Press.
Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.
Melara, R. D. (1992) The concept of perceptual similarity: From psychophysics to cognitive psychology. In D. Algom (Ed.), *Psychophysics approaches to cognition.* New York: Elsevier.
Nosofsky, R. M. (1984). Choice, similarity and context theory of classification. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *10*, 104-114.
Nosofsky, R. M. (1992). Similarity scaling and cognitive process models. *Annual Review of Psychology, 43*, 25-53.
Penry, J. (1971a). *Looking at faces and remembering them: A guide to facial identification.* London: Blek Books.
Penry, J. (1971b) *Photofit kit.* Leeds: John Wadington of Kirkstall.
Popper, K. R. (1972). *The logic of scientific discovery.* New York: Wiley.
Rakover, S. S. (1990). *Metapsychology: Missing links in behavior, mind, and science.* New York: Paragon/Solomon.
Rakover, S. S. (1997). Can psychology provide a coherent account of human behavior? A proposed multiexplanation-model theory. *Behavior and Philosophy*, *25*, 43-76.
Rakover, S. S. (2002). Featural vs. configurational information in faces: A conceptual and empirical analysis. *British Journal of Psychology, 93*, 1-30.
Rakover, S. S., & Cahlon, B. (1989). To catch a thief with a recognition test: The model and some empirical results. *Cognitive Psychology, 21*, 423-468.
Rakover, S. S., & Cahlon, B. (1999). The Catch model: A solution to the problem of saliency in facial features. *Spatial Vision, 12*, 73-81.
Rakover, S. S., & Cahlon, B. (2001) *Face recognition: Cognitive and computational processes.* Amsterdam/Philadelphia: John Benjamins.
Rakover, S. S., & Teucher, B. (1997). Facial inversion effects: Parts and whole relationship. *Perception & Psychophysics, 59*, 752-761.

Rappaport, S. (1996). Inference to the best explanation: Is it really different from Mill's methods? *Philosophy of Science, 63,* 65-80.
Ross, M. (1997). Validating memories. In N. L. Stein, P. A. Ornstein, Tversky, B., & C. Brainerd (Eds.), *Memory for everyday and emotional events.* Mahwah, NJ: LEA.
Salmon, W. C. (1967). *The foundations of scientific inference.* Pittsburgh: University of Pittsburgh Press.
Suppe, F. (1998). The structure of a scientific paper. *Philosophy of Science, 65*, 381-405.
Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46A*, 225-245.
Tversky, A. (1977). Features of similarity. *Psychological Review, 84,* 327-352.
Valentine, T. (1988). Upside-down faces: A review of the effect of inversion upon face recognition. *British Journal of Psychology, 79,* 471-491.
Winer, B. J. (1971). *Statistical principles in experimental design*. New York: McGraw-Hill.