A Concept of *Critical Thinking*

A Proposed Basis for Research in the Teaching and Evaluation of Critical Thinking Ability*

The author has attempted to fill a gap which he perceives to exist in the literature on thinking. He has identified twelve aspects of critical thinking (construed as "the correct assessing of statements") and elaborated a system of criteria to be applied in it. The relevance of his enquiry for the schools is implied in the title and is close to the author's attention throughout the article.

Robert H. Ennis has studied at the Universities of Wisconsin (A.B.), New York (A.M.), and Illinois (Ph.D.) and is now Assistant Professor in the School of Education, Cornell University. With B. Othanel Smith he edited Language and Concepts in Education *(Chicago: Rand McNally, 1961) in which two of his essays were published; one of these, "The 'Impossibility' of Neutrality," first appeared in the* Harvard Educational Review, *XXIX, 128-136.*

ROBERT H. ENNIS
Cornell University

Introduction

Research in thinking has frequently been conducted in the fields of psychology and education, but in both fields there is a significant gap. There has been a lack of careful attention to the concept, *critical thinking*. Furthermore no comprehensive, thorough, up-to-date treatment of this concept is available.

Psychologists have been concerned with associative thinking, concept formation, problem solving, and creative thinking, all of which are important aspects of thinking. But psychologists have not chosen to go any deeper into the evaluation of the products of thought than John Dewey's analysis of reflective thinking. This is understandable. Until psychologists can establish predictable, theoretically-explainable regularities between their variables (set, motivation, aspiration, ego-involvement, flexibility, direction, drive-reduction, etc.) and the successful solution of simple problems, they are reluctant to

* The preparation of this work was facilitated by the Cornell Social Science Research Center through a grant from the Ford Foundation.

investigate the relationships between their variables and the acquisition of the knowledge and mental skills needed for judging solutions to complex problems. Thus they have not felt the need for a careful study of the knowledge and mental skills involved in making such judgments. Somewhat of an exception to this general pattern is the report of Bruner, Goodnow, and Austin (6), which did deal with such knowledge and skills in the rather restricted area in which a subject judges whether an hypothesized selection from a small number of given values of given variables is the one that the experimenter has picked.

Research in education, directed toward thinking, has often followed the concerns of research in psychology, since most educational researchers use this discipline as their base. There have, however, been a number of efforts aimed at the teaching of critical thinking. This is to be expected since such teaching is thought to be one of the educator's main jobs. But even in education such efforts have for the most part been deficient in an important respect: they have not been based on a comprehensive and detailed examination of what is involved in making judgments about the worth of statements or answers to problems. Propaganda analysis has amounted to an indiscriminate condemnation of attempts to influence people. Problem solving has been essentially a combination of a plausible method of motivation and a sensible but very general way of organizing an attack on a question. Some statements, including Dewey's (14, pp. 107, 109, 114), unfortunately suggest that the problem is solved when the solver thinks it is solved, thus providing a psychological instead of a logical criterion for the solution of a problem.

Though efforts in the field of education concerned with critical thinking have generally used a hasty conception of what is involved in judging statements, there are exceptions. Notable are the works of Henderson and Smith (27), Dressel and Mayhew (16), Hood (28), and Conant (10, 11). But even these efforts have not produced to a degree that is desirable an explicit comprehensive consideration of what is involved in judging statements.

A number of works in philosophy have dealt successfully with specific parts of what is involved (1, 2, 4, 5, 8, 12, 15, 17, 18, 19, 22, 23, 24, 25, 26, 29, 34, 35, 36, 37, 38, 41, 42, 43, 44, 45, 46, 49), but no one of them is as comprehensive as the present effort. Some books on logic are about as comprehensive but no one of them takes adequate account of the recent philosophical insights in a number of the works to which reference was just made. Furthermore no logic book attempts the dimensional simplification of the concept, *critical thinking,* which simplification is desirable as a basis for research. The present effort attempts such a simplification.

The program for developing this concept of *critical thinking* was as follows: (1) to examine the literature on the goals of the schools and the literature on the criteria of good thinking; (2) to select from this literature those aspects which come under the following basic notion of *critical thinking*: 'critical

thinking' as 'the correct assessing of statements';[1] (3) to elaborate the criteria to apply in making such assessments—to the extent that there are criteria; (4) to simplify the rather elaborate result by classifying some aspects under others that are logically more basic; (5) then to simplify further by logically analyzing them into basic factors or dimensions of *critical thinking*.

What follows is an attempt to describe the results of this program. First there will be a list of twelve aspects which come under the basic notion of *critical thinking* as the correct assessing of statements, and which commonly appear on lists of aspects of critical thinking. Some of these overlap or utilize others. Next comes a proposed analysis of *critical thinking* in a three-dimensional scheme composed of logical, criterial, and pragmatic dimensions. Third, the main body of this paper is a detailed consideration of each aspect showing (with sufficient detail to serve as a guide for instruction and evaluation) criteria that are appropriate for judging statements of the type covered by it, explaining its dimensional analysis, and showing the interrelationships between aspects. In the last section, possible research steps will be suggested.

To argue fully the claim that this is a defensible concept of *critical thinking* would take more time and space than is here available. A defense will be sketched in only at a few crucial points. The main task of this paper is to present a clear and detailed account of a concept of *critical thinking*. This account can be judged for its defensibility, and, if found defensible, can serve as a basis for research in the teaching of and testing for critical thinking ability.

A List of Aspects of Critical Thinking

As a root notion *critical thinking* is taken to be *the correct assessing of statements*. Since there are various kinds of statements, various relations between statements and their grounds, and various stages in the process of assessment, we can expect that there will be various ways of going wrong when one attempts to think critically. The following list may be looked upon as a list of specific ways to avoid the pitfalls in assessment.

As indicated before, this list is not intended to provide mutually exclusive categories. Instead it shows common pitfalls, and items about which people are concerned. Criteria of classification are sacrificed in the interest of plausibility and intelligibility. The dimensional categories meet the criteria of classification, but sacrifice initial intuitiveness in order to do so.

[1] This basic notion was suggested by B. Othanel Smith (39, p. 130): "Now if we set about to find out what . . . [a] statement means and to determine whether to accept or reject it, we would be engaged in thinking which, for lack of a better term, we shall call critical thinking." Since Smith's definition does not use any words like 'correct,' his notion is slightly different.

Smith's concept of *critical thinking* permits us to speak of 'good critical thinking' and 'poor critical thinking' without redundance or contradiction. Though this is an accepted manner of speaking, the predominant manner of speaking presumably builds the notion of correct thinking into the notion of critical thinking. Though the latter interpretation is used in this paper, it would be easy to restructure what follows and use Smith's concept. 'Good critical thinking' in Smith's sense means 'critical thinking' as used in this paper.

Twelve aspects of critical thinking are:

1. Grasping the meaning of a statement.
2. Judging whether there is ambiguity in a line of reasoning.
3. Judging whether certain statements contradict each other.
4. Judging whether a conclusion follows necessarily.
5. Judging whether a statement is specific enough.
6. Judging whether a statement is actually the application of a certain principle.
7. Judging whether an observation statement is reliable.
8. Judging whether an inductive conclusion is warranted.
9. Judging whether the problem has been identified.
10. Judging whether something is an assumption.
11. Judging whether a definition is adequate.
12. Judging whether a statement made by an alleged authority is acceptable.

Although the root notion calls for its inclusion, the judging of value statements is deliberately excluded from the above list. This exclusion admittedly weakens the attractiveness of the presented concept, but makes it more manageable. So long as we remember that this exclusion has occurred, we should not be confused by the truncated concept. Perhaps this gap can at some future time be at least partially filled.

The exclusion of other important kinds of thinking (creative thinking, for example) from this basic concept of *critical thinking* does not imply that the others are unimportant, nor does it imply that they are separable from it in practice. This exclusion is simply the result of an attempt to focus attention on one important kind of thinking.

The Three Dimensions of Critical Thinking

There are three basic analytically distinguishable dimensions of the proposed concept of *critical thinking*: a logical dimension, a criterial dimension, and a pragmatic dimension.[2]

The *logical dimension,* roughly speaking, covers judging alleged relationships between meanings of words and statements. A person who is competent in this dimension knows what follows from a statement or a group of statements, by virtue of their meaning. He particularly knows how to use the logical operators, 'all,' 'some,' 'none,' 'not,' 'and,' 'if . . . then,' 'or,' 'unless,' etc. He knows what it is for something to be a member of a class of things. Further-

[2] For those who are interested in comparing this analysis with Charles Morris' analysis of the dimensions of language (32) this logical dimension includes Morris' syntactic and semantic dimensions. And it includes more than these, since many terms do not refer to objects, but serve other functions. Morris' pragmatic dimension is quite different from this pragmatic dimension.

more he knows the meaning of the basic terms in the field in which the statement under consideration is made.[3]

The *criterial dimension* covers knowledge of the criteria for judging statements (soon to be presented), except for the logical criteria, which are covered by the logical dimension.

The *pragmatic dimension* covers the impression of the background purpose on the judgment, and it covers the decision as to whether the statement is good *enough* for the purpose. Including this dimension does not constitute endorsement of the doctrine often attributed to pragmatism: "a statement is true if it fulfills the purpose of the speaker." But the inclusion of this dimension does constitute recognition of the legitimate function played by the background purpose in making decisions about the acceptability of statements. It does constitute recognition of the necessity for the balancing of factors preceding the judgment, "This is enough evidence." Furthermore, inclusion of this dimension requires the admission that complete criteria can not be established for critical thinking. An element of intelligent judgment is usually required in addition to applying criteria and knowing the meaning.

To clarify what is meant by the function of the background purpose, let us consider two examples. In the first, the purpose helps us judge how important it is to be right, how strict must be our standards, and when there is enough evidence; in the second, the purpose enables us to judge how precise a statement must be.

Consider the problem in educational experimentation of deciding on the level of statistical significance to be demanded of the data. In making this decision about the level of significance, the background purpose of the investigation must be considered. Is the purpose to gather information on which to base a decision about how to teach reading in a large school system? Or is the purpose to gather information on the optimum time to keep a school bus? Certainly the first purpose imposes more stringent requirements on the data since much more is at stake. We have higher standards for assenting to a proposition in the first area than in the second. The judgment, "That's enough evidence to settle the matter," requires stronger evidence in the first area than in the second.

The importance of this point is generally recognized by statisticians. Guilford (21) remarks, "An investigator may choose any level of confidence he prefers. But he must defend it. It may depend upon the kind of problem being

[3] The requirement, 'knows the meaning of the basic terms in the field,' might better be partly thought of as a presupposition of critical thinking since, if a person does not understand a statement at all, the question of whether he can think critically about it does not arise. However, *knowing the meaning of a statement* involves *knowing the implications of the statement;* the latter is directly a part of the assessing of statements. Since a line between *bare understanding* and *knowing implications* is a difficult one to draw, the former is also included in critical thinking via the logical dimension, for simplicity's sake.

investigated and upon the seriousness of being wrong in concluding either for or against the null hypothesis."

In our second example a teacher is evaluating homework papers. She is trying to figure out who studied together. She says, *"John and Frank have the same answers to question four* and they are both way off." In that situation with her purpose, what she has said is all right even if John and Frank arranged the ideas in different order and used different words.

But suppose that she is interested in seeing if any of her students blindly copied from another. In that situation she should say, *"John and Frank do not have the same answer."* Greater precision is demanded by the second purpose. When faced with the same objective facts, we can justifiably make statements at different levels of precision, depending on the background purpose.[4]

The next section elaborates the list of twelve aspects and presents the dimensional makeup of each.

Clarification of the Twelve Aspects of Critical Thinking

1. Grasping the meaning of a statement.

If a person knows the meaning of a statement, he should know what would count as evidence for and against it. He should know what, if anything, it would imply in a situation and what would imply it. And he should know what statements contradict it.

The things to be known should, of course, not be expected to be more sophisticated than the statement in question, nor should they be expected to include things that are distantly removed. For example, it should not be required that in order to know the meaning of, "Wood floats on water," a person be able to recognize that this statement is implied by Archimedes' Principle together with certain facts about the densities of wood and water. These things are more sophisticated than the sentence about the floating wood and are a bit distant from it, logically speaking.

But a person should, if he claims to know the meaning of, "Wood floats on water," be able to recognize that it is implied by the combination: "Wood is lighter than water" and, "Whatever is lighter than water floats on it."

This aspect is the core of the other logical aspects, but they differ insofar as each distinguishes a different particular phase of the concept of *critical thinking.*

2. Judging whether there is ambiguity in a line of reasoning.

This aspect is a more complicated form of the first, since it requires two distinct apprehensions of the meaning of a statement—one as it is proved and

[4] Rupert Crawshay-Williams (12) has insightfully discussed this role of the background purpose.

one as it is applied. The two meanings are compared to see if they are the same. If not, the appropriate judgment is that there is an ambiguity.

There is an ambiguity in the following line of reasoning:

> There are people who sincerely believe on religious grounds that medication is wrong. They believe this because they believe that any treatment of human beings with medicine is a violation of their religious principles. 'Medication' means 'anything intended for the prevention, cure, or alleviation of disease.' Since the chlorination of water is intended for the prevention of disease, it is medication. To chlorinate water is thus to violate their religious principles.

The statement, "Chlorination is medication," is proven when the statement has one meaning: "Chlorination is something intended for the prevention, cure, or alleviation of disease." And it is applied with a different meaning: "Chlorination is treatment of human beings with medicine."

This aspect as exemplified has primarily a logical dimension since it requires grasp and comparison of meanings. Often, however, the ambiguity is not so blatant and one or both meanings must be inferred inductively. In such cases, all three dimensions are involved. The dimensional analysis of inductive reasoning will be explained later.

3. Judging whether certain statements contradict each other.

Judging contradictions is useful in making decisions about alleged deductions and in making decisions to reject the contradictory of an acceptable statement. Since contradiction is a matter of incompatible meanings, this aspect is on the logical dimension.

4. Judging whether a conclusion follows necessarily.

This is the one most often referred to when 'logical reasoning' is mentioned. It is the judging of deduction. Reasoning in mathematics, "if-then" reasoning, and syllogistic reasoning all exemplify deduction.

The basic criterion is this: "A conclusion follows necessarily, if its denial contradicts the assertion of the premises." Various rules have been developed for different types of deduction, but all see to it that this requirement is fulfilled. Well-developed sets of rules include:

- 4.1 The rules for handling equations and inequalities.
- 4.2 The rules of "if-then" reasoning:
  - 4.21 Denial of the "then-part" requires denial of the "if-part," but not necessarily vice versa.
  - 4.22 Acceptance of the "if-part" requires acceptance of the "then-part," but not necessarily vice versa.
  - 4.23 Instances of an "if-then" statement are implied by the "if-then" statement.
- 4.3 The rules for categorical reasoning. These rules may be sum-

> marized by the following: "Whatever is included in a general class is included also in whatever that general class is included in, and is excluded from whatever the general class is excluded from."

This aspect extends along the logical dimension only. There are no extenuating circumstances. Either a conclusion follows necessarily or it does not. It is partly because these decisions can be carefully and neatly systematized that deductive logic is given so much attention in logic courses.

5. Judging whether a statement is specific enough.

For the purposes of a given situation, a particular statement might be too vague to provide guidance. In such situations the statement should be rejected or inquired into, since in its condition its truth or falsity is irrelevant.

The statement, "Education has disappeared from the schools" (or, "There is more education in the schools than ever before") is useless in decision-making about curriculum and school finance until the terms, 'education,' 'disappeared,' and, 'the schools,' are clarified. The statements are not specific enough to be tested and applied.

On the other hand, in a war-ravaged country it might be quite meaningful to say that education has disappeared from the schools (since they are now used for hospitals or housing). In this situation, 'education,' even loosely defined, *has* disappeared from the schools.

This aspect requires consideration of the purpose of the discourse and requires the judgment, "This is (or is not) specific enough for our purpose." If the purpose is to come up with curriculum and budgetary recommendations for a school system long in existence, the statement is not specific enough. If the purpose is to make a report to the leader of a war-ravaged country, it is specific enough.

Another example of this aspect was given in the explanation of the pragmatic dimension—the teacher reading homework assignments. "The answers are the same," was specific enough for one purpose, but not for another.

This aspect extends on both the pragmatic and logical dimensions. That the pragmatic dimension applies was shown in the above discussion of the role of purpose. The logical dimension applies through the need for grasp of the meaning of the statement.

It might be thought that this aspect of critical thinking is one in which people do not make mistakes. In concrete situations this tends to be true, but in abstract situations it is easy to go wrong by forgetting to put questions and answers in the context of situations with purposes.[5]

[5] Crawshay-Williams (12) develops this point well.

6. Judging whether a principle establishes a statement that is alleged to be an application of it.

Decisions about the application of principles and hypotheses to the world of things, men, and events are similar to decisions about deduction, but there is a significant difference between them. Principles and hypotheses of the former type do not hold universally; they all have exceptions and limits. Sometimes these exceptions and limits are so far removed that we do not have to worry about them, and in such cases we can proceed as in deduction without fear of going wrong. Sometimes the limits and exceptions are close by, in which case, still approximating the deductive model, we use words like 'probable,' 'likely,' 'barring unforseen circumstances,' etc., in the conclusion.

For an example of the latter case, consider the application of that standard law of economics, "If the supply is constant and the demand for a product decreases, the price will decrease." Two of the limits of the application of this law are within the knowledge of all of us. It is intended to apply to an economy free of government control and to a sector of it that is free of monopolistic control. Mention of these limits will suffice for present purposes, although there are others.

Now let us apply this law to a situation in which there is a decrease in demand for microscopes. Applying the law deductively, we are unalterably committed to a prediction of a price decrease. But it is not wise to be unalterably committed to such a prediction. For one thing, the well-known limits of the law might be breached: the government might decide to maintain the price of microscopes and buy up the extras; or a monopoly might be formed which would decrease production without decreasing price.

But secondly, other things that are not yet explictly built into the limits might go wrong. The makers of microscopes might form a trade association and decide that with good advertising they can create a demand much greater than ever before, so that they can afford to raise prices.

It is because of considerations like this that qualifiers like 'probable' must be included in the application of many principles. The application of that law in that situation might be, "It is probable that there will be a lowering of price." But the application would not be this at all if it can be seen that a known limit is breached or that there is some other extenuating circumstance. The point is that the application of such principles should often not be stated any more strongly than this, even though the steps in reasoning parallel those of deduction.

This aspect of *critical thinking* is made up of two dimensions: the logical dimension, since one must decide what could be deduced, neglecting limits and extenuating circumstances; and the pragmatic dimension, since the question of how important it is to be right and how sure we must be in order even to say 'probable' must be faced.

7. Judging whether an observation statement is reliable.

An observation statement is a specific description. Over the years, those fields most concerned with accuracy of observation have built up a set of rules for judging the reliability of observation statements. These rules give a criterial dimension to this aspect of critical thinking. In addition, application of them to a particular situation, like the application of principles discussed previously, gives this critical thinking ability logical and pragmatic dimensions.

Here is a combined list of principles from the fields of law, history, and science:

- 7.1 Observation statements tend to be more reliable if the observer:
  - 7.11 Was unemotional, alert, and disinterested.
  - 7.12 Was skilled at observing the sort of thing observed.
  - 7.13 Had sensory equipment that was in good condition.
  - 7.14 Has a reputation for veracity.
  - 7.15 Used precise techniques.
  - 7.16 Had no preconception about the way the observation would turn out.
- 7.2 Observation statements tend to be more reliable if the observation conditions:
  - 7.21 Were such that the observer had good access.
  - 7.22 Provided a satisfactory medium of observation.
- 7.3 Observation statements tend to be more reliable to the extent that the statement:
  - 7.31 Is close to being a statement of direct observation.
  - 7.32 Is corroborated.
  - 7.33 Is corroboratable.
  - 7.34 Comes from a disinterested source with a reputation for veracity.
- 7.4 Observation statements, if based on a record, tend to be more reliable if the record:
  - 7.41 Was made at the time of observation.
  - 7.42 Was made by the person making the statement.
  - 7.43 Is believed by the person making the statement to be correct—either because he so believed at the time the record was made, or because he believes it was the record-maker's habit to make correct records.
- 7.5 Observation statements tend to be more reliable than inferences made from them.

8. Judging whether an inductive conclusion is warranted.

Inductive conclusions are of three types, simple generalizations, explanatory hypotheses, and theoretic systems. Though similar in many aspects, they are different enough to warrant separate treatment.

- 8.1 A simple generalization about experience. Such a generalization is warranted:

8.11 To the extent that there is a bulk of reliable instances of it. The greater the variability of the population, the greater the bulk needed.
8.12 To the extent that it fits into the larger structure of knowledge.
8.13 To the extent that the selecting of instances is unbiased.
  8.131 A pure random sample is unbiased.
  8.132 A systematic sample is unbiased if a careful investigation suggests that there is not a relevant cycle or trend followed by the sampling procedure.
  8.133 Stratification of a population on relevant variables and unbiased sampling within the strata, is likely to be more efficient than 8.131 or 8.132 alone.
  8.134 An unbiased sampling of clusters of the population and unbiased sampling (or complete enumeration) within the clusters is likely to be an efficient way of sampling when access to separate individual units is difficult.
8.14 To the extent that there are no counter-instances.

For example, the generalization that red-headed people tend to have hot tempers would be warranted to the extent that there is a large number of reliable instances of red-heads with hot tempers, to the extent that we are able to account for red-heads being hot-tempered, to the extent that our instances of red-heads are picked without bias, and to the extent that there is a lack of reliable instances of red-heads with even tempers.

The criterial dimension is involved here through knowledge of the above criteria. There is a vast literature on judging the adequacy of samples, which very rarely are purely random.[6]

The logical dimension is involved in the recognition of instances and the application of the above principles. And the pragmatic dimension is invoked in deciding that there is or is not enough evidence for the purposes of the inquiry.

8.2 An hypothesis which is related to its support through its explanatory power. An hypothesis is warranted to the extent that:
  8.21 It explains a bulk and variety of reliable data. If a datum is explained, it can be deduced or loosely derived (in the fashion of the application of principles) from the hypothesis together with established facts or generalizations.
  8.22 It is itself explained by a satisfactory system of knowledge.
  8.23 It is not inconsistent with any evidence.
  8.24 Its competitors are inconsistent with the evidence. This principle is the basis of controlled experiments.
  8.25 It is testable. It must be, or have been, possible to make predictions from it.

[6] A valuable introductory pamphlet on this topic was prepared by Philip J. McCarthy (31).

The hypothesis can be either specific (as is the case in law and usually in history) or it can be general (as is ordinarily the case in physical sciences and the social sciences of economics, sociology, and psychology). Here is an example of a specific hypothesis:

> Hissarlik is located at the site of Troy.

And here is an example of a general hypothesis:

> The pressure in a liquid varies directly as the depth, assuming the pressure at the surface to be zero.

For purposes of illustration let us consider the bearing of each of the criteria on each hypothesis.

Explaining a bulk and variety of reliable data (8.21). Since Hissarlik is only an hour's walk from the sea, the Hissarlik hypothesis explains the reported ability of the Greeks to go back and forth from Troy several times a day. It explains why there are ruins at Hissarlik. These explained reports, it should be noted, can be derived from the Hissarlik hypotheses together with established facts or generalizations:

H-1. Hissarlik is at the site of Troy.
Hissarlik is one hour's walk from the sea.
People are able to walk back and forth several times a day between places that are one hour's walk apart.

Therefore, it is probable that the Greeks were able to go back and forth from Troy to the sea several times daily.

H-2. Hissarlik is at the site of Troy.
A large city when abandoned tends to leave ruins.

Therefore, it is probable that there would be ruins at Hissarlik.

Of course explaining only those two pieces of evidence is not enough to establish the hypothesis. More evidence of different types must be provided.

The pressure hypothesis explains why water spurts farther from a hole near the bottom of a tank than from a hole in the middle of a tank. It also explains the proportional relationships between the following sets of readings of pressure gauges attached to the supply tank in a water system:

Distance from top of tank (in ft.)	Pressure reading (in lbs./sq. in.)
0	0
5	2.1
10	4.2

These data can be derived from the hypothesis together with established facts or generalizations:

P-1. The pressure varies directly as the depth.
The greater the pressure at a hole, the farther the liquid will spurt.
The bottom hole is at a greater depth than the middle hole.

Therefore, the water spurts farther from the hole near the bottom.

P-2. The pressure varies directly as the depth.
The depth at 10 ft. is twice that at 5 ft.

Therefore, the pressure at 10 ft (4.2 lbs./sq. in.) is twice that at 5 ft. (2.1 lbs./sq. in.).

Again the explanation of these data alone does not establish the hypothesis. More explained data of various types are needed.

Being explained by a satisfactory system of knowledge (8.22). If the Hissarlik hypothesis could itself be tentatively explained by established facts and generalizations, it would then be more acceptable. For example, suppose it were possible to show that the traits of the Trojans and the facts about the geography, climate, and nearby civilization at the time make it probable that the Trojan city would have developed at Hissarlik at the time that Troy was supposed to have existed. If it were possible to show that, the Hissarlik hypothesis would thereby receive support.

Similarly the pressure hypothesis is supported by showing that it can be explained, and thus derived, as follows:

Pressure in a liquid is the equivalent of the weight of a regular column of liquid extending to the top of the container.
The weight of a column of liquid varies directly with its depth.

Therefore, the pressure in a liquid varies directly with the depth.

Not being inconsistent with any evidence (8.23). The Hissarlik hypothesis would be weakened if no springs could be found in the area of Hissarlik, since the *Iliad* mentions two springs in the area, one hot and one cold. The reasoning might go as follows:

Hissarlik is at the site of Troy.
There were probably at least two springs at Troy, one hot and one cold.
Springs tend to remain in existence over the years.

Therefore, it is probable that there are at least two springs at Hissarlik, one hot and one cold.

Note that in using the absence of springs as evidence against the hypothesis, we are assuming that springs tend to remain and that the report of the *Iliad* is reliable. Either of these could be wrong. The less dependable these auxiliary assumptions are, the less dependable is our counter-evidence.

The pressure hypothesis would be weakened by the discovery that water

spurted out the same amounts at the middle and the bottom, since the hypothesis implies otherwise. That is, it would be weakened if we did not previously have so much by way of other evidence built up in favor of the hypothesis—so much that in this case, one would have a right to suspect such data.

Its competitors' being inconsistent with the data (8.24). A competitor of the Hissarlik hypothesis is the hypothesis that Bunarbashi is at the site of Troy. This competing hypothesis is not consistent with the data that Bunarbashi is a three-hours' walk from the sea and that the Greeks were able to go back and forth several times daily, if we assume that the Greeks walked.

A competitor of the earlier-stated pressure hypothesis might be one to the effect that the pressure increases directly as one gets closer to the surface of the earth. This hypothesis is inconsistent with pressure gauge readings on two independent tanks, one over the other, when the top tank has the pressure gauge at its bottom, and the bottom tank has its gauge at the top. The alternative hypothesis implies that the gauge in the upper tank would give the smaller reading. The data are just the opposite.

A controlled experiment is designed to rule out competing hypotheses by producing data inconsistent with them. When we test the hypothesis that a new fertilizer will increase the growth of corn, we put the fertilizer in a corn patch, develop a companion corn patch, the control, identical in every respect possible except for fertilizer, and watch the results. If there is a difference, the fertilizer hypothesis can explain it. But it would not be explained by heavy rainfall, warm weather, sunlight, etc., since both patches supposedly received the same amount. These alternative hypotheses would justify a prediction of no difference and would thus be inconsistent with the data.

It is, of course, impossible to develop a perfectly controlled experiment, since the perfect isolation and variation of a single variable is not possible. The important thing might be a *combination* of weather and fertilizer, or the important thing might have slipped by unnoticed. But we can still see in the controlled experiment an attempt to approximate the logical goal of eliminating hypotheses by turning up data that is inconsistent with them. The controlled experiment is an efficient way of eliminating hypotheses by this method.

Being testable (8.25). This is a logical criterion, not a criterion of practicality or even physical possibility. The criterion requires only that it must be possible to *conceive* of what would count as evidence for, and what would count as evidence against, the hypothesis. We have already seen that this is possible for each of our hypotheses. The fact that some conceivable tests are not practically possible is not important so far as this criterion is concerned. A conceivable, though presumably physically impossible, test of the pressure hypothesis would involve swimming to the bottom of the ocean with pressure and depth gauges, recording readings at various points along the way.

An hypothesis that appears untestable is this one: "Airplane crashes are

caused by gremlins," since it does not appear to be possible to conceive of something that would count as evidence for and to conceive of something that would count as evidence against the hypothesis. The word "appear" is used deliberately, since the conceiving of evidence for and evidence against would immediately make the hypothesis testable—and would reveal to some extent the meaning of the hypothesis.

Most hypotheses that we consider are testable in this logical sense, so this criterion does not often discredit an hypothesis. But its fulfillment is absolutely essential for hypotheses about the world of things, men, and events.

Having thus considered the bearing of each criterion on each hypothesis, we may turn to the dimensional analysis of this aspect of *critical thinking*. Judging whether an explanatory hypothesis is warranted incorporates all three dimensions. The logical dimension is needed in the comprehension of the meaning of the hypothesis, in judging whether the evidence can be derived from the hypothesis, in judging whether the hypothesis can be derived from a broader and satisfactory system, in judging whether it and its competitors are inconsistent with the data, and in judging whether it is testable.

The criterial dimension functions in that the criteria, 8.21-8.25, should be known.

The pragmatic dimension is utilized in judgments about the derivation of, and from, hypotheses when matters of explanation are at hand. It is invoked here just as it was invoked in the application of principles—since one must judge that the warrant is strong *enough* to give an explanation. Similarly judgments about inconsistency have a pragmatic dimension, since the implications are only more or less probable. When a potential inconsistency is faced, one must decide whether the other material used in the derivation is strongly enough grounded so that the fault lies with the hypothesis—and not the assumptions. And one must decide whether a face-value inconsistency is to count as refuting the hypothesis or to be considered beyond the range over which the hypothesis is to hold.[7] This decision is ordinarily based on decisions about convenience or simplicity, which in turn depend on the purpose for which the hypothesis might be used.

But most important of all, the so-called 'inductive leap' utilizes this dimension. Deciding that the evidence at hand is *enough* to establish the hypothesis requires consideration of the degree of satisfaction of the criteria, the purpose, and how important it is to be right; then the leap may or may not be made.

8.3 A theoretic system.

The difference between a theoretic system and an hypothesis of the type we have been considering is that the former is an involved network of relations between concepts, many of which are abstract and technical, while the latter is a simple relation between two or a small number of concepts, often less abstract and technical. Examples of theoretic systems are the kinetic theory

[7] Toulmin (43) provides a valuable discussion of this matter.

of matter, the atomic theory, Gestalt psychology, the theory of evolution, Keynesian economics, Turner's frontier theory, and classical English grammar. Obviously evaluation of theories is a demanding task. It demands more than we can ordinarily expect of elementary and secondary students. Undergraduates are sometimes better equipped, and graduate students are expected to become equipped to perform this task.

Evaluating theories is comparable to evaluating hypotheses but much more complex. In general the same criteria apply but on a broader scale. Two modifications should be noted: the addition of the criterion of simplicity and the weakening of the effect of contrary evidence.

The criterion of simplicity calls for choosing the simpler of two competing systems, other things being equal. The classic example of the application of this criterion is the preference of the Copernican system, which considered the sun the center of the universe, to the Ptolemaic system, which looked upon the earth as the center. The Copernican system was simpler since it needed fewer cycles and epicycles to explain the movements of the planets.

Since theories have so many parts, contrary evidence does not usually result in outright rejection, but rather in adjustment to fit the contrary evidence—until the whole system becomes more complex than a competitor. The criterion of simplicity then functions.

Following are the criteria for theoretic systems. There will be brief comments, but no attempt to exemplify the operation of each will be made, because to do so would be a monumental task and would make rather laborious reading for those not versed in the fields chosen. The reader is invited to provide examples from a theoretic system in a field he knows.

A theoretic system is warranted to the extent that:

8.31 It explains a bulk and variety of reliable data. Within the system, furthermore, the less abstract statements should be explained by the more abstract ones.

8.32 It is explained by broader theories. Some theories are so broad already that, with our present state of knowledge, to demand fulfillment of this criterion is often to demand speculation.

8.33 It is not inconsistent with any evidence. As indicated earlier, occasional inconsistency can be handled by adjusting the theory. Sometimes the inconsistency must just be accepted for lack of a better theory, and we say, "The theory does not hold for this kind of case."

8.34 Its competitors are inconsistent with the data. Again a single inconsistency does not destroy a competitor, for it too can be adjusted, but a large number of inconsistencies damage it.

8.35 It is testable. When adjusting a theory to fit the data, people are sometimes tempted to make the theory impregnable by making it untestable. Freudian psychology is sometimes accused of being untestable.

8.36 It is simpler than its rivals. As theories are adjusted to fit new

> data, they may become extremely complicated, as had happened to the entire Ptolemaic system at the time of Copernicus.

All three dimensions are used in judging theories for reasons like those advanced in the discussion of hypotheses.

9. Judging whether the problem has been identified.

Different kinds of judgments go under this label:

a) Judging that a want has been identified. As when someone says, "My problem is to learn to appreciate poetry." In this sense the judgment that the speaker has identified his problem is tantamount to judging that this is something the speaker, who might also be the judge, really wants to do. Problem identification here is identification of wants, either one's own, or someone else's. If they are one's own introspected wants, then critical thinking is not involved. For a person to know his wants (felt needs) is something that he can not fail to do.

If they are someone else's wants, then identifying problems is the same as establishing inductive hypotheses, as is the case for all subconscious wants, one's own or someone else's: for example, "Mark's problem is to get attention." Judging the identification of someone else's problem and of subconscious wants are then critical thinking of a type already discussed—judging inductive hypotheses.

b) Judging that a valuable goal has been selected. Here is such a problem identification: "Our problem in Culver City is to increase respect for law and order." Insofar as that is a statement of an end rather than a means, the judgment that it is an adequate identification of a problem is a value judgment. For reasons indicated earlier, this type of judging, though important, is excluded from this concept of *critical thinking*.

c) Judging that a means decision is adequate. For example, if the broader objective were respect for law and order the following might be a statement of a means decision: "Our problem in Culver City is to establish a youth bureau." The judgment here that the problem has been identified does at least these two things: (1) implies endorsement of the goal of respect for law and order (this part of the judgment then is a value judgment); (2) says that the means selected will facilitate achievement of the goal and that they will be at least more likely to facilitate it than any other course of action, within the limits of existing resources and goals. These limiting goals, by the way, are another place where values are impressed on problem identification. To judge that the problem has been identified is to judge that no unjustified goal violation would take place if the problem were solved.

To apply the means interpretation to our example: it is there implied that establishment of a youth bureau would increase the likelihood of winning respect for law and order, and would be more likely to do so than any other

course, such as giving the job to the schools (given the limits of the school budget, the state funds available for youth bureaus, trained personnel available, the goals of the school, etc.).

Judging a means decision is judging the application of a principle and judging the acceptability of the principle. To judge whether a youth bureau in Culver City would result in increased respect for law and order is to judge whether a principle about the effectiveness of youth bureaus, applied to this situation, gives us this statement with sufficient probability; and to judge whether the principle is acceptable.

In summary, problem identification is many different things and often a combination of them. Elements capable of being treated under the proposed notion of *critical thinking* are (1) judging the alleged identification of the wants of others and of subconsious wants (inductive hypotheses), and (2) judging the assertion of a means of reaching a goal (judging the application of principles and judging the principles themselves). Since each of these types of judging is treated elsewhere, nothing further need be said here.

10. Judging whether something is an assumption.

This topic is also complicated because there are various logically-different abilities that go under this title. These can be best approached through an examination of various uses of the word 'assumption': the deprecatory use, the concluding use, the premise use, and the presupposition use.

10.1 The deprecatory use and the concluding use.

The deprecatory use is the use when the charge is made that there is little or no evidence to support a given belief and that the belief is questionable. Here is an example: "You're just assuming that Frank didn't read the assignment." This deprecatory use is often found to be incorporated in the other uses, but sometimes it stands alone. As such its appearance is tantamount to a judgment that the view should perhaps be rejected, or at least be held in abeyance because of lack of support. No further discussion of the evaluation of this kind of assumption-claim is necessary here, since this is a general charge and is covered under discussion of the various other abilities.

In the concluding use, the term 'assumption' is used to mark a conclusion, but the deprecatory use is involved too, since the conclusion is implied not to be fully established. Here is an example: "My assumption is that Hissarlik is at the site of Troy." We need not be concerned with discussing whether something is an assumption in the concluding sense; the important question is whether the assumption is justified, and that question is covered elsewhere in this paper.

The first two uses of 'assumption' were specified in order to keep them out of the discussion of the next two; the following discussion, therefore, does not apply to them.

10.2 The premise use.

This kind of assumption stands anterior to a conclusion in a line of reasoning, whether the conclusion be inductive or deductive. To call something an assumption is to say that the conclusion depends upon it, and that if the conclusion is to be accepted, the alleged assumption should also be accepted. Thus the location of assumptions (in this sense) is a useful step in the evaluation of conclusions.

Here are criteria for premise-type assumptions:

10.21 Of the various possible gap-fillers, the alleged assumption should come closest to making the completed argument, proof, or explanation, a satisfactory one. (This criterion is necessary and sufficient.)
- 10.211 The simplest gap-filler is ordinarily the one to choose.
- 10.212 If there is a more plausible gap-filler among the more complex ones, it should be chosen. Plausibility, however, requires fitting in with existing knowledge—not being a special case.

10.22 Other conditions remaining the same, the state of affairs that is predicted could not occur (or probably would not occur) if the alleged assumption were false. (This criterion applies only to alleged empirically necessary assumptions, but for them it is necessary and sufficient.)

10.23 The community of experts in the field would not accept the position, conclusion, or argument, without first believing the assumption to be true. (This criterion is neither necessary nor sufficient, but is a good ground.)

What is a gap-filler? Consider this piece of reasoning:

> Since the demand for microscopes has decreased, the price may be expected to decrease.

A gap-filler here would be the principle mentioned earlier:

1. When the demand for a commodity decreases, the price will decrease.

It fills a gap in reasoning from the decrease in demand for microscopes to a decrease in price.

It is not the only way to fill the gap, however. Consider these alternatives:

2. When the demand for goods and services decreases, the price decreases.
3. When the demand for optical instruments decreases, the price decreases.
4. When the demand for optical instruments (other than field glasses) decreases, the price decreases.

Since all three of these will fill the gap, it should be clear that being a gap-filler is not by itself a sufficient condition for being an assumption. The sim-

plest gap-filler is ordinarily the one to attribute, thus ruling out *4*. Simplicity might also be a ground for not accepting *2* as the assumption, since there is a conjunction of two things (goods and services) mentioned. But if the prevailing knowledge in economics admits no basis for distinguishing goods from services in the context of this "law," then simplicity is counterbalanced by the need to fit into existing knowledge. Gap-fillers *1* and *2* would then be equally defensible (or indefensible) and either could be called the assumption.

Gap-filler *3* introduces a new twist, talking only about optical instruments. It is as simple as *1* but is not as general. Other things being equal, generality is to be preferred. A system of knowledge is better if it covers more cases. But if the more general gap-filler (*1*) should be false, and the less general one (*3*) true (or more likely to be true), the less general gap-filler is the one to choose.

Assumption-finding then is the locating of a gap-filler, the simpler the better, provided that it fits into and contributes to a system of knowledge. The assumption-finder should try to be generous to the position whose assumptions he is locating, generous in that he should try to find the best candidate for participation in a knowledge system. He should not accept a false gap-filler as the assumption until he has searched for one that fits into an acceptable body of knowledge. Put more simply in a way that covers most cases, he should search for one that is true.

While discussing gap-filling it would be well to note that there is one sometimes-used criterion that is inapplicable: *logical* necessity. As exemplified by the four gap-fillers previously discussed, there is no single premise-type gap-filler which is logically necessary. It is always *logically* possible (though it may be extremely implausible) to complete an argument in more than one way.[8]

Empirical necessity (10.22) is different. To the extent that empirical statements can be necessary, there can be empirically necessary assumptions. For example, an argument which predicts that the pressure in a fixed cylinder of confined air will increase is assuming that there will be a temperature increase. Since an increase in temperature in that situation *is* necessary for there to be an increase in pressure, the assumption is empirically necessary and can be pinned on the argument with confidence.

Criterion 10.23 mentions the experts. Although their considered opinions can be wrong, they do ordinarily know what fits into their body of knowledge. And they do know what is used successfully in the field to back up arguments and conclusions. So they can ordinarily be expected to know what an argument would need in order to be a good one.

Premise-type assumption-checking extends in all three dimensions. The logical dimension is used in judging whether the alleged assumption would fill the gap. The criterial dimension is used in knowing the criteria of assumptions. And the pragmatic dimension operates when the decision is made that the search for other satisfactory gap-fillers has been thorough *enough*.

[8] This point has been developed elsewhere by the author (17).

10.3 The presupposition use.

Presuppositions are sentences which must be true for a given statement even to make sense.[9] The claim, "The governor's mistakes have caused our present plight," presupposes that the governor has made mistakes. His not having done so would make nonsense out of either the affirmation or denial of the claim. If the governor has made no mistakes, it does not even make sense to say that his mistakes have caused our plight; nor does it make sense to say that his mistakes have not caused our plight.

Presupposition-finding is useful in avoiding being swayed by false presuppositions (if the governor has made no mistakes, we should be able to react to the presupposition that he has). And presupposition-finding is useful in grasping a verbal picture, and a part or the whole of a theory.

Judging whether something is presupposed by something else is simply a matter of stating the meaning of the "something else." Thus presupposition-finding has only a logical dimension.

11. Judging whether a definition is adequate.

There are various kinds of activities that are called 'judging the adequacy of a definition.' One major distinction to be drawn is that between judging the adequacy of a concept and judging the explanation of the meaning of a term.

11.1 Judging the adequacy of a concept.

A concept in fact-stating discourse is valuable to the extent that it plays a contributing role in a worth while system of knowledge. For example, consider two concepts which are competitors for the term 'intelligence': *innate problem-solving ability* and *scholastic aptitude*. A judgment about the claim that 'innate problem-solving ability' is a more adequate definition of 'intelligence' is a judgment about the worth of the concept, *innate problem-solving ability*. And this depends on the extent to which this concept contributes to a worthwhile theoretic system. The value of a theoretic system has been discussed before (8.3) so no more will be said here. Judging adequacy of concepts in fact-stating discourse involves all three dimensions, as with the judging of systems.

A judgment about the adequacy of a "persuasive definition"[10] is a special kind of judgment of a concept. A persuasive definition is an attempt to put across a proposal by attaching desired criteria to a term which arouses emotions. For example the definition of 'freedom' as the 'power to do' is a persuasive definition. The term 'freedom' is a term which arouses positive emotions. To try to attach the criterion, 'power to do,' to that word, instead of 'lack of government restraint' is to propose in effect that power to do is more important than lack of government restraint. Thus a judgment about the adequacy of

[9] This is the sense of the term 'presupposition' that was presented by P. F. Strawson (42).
[10] A term proposed by C. L. Stevenson (41).

this kind of definition is in effect a judgment about the worth of the proposal built into the concept.

To the extent that such judgments are value judgments, as is the case with the definitions of 'freedom,' making them is excluded from the present notion of *critical thinking*. This exclusion itself, by the way, exemplifies the choice of one concept over another. In order that prediction and control of students' behavior be facilitated, an unwieldy area (evaluating value statements) has been eliminated from the concept, *critical thinking*.

11.2 Judging the explanation of terms.[11]

In addition to concern about the adequacy of the underlying concepts, three other things can be involved in judging the adequacy of a definition. One might judge whether a stipulation of a short term in place of a longer term provides efficiency, one might judge whether the standard meaning of the term is being reported, or one might judge whether the form of the definition is appropriate. Each of these areas of judgment will be discussed.

11.21 Stipulated definitions. Stipulated definitions (example: "Let us use 'DDT' for 'dichloro-diphenyl-trichloro-ethane' ") can go wrong by failing to warn us that they are stipulations rather than reports, by being inefficient (by loading us down with new terminology), or by being in poor form. Judgments about the stipulation-warning and about efficiency involve the application of principles, and thus involve all three dimensions: the application of principles requires the logical and pragmatic dimensions, and knowledge of the principles brings in the criterial dimension. Judgments about form will be discussed later.

11.22 Reported definitions. Judging whether the standard meaning of the term has been *reported* is a matter of applying the definition to see if it excludes what should be excluded and includes what should be included. To do this requires knowledge of the standard meaning (the logical dimension) shown either through examples and non-examples, or through another definition; comprehension of the offered definition (the logical dimension); and either application of the offered definition to the known examples and non-examples (the logical and pragmatic dimensions), or comparison of the offered definition with the known definition (logical and pragmatic dimensions again). Reported definitions can go wrong by failing to represent the standard meaning (in which case it is appropriate to say that the definition is false) or by being in poor form.

This is an example of a reported definition:

> 'Tissue' means the same as 'group of similar cells performing the same function.'

If this definition either fails to represent the standard meaning or if it is in poor form, then it is inadequate. Both of these are matters of degree.

[11] Robinson's treatment of definitions (35) is enlightening and the source of many of the ideas expressed here.

No reported definition of a term that applies to the world of things, men, and events can be expected to do a perfect job of inclusion and exclusion. There will always be borderline cases and exceptions. So the pragmatic question must be answered: "It this good enough for our purposes in this context?"

For example, the above definition of 'tissue' implies that the cells of the skin of two people form *a* tissue since these cells all perform the same functions. In the context of a general science class, no heed need be paid to this exception. But there must not be exceptions which will mislead the audience in the enterprise in which it is engaged.

11.23 Forms of definition. The forms of definition are various. Decisions about which is appropriate depend on the nature of the term being defined and the context in which the definition is offered. Only the major forms are considered here; there are many variations which are legitimate and not considered. Equivalence of the term being defined and that which is offered in definition is usually desirable. The classification and synonym forms come closest to achieving this goal. The equivalent-context form and the range form achieve it more or less. And the example form and operational form do not achieve it at all, since they ensure a concrete interpretation.

11.231. Definition through classification. Classification definition is the prevalent form of definition and justifiably so. It provides brevity, rigor, and substitutability. The definition of 'tissue' given previously is a classification definition. Tissues were classified as groups of cells and were then distinguished from other groups of cells.

Here are the rules for classification definition:

11.2311 The defining part should contain (a) a general class, and (b) a feature or features that set this member off from members of the general class.
11.2312 The defining part should be equivalent to the part being defined.
11.2313 The defining part should not use the term to be defined.
11.2314 The defining part should not give more than enough to provide a complete classification.

The obvious exception to rule 11.2313 is that situation in which the term to be defined is composed of two or more terms (e.g. 'parallel lines'), one of which is already familiar. The term 'lines' can be repeated in the defining part, since it is not in question:

> 'Parallel lines' means the same as 'straight lines in the same plane which do not meet however far extended.'

The exception to rule 11.2314 is the situation in which the definition is to serve as a means of teaching a concept. Redundancy and even non-defining characteristics are often permissible because they are often useful in teaching.

The following definition is redundant:

> 'Parallelogram' means the same as 'four-sided plane figure with opposite sides equal and parallel.'

Either the reference to being equal or the reference to being parallel could have been omitted without any logical loss to the definition. But teaching the concept *parallelogram* is often facilitated by reference to both.

Here is an example of the addition of a non-defining characteristic:

> 'Concave lens' means the same as 'a lens with both surfaces curved inwards, which diverges parallel rays.'

The last clause gives a non-defining characteristic. A lens with both surfaces curved inward would be concave, even if it did not diverge parallel rays. But for purposes of teaching the concept at a low level of sophistication, it helps to include this characteristic in the definition.

Although classification definitions are appropriate for most situations there are certain kinds of situations for which they are not appropriate. They can be used only inconveniently to define relationship; they can not be used to give a concrete interpretation of abstract terms; and they are ill-suited in spirit to terms that have boundaries that are quite indefinite. To handle each of these weaknesses there is another form of definition which is more appropriate than classification definition. These forms will be considered after synonym definitions are mentioned.

11.232 Synonym definition. Where close synonyms exist, the meaning of which is known to the audience, this is an efficient and thus useful form. Synonyms are only rarely really equivalent, so caution should be exercised.

11.233 Definition through equivalent contexts. An equivalent-context definition puts the term in question in a standard context and equates this context with another context, presumably better understood. This form of definition is particularly appropriate for terms that express relationships between two things. The term 'anonymously' expresses a relationship between, say, a poem and its author, and may be defined as follows:

> 'The poem was written *anonymously*' means the same as, 'The writer of the poem did not reveal his name.'

11.234 Definition by example. This way of giving the meaning of a term is a very commonly used method. It is usually the best way to *teach* the meaning of a new term, if the new term refers to a new concept. Loose general terms which function largely in systems that organize facts (as opposed to providing prediction and control) are ordinarily best defined for the beginner by a list of examples *and* non-examples. 'Romanticism,' 'scientific method,' and 'nationalism' are terms that can be most thoroughly defined by abundant examples and non-examples. Shorter definitions for these terms can of course be framed. These will be considered later under the topic, "range definition."

A criterion for definition by example can be stated only vaguely: the examples and non-examples should be so selected that they clear up questions that are likely to arise. This requires enough variety to cover approximately the same area covered by the term. To call this a 'definition' may bother some people. It is called a definition here because it is a significant way of showing meaning, and is an alternative to the others; but it could be called something else and the important things said about it would still hold.

When economy of space is important, definition by example should not be used, unless it takes only a few examples to clear up a few doubtful areas. Definition by example is helpful in teaching terms used in systems of knowledge that are used for prediction and control, but for people working in the field it should be superseded by a more rigorous form. Probably the classification definition should be used if concrete interpretation is not needed, and if it is needed, the operational definition should be applied.

11.235 Operational definition. This form ordinarily puts in an if-clause, an operation performed by an investigator, and then gives with as much rigor as possible the relationship between the concept and the observation that would be made subsequent to the operation. For a definition of 'intelligence quotient,' the operation could be the administration and grading of a test; and the observations could be reports of the answers, summarized by the I. Q. score:

> If the "California Test of Mental Maturity" is administered according to directions, a person's score approximates his intelligence quotient.

Here is an operational definition of 'sovereignty':

> If Nation A orders the fishing vessels of Nation B to leave an area, the sovereignty of Nation A in that area would imply that the vessels probably will leave, or be apprehended, destroyed, or otherwise interfered with.

There are three interesting things to note about this definition:

1) The qualifier 'probably' is explicitly present. Some qualifier like this is always to be understood as part of an operational definition. To make the qualifier explicit simply suggests more strongly that the implied result might not occur.

2) Even after the antecedent condition is fulfilled, an equivalence is not suggested. A necessary condition of *sovereignty* is given instead. In the previous definition, a condition was given that was roughly both necessary and sufficient after the fulfillment of the antecedent condition.

3) No actual operation by an investigator is mentioned. We have here an extension of operational definition to cover those observations where the only operation involved is looking to see what there is to see. But the essential parts are there: concrete interpretation with a somewhat rigorous relationship be-

tween the concept and the interpretation, the relationship being alleged to hold if a condition is fulfilled.

11.236 Range definition. When definition by example was being considered we noted the problem of conveying the meaning of loose general terms with less consumption of time and space. *Range definitions*[12] do this task. They give a list of characteristics most of which apply whenever the term applies. The boundaries of range concepts are indefinite, as are the boundaries of mountain ranges. Range definitions show this indefiniteness. Here is a range definition of 'romanticism,' which shows the indefinite boundaries by the use of the expression, 'characterized by':

> 'Romanticism' is characterized by (1) a sturdy and plainly expressed belief in the brotherhood of man; (2) a deep sympathy with humble lives, human and animal alike; and (3) a sense of the independent spirit of man and his natural right to freedom.[13]

The criteria for judging range definitions are that they should be as loose as the concept and cover the same indefinite territory.

Judgments about form of definitions require all three dimensions of critical thinking ability. Knowledge of the rules for selecting forms as well as the rules of the form to be used is required, thus giving the criterial dimension. Application of these rules involves the logical and pragmatic dimensions.

In summary, judging the adequacy of a definition involves judging the worth of the concept and/or judging the way it is put. All three dimensions are involved, as the analysis has shown.

12. Judging whether a statement made by an alleged authority is acceptable.

This aspect requires development in all three dimensions since it is concerned with the application of principles and, insofar as the evaluator is able, requires evaluation of the statement irrespective of its source. The criteria that apply are fairly straightforward, though they must be applied with discretion:

- 12.1 To the extent that one is able to apply them, the relevant criteria from *1* through *11* should be applied.
- 12.2 Furthermore the credentials of the alleged authority should be considered. An alleged authority should be accepted to the extent that:
  - 12.21 He has a good reputation.
  - 12.22 The statement is in his field.
  - 12.23 He was disinterested—that is, he did not knowingly stand to profit by the results of his statements (except that he may have stood to have his reputation affected).

[12] A term and concept proposed by Max Black (4).

[13] Adapted from Rewey Belle Inglis, *et al., Adventures in English Literature* (New York: Harcourt, Brace and Company, 1952), p. 262.

12.24 His reputation could be affected by his statement and he was aware of this fact when he made his statement.
12.25 He studied the matter.
12.26 He followed the accepted procedures in coming to his conclusion (although there are legitimate exceptions to this requirement).
12.27 He was in full possesion of his faculties.

All three dimensions are involved, since these judgments call for the application of criteria.

Summary and Comment

A range definition of *critical thinking* has been provided by listing twelve aspects which characterize a critical thinker. It should now be clear that there is considerable overlap in the list of twelve aspects. For example, a statement that the problem has been identified, in one of its senses, is a statement that a principle has been applied correctly. This duplication has been permitted in order to present a more comprehensible list, and to make more apparent the conformity to the ordinary sense of *critical thinking*.

Under most of the aspects of *critical thinking* a list of criteria was presented; criteria which may be used in making the judgment called for. In all cases requiring judgments about the world of things, men, and events, these criteria must be applied with discretion; they are not to be applied mechanically. This is partly because numerous exceptions and qualifications which could have been mentioned were not mentioned in the interest of brevity; partly because a list of exceptions and qualifications would be endless, if we were to attempt to list them all; and partly because some difficult problems are still in need of further study.[14]

Finally, a logical analysis of the twelve abilities was made along three dimensions: the logical, the criterial, and the pragmatic. Appropriately weighted and empirically supported, such a simplification would suggest foci for evaluation and teaching.

Next Steps

Assuming that the foregoing range definition of *critical thinking* is acceptable, what are the next steps in research into evaluating and teaching *critical thinking*? General suggestions can be made; suggestions which require elaboration and adjustment for particular situations, and further adjustment as new information feeds back into the theoretical structure.

Evaluation

The list of aspects and criteria can be used to develop tables of specifications for critical thinking tests. Weighting decisions are yet to be made, as are deci-

[14] For example, the meaning of 'cause'; though on this problem Gasking's analysis (19) is very helpful.

sions about the levels at which the different abilities are possible. The latter decisions will require considerably more research than has yet been done in the area of readiness.

Checking the validity of these critical thinking tests should proceed along the lines advocated by Cronbach and Meehl in their article on construct validity (13). A check for agreement between the presented dimensional analysis and an empirical factor analysis would be worthwhile.

A check of certain hypothesized correlations with other traits would also be valuable, since construct validity depends on showing relationships with known variables, relationships that conform to theory-generated expectations. The following relationships are tentatively suggested. They are put vaguely, but necessarily so at this early stage.

Since all three dimensions appear to be learnable somewhat as academic subject matter is learnable, a smaller correlation is expected with I. Q. tests than is found between I. Q. tests. Since mental ability is obviously required for these dimensions, however, a sizeable correlation with I. Q. at a given age level is still expected. Of the three dimensions the logical dimension is expected to have the highest correlation with I. Q. at a given age level, followed by the criterial dimension and the pragmatic dimension in that order. Instruction in *critical thinking* should increase all of these correlations on the assumption that I. Q. tests measure potential to learn.

Since experience might be expected to strengthen one's pragmatic and criterial dimensions more than one's logical dimension after sixteen years of age, the former two are expected to correlate more highly with age beyond sixteen. The works of Burt, Moore, and Valentine suggest that logical ability appears early in life and develops in a fairly even manner (7, 33, 47). If this is so, the logical dimension may be expected to correlate substantially with age in elementary and secondary school. Of course the other dimensions should correlate substantially with age in elementary and secondary school also.

The correlations hypothesized above assume that there is some improvement of critical thinking without instruction. On the assumption that critical thinking is teachable, there should be greater improvement on all three dimensions if deliberate instruction is given.

There is probably a negative correlation between the degree to which a personality is authoritarian and the logical dimension, since the logical dimension is understood to include ability to judge not only what does follow and is implied, but what would follow, and what would be implied, if—. There is probably also a negative correlation between the degree to which a personality is authoritarian and the pragmatic dimension, since the pragmatic dimension is antithetical to all-or-none thinking.

Social-class status of students outside of the upper class, is probably correlated with *critical thinking*. No correlation is expected between *critical think-*

ing and political, social, and moral values, if the previously-mentioned factors are partialed out.

A construct validity judgment would also be helped by determining relationships between test scores and variables in methods of teaching. Thus experiments in the teaching of *critical thinking* would be helpful in the development of evaluation instruments and vice versa.

Instruction

We need to learn at what age students of various types can efficiently master the various aspects and criteria. Alternatively, or as a supplement, we need to know at what ages various types of students can master at various levels the three dimensions of *critical thinking*. It can not be said now which approach would be more fruitful.

We need to know in what curriculum patterns the aspects and/or dimensions are most effectively presented. Should they be integrated into existing courses or presented in a separate course? If integrated, which courses should be the vehicles? Or should all courses be involved, since *critical thinking* is needed in all areas?

What methods of teaching are most appropriate? Should the criteria of *critical thinking* be made explicit or left implicit? Do different groups need different approaches: boys and girls; social class groups; high, medium, and low mental ability groups? Must class size be kept at twenty-five, or can classes be increased in size—perhaps to hundreds or thousands? And how can teachers be prepared to teach their students to think critically? A course in logic? A course in *critical thinking*? A course in the logic of teaching? A special methods course? On-the-job training? A heavy dose of their subject-matter?

Since developments in evaluation and instruction complement each other, they must proceed apace. Let us hope that the concept of *critical thinking* presented in this essay will help us to push forward on both fronts.

Selected References

1. Beck, Lewis White. "Constructions and Inferred Entities." In Herbert Feigl and May Brodbeck (eds.), *Readings in the Philosophy of Science*. New York: Appleton-Century-Crofts, 1953. Pp. 368-381.
2. Black, Max. *Critical Thinking*. New York: Prentice Hall, 1952.
3. ———. "Definition, Presupposition, and Assertion." In Max Black, *Problems of Analysis*. Ithaca, New York: Cornell University Press, 1954. Pp. 24-45.
4. ———. "The Definition of Scientific Method." In Max Black, *Problems of Analysis*. Ithaca, New York: Cornell University Press, 1954. Pp. 3-23.
5. Braithwaite, R. B. *Scientific Explanation*. London: Cambridge University Press, 1953.

6. Bruner, Jerome S., Jacqueline J. Goodnow, and George A. Austin. *A Study of Thinking.* New York: John Wiley, 1956.
7. Burt, Cyril. "The Development of Reasoning in School Children," *Journal of Experimental Pedagogy,* V (1919), pp. 68-77, 121-127.
8. Carnap, Rudolph. "The Interpretation of Physics." In Herbert Feigl and May Brodbeck (eds.), *Readings in the Philosophy of Science.* New York: Appleton-Century-Crofts, 1953. Pp. 309-318.
9. Cohen, Morris R. and Ernest Nagel. *An Introduction to Logic and Scientific Method.* New York: Harcourt Brace, 1934.
10. Conant, James Bryant. *On Understanding Science.* New York: Mentor Books, 1951.
11. ——— (ed.). *Harvard Case Histories in Experimental Science.* Cambridge: Harvard University Press, 1950-55.
12. Crawshay-Williams, Rupert. *Methods and Criteria of Reasoning.* London: Routledge and Kegan Paul, 1957.
13. Cronbach, Lee J. and Paul E. Meehl. "Construct Validity in Psychological Tests," *Psychological Bulletin,* VII, 4 (July, 1955), pp. 281-302.
14. Dewey, John. *How We Think.* Boston: D. C. Heath, 1933.
15. Dray, William. *Laws and Explanation in History.* London: Oxford University Press, 1957.
16. Dressel, Paul L. and Lewis B. Mayhew. *General Education Explorations in Evaluation.* Washington, D. C.: American Council on Education, 1954.
17. Ennis, Robert H. "Assumption-Finding." In B. Othanel Smith and Robert H. Ennis (eds.), *Language and Concepts in Education.* Chicago: Rand McNally, 1961. Pp. 161-178.
18. Gardiner, Patrick. *The Nature of Historical Explanation.* London: Oxford University Press, 1952.
19. Gasking, Douglas. "Causation and Recipes," *Mind,* LXIV, 256 (Oct., 1955), pp. 479-487.
20. Gottschalk, Louis. *Understanding History.* New York: Alfred A. Knopf, 1951.
21. Guilford, J. P. *Fundamental Statistics in Psychology and Education.* New York: McGraw-Hill, 1950. P. 209.
22. Hall, Roland. "Assuming: One Set of Positing Words," *Philosophical Review,* LXVIII, 1 (January, 1958), pp. 52-75.
23. Hart, H. L. A. and A. M. Honore. *Causation in the Law.* Oxford: Oxford University Press, 1959.
24. Hart, H. L. A. *Definition and Theory in Jurisprudence.* London: Oxford University Press, 1953.
25. Hempel, Carl G. *Fundamentals of Concept Formation in Empirical Science.* Chicago: University of Chicago Press, 1952.
26. ———. "The Function of General Laws in History." In Herbert Feigl and Wilfrid Sellars (eds.), *Readings in Philosophical Analysis.* New York: Appleton-Century-Crofts, 1949. Pp. 459-471. Also in Patrick Gardiner (ed.), *Theories of History.* Glencoe, Illinois: The Free Press, 1959. Pp. 344-356.
27. Henderson, Kenneth B. and B. Othanel Smith. *A Study in the Teaching of Critical Thinking.* Urbana, Illinois: University of Illinois Press, in press (publication date indefinite).

28. Hood, Fred C. *Developing the Ability to Assess the Results of Thinking.* Urbana, Illinois: University of Illinois Bulletin, 1949.
29. Hospers, John. "What is Explanation?" In Antony Flew (ed.), *Essays in Conceptual Analysis.* London: Macmillan, 1956. Pp. 94-119.
30. Keynes, John Neville. *Formal Logic* (4th edition). London: Macmillan, 1906.
31. McCarthy, Philip J. "Sampling." Bulletin No. 15 of the New York State School of Industrial and Labor Relations. Ithaca, New York: 1951.
32. Morris, Charles W. *Foundations of the Theory of Signs.* Chicago: University of Chicago Press, 1938.
33. Moore, Thomas Verner. "The Reasoning Ability of Children in the First Years of School Life," *Studies in Psychology and Psychiatry from the Catholic University of America,* II, 2 (April, 1929).
34. Nagel, Ernest. "Some Issues in the Logic of Historical Analysis." In Patrick Gardiner (ed.), *Theories of History.* Glencoe, Illinois: The Free Press, 1959. Pp. 373-385.
35. Robinson, Richard. *Definition.* Oxford: Oxford University Press, 1954.
36. Scheffler, Israel. "Explanation, Prediction, and Abstract Entities," *British Journal for the Philosophy of Science,* VII (1956-7), pp. 293-309.
37. Scriven, Michael. "Definitions, Explanations, and Theories." In Herbert Feigl and Michael Scriven (eds.), *Concepts, Theories, and the Mind-Body Problem.* Minneapolis: University of Minnesota Press, 1958. Pp. 99-195.
38. ———. "Truisms as the Grounds for Historical Explanation." In Patrick Gardiner (ed.), *Theories of History.* Glencoe, Illinois: The Free Press, 1959. Pp. 443-475.
39. Smith, B. Othanel. "The Improvement of Critical Thinking," *Progressive Education,* XXX, 5 (March, 1953), pp. 129-134.
40. ———, "Logic, Thinking, and Teaching," *Educational Theory,* VII, 4 (October, 1957), pp. 225-233.
41. Stevenson, Charles L. *Ethics and Language.* New Haven: Yale University Press, 1944.
42. Strawson, P. F. *Introduction to Logical Theory.* London: Methuen, 1952.
43. Toulmin, Stephen Edelston. *The Philosophy of Science.* London: Hutchinson, 1953.
44. ———. "Probability." In Antony Flew (ed.), *Essays in Conceptual Analysis.* London: Macmillan, 1956. Pp. 157-191.
45. ———. *The Uses of Argument.* Cambridge: Cambridge University Press, 1958.
46. Urmson, J. O. "Parenthetical Verbs." In Antony Flew (ed.), *Essays in Conceptual Analysis.* London: Macmillan, 1956. Pp. 192-212.
47. Valentine, Charles W. *Intelligence Tests for Children* (6th edition). London: Methuen, 1958.
48. Wigmore, John Henry. *A Students' Textbook of the Laws of Evidence.* Brooklyn: The Foundation Press, 1935.
49. Will, Frederick H. "Generalization and Evidence." In Max Black (ed.), *Philosophical Analysis.* Ithaca, New York: Cornell University Press, 1950. Pp. 384-413.