

**QUERIES 'N THEORIES:
AN INSTRUCTIONAL GAME ON
THE DOT, DOT, DOT, . . . APPROACH TO SCIENTIFIC METHOD**

LAYMAN E. ALLEN

University of Michigan

ABSTRACT

QUERIES 'N THEORIES provides a parallel to the strong inference approach to scientific method – designing experiments, observing data, and theorizing. The reiterated use of the DOT approach (Design, Observe, Theorize) in the problem-solving required by the game mirrors the regular, systematic application of strong inference in some areas of science (e.g., high energy physics and molecular biology) that have moved ahead much more rapidly than others. Moreover, the game embodies and provides practice in two aspects of scientific theorizing and designing which John Platt has pointed out as central to scientific advance: (1) the usefulness of multiple hypotheses and (2) disproof as science's mode of advance.

Players of QUERIES 'N THEORIES assume the roles of a “native” and of linguists (“querists”) who are attempting to understand the native's language in a defined sense. Using strings of colored chips, the native secretly constructs the basic sentences and the replacement rules (if any) of his language. The querists ask questions by constructing queries (strings of colored chips) on the query mat. They seek to achieve an understanding of the language such that when the native in turn asks them about it by constructing queries on the query mat, they will be able to answer correctly. The goal is to ask the fewest number of questions necessary to achieve the specified understanding.

By presenting a programmed series of sample games that gradually and steadily increase in complexity, the author attempts to demonstrate that QUERIES 'N THEORIES offers its players the opportunity to learn a powerful analytic skill. He concludes that the examples are sufficiently persuasive to indicate the value of rigorous empirical investigation of a series of hypotheses about the merits of QUERIES 'N THEORIES as a vehicle for practicing and improving skill in the use of the strong inference approach to scientific method.

Observing and theorizing and designing experiments – and doing all three well – are the intellectual keys to scientific progress. In his classic article on scientific method, John Platt has convincingly set forth the case

for more widespread systematic application of man's most powerful approach to problem solving (Platt, 1964). However, to characterize the crucial processes of scientific research is one thing; to learn how to carefully observe, to imaginatively theorize, and to rigorously design is quite another matter. It is to this learning problem that the instructional game QUERIES 'N THEORIES (Allen et al., 1970) is addressed.

Convinced of the importance of examining the strong inference approach of science to determine whether other groups and individuals might learn to adopt it profitably in their own scientific and intellectual work, Platt observes that some fields of science are moving ahead very much faster than others. He observes further that in such rapidly advancing fields as molecular biology and high energy physics the techniques of strong inference are applied formally and explicitly and regularly and indicates his belief that this systematic application is the primary factor in such advances. Platt characterizes strong inference as follows:

In its separate elements strong inference is just the simple and old-fashioned method of inductive inference that goes back to Francis Bacon. The steps are familiar to every college student and are practiced off and on by every scientist. The difference comes in their systematic application. Strong inference consists of applying the following steps to every problem in science, formally and explicitly and regularly:

- (1) Devising alternative hypotheses;
- (2) Devising a crucial experiment (or several of them), with alternative possible outcomes, each of which will, as nearly as possible, exclude one or more of the hypotheses;
- (3) Carrying out the experiment so as to get a clean result;
- (1') Recycling the procedure, making subhypotheses or sequential hypotheses to refine the possibilities that remain; and so on.

An exact parallel to the strong inference which Platt describes above is called the DOT, DOT, DOT, . . . approach to problem solving in QUERIES 'N THEORIES. The 'D' in DOT stands for Design of experiments (Platt's #2, devising crucial experiments); the 'O' stands for Observation (Platt's #3, carrying out the experiment to generate new data for observation); and the 'T' stands for Theorizing (Platt's #1, devising alternative hypotheses). Calling it the DOT, DOT, DOT, . . . approach emphasizes that its power derives from the repetitive, regular, systematic application of its subsidiary steps. Or as Platt states it more eloquently:

. . . [W]e do not realize the added power that the regular and explicit use of alternative hypotheses and sharp exclusions could give us at every step of our research.

The difference between the average scientist's informal methods and the methods of the strong-inference users is somewhat like the difference between a gasoline engine that fires occasionally and one that fires in steady sequence. If our motorboat engines were as erratic as our deliberate intellectual efforts, most of us would not get home for supper.

Platt finds further testimony to the power of strong inference in the achievements of its use in organic chemistry and yet more in the work habits of some of the world's most respected scientists: Faraday, Fermi, Roentgen, Pasteur, Newton, and Maxwell. He emphasizes the theorizing and designing aspects of strong inference, while in *QUERIES 'N THEORIES* the importance of observing carefully and completely is equally stressed. To illustrate how *QUERIES 'N THEORIES* exemplifies Platt's pair of central themes about theorizing and designing (namely, (1) the usefulness of multiple hypotheses and (2) disproof as science's basic mode of advance), we turn now to a consideration of this game of science and language.

QUERIES 'N THEORIES is a game in the sense of Von Neumann and Morgenstern's definition (Von Neumann and Morgenstern, 1944) — namely, that a game is “the totality of the rules that describe it.” Within the limited and strictly defined universe imposed by the rules of any game, players are given opportunities to test and evaluate the results of many varieties of behavior.

In simple games, players are probably learning only how to play better the game in question; a good instructional game is one that makes it possible for players to try out varieties of behavior that may be useful to them in other contexts. One of the goals in designing *QUERIES 'N THEORIES* was to provide an occasion in which players are confronted in the context of playing the game with a graded series of problems, starting with simple ones and proceeding to ones of increasing complexity. The problems are of a type that furnish an opportunity to practice the use of the strong inference brand of scientific method and to test its usefulness and power in comparison with alternative approaches. Described somewhat more generally, *QUERIES 'N THEORIES* has been designed to develop a basic and uniquely human skill: asking good questions. In the game, players ask questions by constructing queries. But to construct good queries, a player must be a good theorist — skillful in organizing, analyzing, and synthesizing data. These processes of formulating theories and testing them by constructing queries are deliberately parallel to the methods used in science to probe the laws of Nature.

The hope, of course, is that *QUERIES 'N THEORIES* is a good instructional game in the sense that the skill and insight achieved from recurrent practice in theorizing and constructing queries in the game will

prove relevant for enhancing talent in scientific theorizing and experimental design. With the completion and availability of the game, that hypothesized and hoped-for relevance can now be tested

I do not at this time have data to offer from completed experiments that will disconfirm the hypothesis that the hoped-for relevance is absent. What I can do now is to describe in detail what happens in the play of the game so that you can bring your own intuitions to bear in judging the likelihood that the time-compressed experience in designing, observing, and theorizing which occurs in the play will be accompanied by improved performance in the use of scientific method. The question is straightforward and an interesting one: How likely is how much practice with the DOT, DOT, DOT, . . . approach to problem-solving in QUERIES 'N THEORIES to be accompanied by how much improvement in strong inference problem-solving in science?

The play of QUERIES 'N THEORIES involves one player's formulating a language and the other players' seeking to achieve an "understanding" of that language in a defined sense by asking questions of the language-builder (in the game he is called the "native") about the language that he has built. The goal is to ask the fewest number of questions necessary to achieve the specified understanding.

What is a language for purposes of QUERIES 'N THEORIES? One can begin to answer the question by giving some examples of languages in QUERIES 'N THEORIES. The following (L1) is a simple finite language that the native might have formulated by placing yellow (Y) and blue (B) chips in appropriate positions on the language mat.

Language Mat	
Basic Sentences	Replacement Rules
<u>Y</u> <u>B</u>	____ → _____
<u>B</u> <u>B</u>	____ → _____
____	____ → _____

A scientist seeking to understand the native's language (in the game he is called a "querist") is empowered by the game rules to perform certain experiments, the results of which provide data about the native's language. A querist can pose certain questions about the native's language; these the native must answer truthfully. The penalties in the scoring rules are very heavy for "native who speak with forked tongue!" The querist asks questions by constructing queries, that is, by placing colored chips on

the query mat. The following string of colored chips placed in row 1 would ask the question: “Native, is the string of chips consisting of a yellow chip and a blue chip, in that order, a sentence in your language?”

Query Mat	
Queries	Outcomes
<u>Y</u> <u>B</u> _____	1 _____
_____	2 _____
_____	3
	4
	⋮

The native would respond by placing a red chip (to indicate No) or a green chip (to indicate Yes) in the outcome column. In the case of language L1 the native would respond with a green chip.

L1 Y B _____ 1 G

If called upon to prove that yellow–blue was a sentence in his language, the native could do so by merely pointing to the basic sentences in his language and showing that yellow–blue is included there.

If a querist next asked, “Native, is the string yellow–yellow a sentence in your language?” what should the response be?

L1 Y B _____ 1 G
 Y Y _____ 2 R

That’s right, the response should be red, because yellow–yellow is not a sentence in language L1. Similarly, if the queries blue–blue and blue–yellow were constructed in rows 3 and 4, the responses by the native should be green and red, respectively.

L1 Y B _____ 1 G
 Y Y _____ 2 R
 B B _____ 3 G
 B Y _____ 4 R

No matter what other queries are constructed for L1, the responses should be red, because L1 is a finite two-sentence language containing only sentences YB and BB. Now that we have learned language L1 perfectly, let’s consider L2.

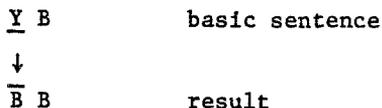
L2 Basic Sentences / Replacement Rules

$$\begin{array}{c} \underline{Y} \ \underline{B} \ / \ \underline{\quad} \ \underline{Y} \ \rightarrow \ \underline{B} \ \underline{\quad} \ \underline{\quad} \\ \underline{\quad} \ \underline{\quad} \ / \ \underline{\quad} \ \underline{\quad} \ \rightarrow \ \underline{\quad} \ \underline{\quad} \ \underline{\quad} \end{array}$$

Language L2 has only one basic sentence – namely, yellow-blue. But L2 has something that L1 does not have: a replacement rule. The single yellow chip to the left of the arrow with the single blue chip to the right of the arrow in the replacement rule in L2 means that anywhere that a yellow chip occurs in a sentence of L2, that yellow chip may be replaced by a blue chip and the result will be a sentence of L2. If the same four questions are asked of L2 as were asked of L1, the responses should be exactly the same.

L2	<u>Y</u> <u>B</u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u>	1	<u> </u> <u>G</u>
	<u>Y</u> <u>Y</u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u>	2	<u> </u> <u>R</u>
	<u>B</u> <u>B</u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u>	3	<u> </u> <u>G</u>
	<u>B</u> <u>Y</u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u>	4	<u> </u> <u>R</u>

If asked to prove his affirmative responses to #1 and #3, the native could do so by indicating: YB is a sentence in L2, because it is a basic sentence; and BB is a sentence of L2 because it is the result of applying the replacement rule, Y→B, to the basic sentence, YB.



No matter what other queries are constructed for L2, the response should be red, because L2 is a finite two-sentence language containing only the sentences YB and BB. Notice that L2 contains exactly the same two sentences as L1 contains, and no others. This means that L2 is really the same language as L1; they are merely different formulations of the same finite language, which contains only the two sentences YB and BB. Thus we see that the same language may be formulated in different ways.

Notice that each of these first two languages can be formulated on the language mat by the use of only four chips and only two different colors of chips. Both of these languages could be built by a native in what is called a 4–2 level game of QUERIES 'N THEORIES. The pair of numbers in a 4–2 level game signifies that the native should use at most four chips and at most two colors of chips in formulating his language.

4-2

: :

: signifies the maximum *number of colors* of chips that native can use in formulating his language

:

: signifies the maximum *number of chips* that native can use in formulating his language.

The first number indicates the maximum number of chips that can be used in formulating the language, and the second number indicates the maximum number of colors of chips that can be used.

Now, let us consider a more interesting language that can be built in a 4-2 level game, language L3.

L3 Y / Y → Y B

What are the appropriate responses to the following series of queries?

L3 Y 1 G

(because it is a basic sentence)

 Y B 2 G

(because it is the result of applying the replacement rule, Y→YB, to the sentence Y)

 Y B B 3 G

(because it is the result of applying the replacement rule, Y→YB, to the sentence YB)

 Y B B B 4 G

(because it is the result of applying the replacement rule, Y→YB, to the sentence YBB)

 Y B B B B 5 G

(because it is the result of applying the replacement rule, Y→YB, to the sentence YBBB)

How many sentences are there in language L3? (an infinite number). How would you characterize which strings of chips are sentences in L3 and which strings are not? (A string of chips is a sentence in L3 if and only if it begins with a yellow chip which is followed by zero or more blue chips.)

Is there a longest sentence in L3? (no)

How long may a string of chips be and still be a sentence in L3? (of any length.)

Hence, at even the relatively simple 4-2 level of game in QUERIES 'N THEORIES it is possible to generate interesting languages which contain an infinite number of sentences and in which sentences may be of any length.

Now, we are ready to move up a game level and consider a language (L4) that could be built in a 5–2 level game. Up to five chips and up to two colors of chips may be used by the native in building a language on the language mat.

L4 Y / Y → Y B B

What are the appropriate responses with respect to L4 to the same five queries that were addressed to L3?

L4	<u> Y </u> <u> </u>	1	<u> G </u>
	<u> Y </u> <u> B </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u>	2	<u> R </u>
	<u> Y </u> <u> B </u> <u> B </u> <u> </u> <u> </u> <u> </u> <u> </u> <u> </u>	3	<u> G </u>
	<u> Y </u> <u> B </u> <u> B </u> <u> B </u> <u> </u> <u> </u> <u> </u> <u> </u>	4	<u> R </u>
	<u> Y </u> <u> B </u> <u> B </u> <u> B </u> <u> B </u> <u> </u> <u> </u> <u> </u>	5	<u> G </u>

The proofs of the affirmative responses to #1, #3, and #5 are that (1) Y is a basic sentence, (3) YBB is the result of applying the replacement rule, Y→YBB, to the basic sentence, Y, and (5) YBBBB is the result of applying the replacement, Y→YBB, to the sentence YBB.

- | | |
|----------------------|--|
| (1) <u> Y </u> | basic sentence |
| ↓ | |
| (3) <u> YBB </u> | result of applying Y→YBB to the sentence Y |
| ↓ | |
| (5) <u> YBBBB </u> | result of applying Y→YBB to the sentence YBB |

How many sentences are there in language L4? (an infinite number)
Is there a longest sentence in L4? (no)

How long may a string of chips be and still be a sentence in L4? (of any length as long as the string contains an odd number of chips). How would you characterize which strings of chips are sentences in L4 and which strings are not? (A string of chips is a sentence in L4 if and only if it begins with a yellow chip which is followed by zero or an even number of blue chips.)

What is the relationship between language L3 and language L4? (L4 is included in L3; every sentence of L4 is a sentence of L3, but not vice-versa.)

With these four examples of languages in mind it is now more meaningful to answer the question: What is a language for purposes of QUERIES 'N THEORIES? In QUERIES 'N THEORIES a language is a set of sentences consisting of one or more basic sentences and zero or more generated sentences. A basic sentence of a language is a string of one or

two colored chips that appears as a basic sentence in the formulation of the language. A generated sentence of a language is the result of applying to a sentence of the language one of the replacement rules that appears in the formulation of the language. In QUERIES 'N THEORIES a language is formulated by the native by placing colored chips on the language mat to specify the basic sentences of the language and the replacement rules of the language.

In the play of QUERIES 'N THEORIES the native secretly formulates a language on the language mat, and the querists (there may be more than one) seek to achieve an understanding of the native's language by asking him questions about it that he must answer. Querists ask questions by constructing queries on the query mat. They seek to achieve an understanding of the language such that when the native in turn asks them about it by constructing queries on the query mat, they will be able to answer correctly. Querists seek to achieve this understanding by asking the fewest number of questions possible, because asking questions costs.

We are now in a position to illustrate the use of the DOT, DOT, DOT, . . . approach to coping with the kinds of problems that querists and the native are confronted with in the play of QUERIES 'N THEORIES. Suppose that the players are engaged in a 4-2 level game and the native has formulated a language on the language mat. Consider the problem that a querist is faced with. By observing carefully (the O phase of the first DOT) and considering the implications of what he observes, a querist can correctly impose some strong constraints upon his theorizing when he reaches that phase of coping with the problem. Observing that they are playing a 4-2 level game and considering the implications of that fact allows a querist to infer (1) that the native has used either one chip, or two chips, or three chips, or four chips in formulating his language and (2) that the set of chips used are either all blue, all yellow, or a mixture of the two. Observing the layout of the language mat and considering the implications of that layout along with the fact that the level of game being played is 4-2, he can correctly infer (3) that the basic sentences are either one chip long or two chips long, (4) that because there are only two blanks to the left of the arrow and three blanks to the right and there is a game rule that all replacement rules must have at least as many chips to the right of the arrow as are to the left of the arrow, the only kinds of replacement rules that can be built on the language mat are the following: $_ \rightarrow _$ or $_ \rightarrow _ _$ or $_ _ \rightarrow _ _ _$ or $_ _ _ \rightarrow _ _ _ _$ or $_ _ _ _ \rightarrow _ _ _ _ _$, and (5) that because at least one chip must be used to build a basic sentence, there are at most only three chips left to build replacement rules and so the only kinds of replacement rules possible in this 4-2 game are $_ \rightarrow _$ and $_ \rightarrow _ _$ and there is at most only one of these.

In moving to the next phase, the T (or theorizing) phase of coping with this problem of seeking to achieve an understanding of the native's language, it is useful to note the useful interaction between the preceding O phase and the thinking in the theorizing phase. A useful boundary is drawn around what theorizing is useful to do by the inferences from the observations in the O phase. The presence of such a limiting boundary is especially useful if the querist is employing strong inference in the manner that Platt counsels – with multiple working hypotheses to explain the data. The virtues of multiple working hypotheses will certainly become convincingly apparent to most experienced players of QUERIES 'N THEORIES. The argument for proceeding with multiple working hypotheses is especially strengthened when it is coupled with the limiting principle of only proceeding with hypotheses that are disconfirmable.

It is easy to agree with Platt, that these twin principles are strongly coupled to rapid scientific progress. Essentially, progress is achieved in this manner by conceiving of likely explanations of the phenomena you are interested in understanding and then designing experiments that will disprove one or more of those explanations. (The experiments sometimes also produce data that motivate researchers to conceive of additional explanations.) You advance to the good theory by disproving the bad ones. And that's really how science makes progress! (Popper, 1959) What's more, the fact that science never really proves anything in the sense of completely confirming one theory as totally sound, is part of what keeps it exciting and adventuresome. A helpful reminder of the nature of this process is embodied in a pair of severe but useful private questions that Platt suggests one ask (oneself):

on hearing any scientific explanation or theory put forward,
“But sir, what experiment could *disprove* your hypothesis?”,

or

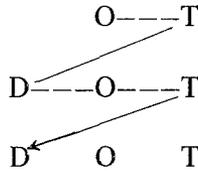
on hearing a scientific experiment described,
“But sir, what hypothesis does your experiment *disprove*?”

Polya also makes the same point in memorable language (Polya, 1954):
Nature may answer Yes or No, but it whispers one answer and thunders the other, its Yes is provisional, its No is definitive.

But limiting the multiple working hypotheses to those that are disconfirmable is not the only means of keeping multiple hypotheses within manageable limits; the implications of careful prior observations are also helpful in this respect. Multiple hypotheses are useful. But how many is multiple, or, perhaps more precisely, how multiple is most useful? Armed with inferences (1)–(5), a querist can easily ascertain that there are exactly ten different disconfirmable theories – and no more – to be explored in the questioning of the native. There are six possible different theories of languages consisting of various combinations of basic

sentences alone and four different theories of languages that are formulated by combinations of both basic sentences and replacement rules:

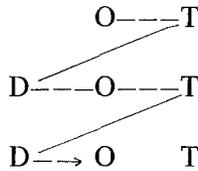
	<i>Theory that native's language consists of:</i>
T1 ___	one basic sentence (a singleton) and no replacement rules Possibilities: (Y) or (B)
T2 ___ ___	one basic sentence (a doubleton) and no replacement rules Possibilities: (BB) or (BY) or (YB) or (YY)
T3 ___, ___	two basic sentences (a pair of singletons) and no replacement rules Possibility: (Y, B)
T4 ___, ___ ___	two basic sentences (a singleton and a doubleton) and no replacement rules Possibilities: (B, BB) or (B, BY) or (B, YB) or (B, YY) or (Y, BB) or (Y, BY) or (Y, YB) or (Y, YY)
T5 ___ ___, ___ ___	two basic sentences (a pair of doubletons) and no replacement rules Possibilities: (BB, BY) or (BB, YB) or (BB, YY) or (BY, YB) or (BY, YY) or (YB, YY)
T6 ___, ___, ___ ___	three basic sentences (a pair of singletons and a doubleton) Possibilities: (B, Y, BB) or (B, Y, BY) or (B, Y, YB) or (B, Y, YY)
T7 ___ / ___ → ___	one basic sentence and a replacement rule with a singleton changing to a singleton Possibilities: (B/B→Y) or (Y/Y→B)
T8 ___ ___ / ___ → ___	one basic sentence (a doubleton) and a replacement rule with a singleton changing to a singleton Possibilities: (BB/B→Y) or (BY/B→Y) or (BY/Y→B) or (YB/B→Y) or (YB/Y→B) or (YY/Y→B) (Note that the other possible rules do not generate any new results.)
T9 ___, ___ / ___ ___	two basic sentences (a pair of singletons) and a replacement rule with a singleton changing to a singleton Possibilities: B, Y / ___ → ___ (Note that no rule generates any new results.)



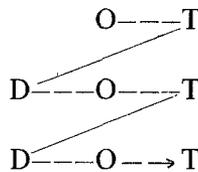
Consider the effects of building a query consisting of a single yellow chip alone as the second query in the context of what is already known from the results of the first query. If the outcome is green, then theories T1, T4, and T10 are disconfirmed; if the outcome is red, then theories T3, T6, T7, and T9 are disconfirmed. Therefore, the query is a good one to reduce the possibilities. Suppose that the outcome is red:

Y _____ 2 _____ R

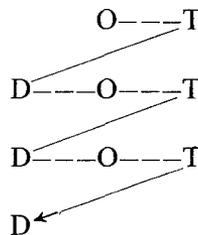
Observe (back to O again), what can the querist infer from this result?



He now knows (back to T again) that T3, T6, T7, and T9 are disconfirmed and that B is a basic sentence, because none of the only three theories that remain as possible explanations of the data generated so far (T1, T4, and T10) can have B as a generated sentence; so B must be a basic sentence.



What is a good question to ask next (back to D again) to disconfirm one or more of the remaining three theories?



Consider the effects of building BB as the third query in the context

of what is already known from the results of the first two queries.

If the outcome is green, then T1 is disconfirmed; but if the outcome is red then none of the three remaining theories are disconfirmed. Therefore, it is not a query that will definitely reduce the possibilities. Is there any such query in the circumstances? Notice that the queries BY, YB and YY all have the same limitations: a red outcome does not disconfirm any of the three remaining theories. In some circumstances there does not exist any query (of the type that has been considered so far) that will guarantee, no matter what the outcome to the query is, disconfirmation of at least one theory. Sometimes the best that can be accomplished by a query regardless of outcome is to reduce the number of queries that may have to be built before reaching one that will definitely disconfirm at least one theory regardless of outcome. This is one example of such circumstances. Each of the four queries above does disconfirm T1 if the outcome is green, but none of them disconfirms any of the three remaining theories if the outcome is red. And there is no other query of this type that will guarantee disconfirmation of at least one theory regardless of the outcome.

Suppose that the query built in fact is BB and that the outcome is green:

 B B _____ _____ _____ _____ _____ _____ 3 G

Observe (back to O again), what can the querist infer from this result? He now knows (back to T again) that T1 is disconfirmed and that the native's language is either (B, BB) of T4 or (B/B→BB) of T10.

The querist is now ready for a crucial experiment. What is a good question to ask next (back to D again) that will disconfirm one of the two remaining explanations of what the native's language is? Consider the effects of building BBB as the fourth query in the context of what is already known from the results of the first three queries. If the outcome is green, then T4 is disconfirmed; if the outcome is red, then T10 is disconfirmed. Therefore, it is a good query (it is the equivalent of a crucial experiment) to determine the language.

Whatever the outcome (suppose it is green), the querist will understand the native's language. When he feels confident that he understands the native's language well enough to answer correctly any question that the native can construct on the query mat, the querist should declare, "Aha," signifying his claim that he can answer all of the native's questions correctly. If the querist fails to answer all of the native's questions correctly, then the querist loses. If he does answer all of the questions correctly, then he wins, provided that the number of questions that he has asked in learning the native's language does not exceed some maximum

specified by the game rules. In detecting the native's language in this 4-2 level game the querist required only four queries to determine that it was (B/B→BB). In doing so he went through exactly four complete DOT cycles (that is, OT, DOT, DOT, DOT, DOT). Even in this relatively simple example, the sharp convergence obtained by reducing the number of theories compatible with the data through the repeated application of strong inference technique is apparent. In more complex examples, with more possibilities, the necessity for such convergence is even more apparent, and players are unlikely to miss the usefulness of strong inference methods to achieve it.

However, while the convergence necessary is affected by the complexity of language possible at the game level being played, its availability is influenced by the generality of questioning permitted by the game rules. At the very next game level it becomes apparent that the kind of questioning considered so far is extremely weak in detecting languages permissible at the 5-2 level – extremely weak in the sense that it may require many questions to detect the language with certainty. Suppose that the native formulated language L5 on the language mat:

L5 B / B → YYY

Although the querist may be asking as good a question as it is possible for him to ask on the basis of the information that he has available at each stage of the play, he may require 14 queries to generate enough data to understand the native's language with certainty if the questioning follows this sequence:

L5	<u> B </u> _____	1	<u> G </u> _____
	<u> Y </u> _____	2	<u> R </u> _____
	<u> B </u> <u> B </u> _____	3	<u> R </u> _____
	<u> B </u> <u> Y </u> _____	4	<u> R </u> _____
	<u> Y </u> <u> B </u> _____	5	<u> R </u> _____
	<u> Y </u> <u> Y </u> _____	6	<u> R </u> _____
	<u> B </u> <u> B </u> <u> B </u> _____	7	<u> R </u> _____
	<u> B </u> <u> B </u> <u> Y </u> _____	8	<u> R </u> _____
	<u> B </u> <u> Y </u> <u> B </u> _____	9	<u> R </u> _____
	<u> B </u> <u> Y </u> <u> Y </u> _____	10	<u> R </u> _____
	<u> Y </u> <u> B </u> <u> B </u> _____	11	<u> R </u> _____
	<u> Y </u> <u> B </u> <u> Y </u> _____	12	<u> R </u> _____
	<u> Y </u> <u> Y </u> <u> B </u> _____	13	<u> R </u> _____
	<u> Y </u> <u> Y </u> <u> Y </u> _____	14	<u> G </u> _____

When limited to asking specific questions of the sort, “Is this string of chips a sentence in the language?” the querist will in some cases obtain

very little data to narrow down the range of possible languages even though he is proceeding systematically and posing perfectly good queries on the basis of what information is available. This points up the need for game rules which will allow the querist to ask more general questions; for example, of the type, "Is there any sentence in the language that has this property?" Permitting this more general type of questioning is accomplished in QUERIES 'N THEORIES by introducing the use of the white chip to construct what are called "strong queries." A white chip stands for a string of chips (possibly empty) of unspecified length with colors of chips unspecified; it indicates an elision, very much like the use of three dots (. . .) in ordinary prose. Every empty space in a strong query, except those that precede an initial white chip or follow a terminal white chip, stands for a single chip of unspecified color. Consider what can be done with the availability of white chips for constructing strong queries.

Y W _____ 15 _____

This query asks the question:

Is there any sentence in the language that has the property of beginning with a yellow chip?

More briefly:

Y . . . ?

On the other hand,

_____ W Y 16 _____

asks the question, "Is there any sentence in the language that has the property of ending with a yellow chip?"; that is, ". . . Y?" More than one white chip can be used in a strong query,

W Y W _____ 17 _____

to ask, "Is there any sentence in the language that contains a yellow chip?"; that is, ". . . Y . . . ?" Blank spaces can also be used to advantage. (Recall that only those blanks preceding initial white chips or following terminal white chips are ignored.) The following strong query,

Y _____ W _____ 18 _____

asks, "Is there any sentence in the language at least two chips long that begins with a yellow chip?"; that is, "Y_ . . . ?" The following,

_____ W _____ B 19 _____

asks, "Is there any sentence in the language at least two chips long that ends with a blue chip?"; that is, ". . . _B?"

The availability of strong queries as well as specific queries permits a much more rapid convergence in narrowing the range of possible languages in a 5-2 (or higher) level of game. With the inferences from the responses in the following sequence of queries one can detect language L5 with just seven questions:

	<i>Queries</i>		<i>Outcomes</i>	<i>Inferences</i>
L5	<u>B</u> _____	1	<u>G</u>	(B) or (Y / Y → B)
	<u>Y</u> _____	2	<u>R</u>	(B), B ↯ Y
	<u>B</u> _____ <u>W</u> _____	20	<u>R</u>	BB:R, BY:R, B ↯ B...
	<u>Y</u> _____ <u>W</u> _____	18	<u>G</u>	(YB or YY or B → Y...) Y ↯ B
	<u>Y</u> _____ <u>W</u> _____	21	<u>G</u>	(B → Y...), B ↯ YY, i.e., (B → YYY or B → YB or B → YBY)
	_____ <u>W</u> _____ <u>B</u>	19	<u>R</u>	B ↯ YB
	<u>Y</u> <u>B</u> <u>Y</u> _____	22	<u>R</u>	(B / B → YYY)

Notice that in the context of the responses to queries 1 and 2 the response to query 20 provides all the data that was provided by the responses to queries 3, 4, 7, 8, 9, and 10. The green response to query 18 indicates that there is a sentence in the language that begins with a yellow chip, which must be so by its being a basic sentence or being generated by a rule. The green response to query 21 indicates that the sentence beginning with a yellow chip is generated by means of a rule – one of the following three: B→YB, B→YBY, or B→YYY. The red responses to queries 19 and 22 indicate that the rule is neither B→YB nor B→YBY; so it must be B→YYY. Hence, the language is: B/B→YYY.

By the time players reach the 6-2 level, the natives can build languages with multiple rules and languages that are universal or nearly universal. Consider, for example, language L6. It is a 6-2 universal language; every possible string of yellow or blue chips, or combinations of them, are sentences in L6.

L6 B / B→_aBB, B→_bY

(The 'a' and the 'b' under B→BB and B→Y are merely to indicate the names of the rules so that they can easily be referred to.) For example, BBYBYYB can be generated from the basic sentence in L6 by applying the rules in the following order: a a a a a b b b. Language L7 is almost a 6-2 level universal language. Every possible string of yellow or blue, except one, is a sentence.

occur in any of the languages formulated by T1.3 or T1.4 (for example, BB, BP, PB, and PP are sentences in the language B, P, Y/P→PP, P→B); so, T1.3 and T1.4 are also disconfirmed. Since any other usage of the five remaining chips than in one of the above four ways will generate even less doubletons, T1 is disconfirmed.

If the three singletons in the language are by virtue of T2, there are only nine possible theories that have just a pair of singleton basic sentences:

- T2.1 __, __ / __ → __
- T2.2 __, __, ____ / __ → __
- T2.3 __, __ / __ → __, __ → __
- T2.4 __, __ / __ → __, __ → ____
- T2.5 __, __ / __ → __, __ → __, __ → __
- T2.6 __, __, ____ / __ → __, __ → __
- T2.7 __, __, ____, ____ / __ → __
- T2.8 __, __ / __ → __, ____ → ____
- T2.9 __, __ / __ → __, __ → ____

However, there are no doubletons in any of the languages formulated by T2.1, T2.3, T2.5, T2.8, or T2.9. There are, at most, four doubletons in any languages formulated by T2.2 or T2.4 (for example, Y, B, BB/B→P). This disconfirms each of them leaving only T2.6 and T2.7 as possibilities.

Since Y is revealed to be a sentence in the Native's language, it must be a sentence by virtue of one of the three following possibilities:

T2.6a ... / B → Y, ...

T2.6b ... / P → Y, ...

T2.6c Y, ... / ...

Since YB is a sentence in the native's language and YY is not, there cannot be a B→Y rule in the formulation of the native's language. Similarly, since YP is a sentence in the native's language and YY is not, there cannot be a P→Y rule in the formulation of the native's language. Hence, T2.6a and T2.6b are disconfirmed.

The doubleton in T2.6c must begin with either B, P, or Y; hence, there are just the three following alternatives:

T2.6c1 __, Y, B ____ / ____ → __, ____ → ____

T2.6c2 __, Y, P ____ / ____ → __, ____ → ____

T2.6c3 __, Y, Y ____ / ____ → __, ____ → ____

Since YP is revealed to be a sentence in the native's language, and in T2.6c1 B — is the only doubleton that is a basic sentence, YP must be

directly or indirectly generated from B___ by either $B \rightarrow Y$ or by the pair of rules, $B \rightarrow P$ and $P \rightarrow Y$. However, it has already been shown above that neither $B \rightarrow Y$ nor $P \rightarrow Y$ can be in the formulation of the language; hence, T2.6c1 is disconfirmed. Similarly, T2.6c2 is disconfirmed, leaving only T2.6c3.

The other singleton in T2.6c3 is either B, P, or Y; thus, there are three possibilities:

T2.6c3A B, Y, Y ___ / ___ \rightarrow ___, ___ \rightarrow ___

T2.6c3B P, Y, Y ___ / ___ \rightarrow ___, ___ \rightarrow ___

T2.6c3C Y, Y, Y ___ / ___ \rightarrow ___, ___ \rightarrow ___

Because there are three singletons in the native's language, the pair of 1-to-1 rules of T2.6c3C must be $Y \rightarrow B$ and $Y \rightarrow P$. However, if that is so, the maximum number of doubletons there can be in any language formulated by T2.6c3C is three, because YY cannot be the basic sentence since it is not a sentence of the native's language. Since the native's language has six doubletons, T2.6c3C is disconfirmed.

Because YY is not a sentence in the native's language, the second chip in the doubleton of T2.6c3A and of T2.6c3B must be either B or P; thus, there are four possibilities:

T2.6c3A1 B, Y, YB / ___ \rightarrow ___, ___ \rightarrow ___

T2.6c3A2 B, Y, YP / ___ \rightarrow ___, ___ \rightarrow ___

T2.6c3B1 P, Y, YB / ___ \rightarrow ___, ___ \rightarrow ___

T2.6c3B2 P, Y, YP / ___ \rightarrow ___, ___ \rightarrow ___

In order to generate the singletons and doubletons revealed to be sentences of the native's language (i.e., B, P, Y, BB, BP, PB, PP, YB, and YP), the pair of rules to T2.6c3A1 and T2.6c3B1 must be $Y \rightarrow B$ and $B \rightarrow P$ and the pair of rules to T2.6c3A2 and T2.6c3B2 must be $Y \rightarrow P$ and $B \rightarrow B$. This results in the following four formulations of a language:

T2.6c3A1 B, Y, YB / $Y \rightarrow B, B \rightarrow P$

T2.6c3A2 B, Y, YP / $Y \rightarrow P, P \rightarrow B$

T2.6c3B1 P, Y, YB / $Y \rightarrow B, B \rightarrow P$

T2.6c3B2 P, Y, YP / $Y \rightarrow P, P \rightarrow B$

Each of these four alternatives formulates exactly the same language.

It is the language that contains just the revealed singletons and doubletons of the native's language; it does not have any sentences containing three or more chips. Hence, the language formulated by these four T2.6c3 formulations is compatible with the data revealed, and these four alone among the T2.6 formulations result in a language compatible with the data.

Similarly, the following pair of T2.7 formulations are the only ones compatible with the three singletons and six doubletons revealed to be sentences in the native's language:

T2.7a B, Y, BB, YB / B → P

T2.7b P, Y, PP, YP / P → B

These two T2.7 alternatives formulate exactly the same language as that formulated by the four T2.6c3 alternatives above. These six formulations are the only T2 formulations that are compatible with the native's revelations; they all formulate the same language, one which has as its only sentences those revelations.

Turning now to the T3 formulations, there are only four possibilities having just one singleton basic sentence.

T3.1 ___ / ___ → ___, ___ → ___

T3.2 ___ / ___ → ___, ___ → ___, ___ → ___

T3.3 ___, ___ ___ / ___ → ___, ___ → ___

T3.4 ___ / ___ → ___, ___ → ___, ___ → ___ ___

Since there are no doubletons in the languages formulated by T3.1 or T3.2, they are both disconfirmed.

There are two T3.3 formulations that are compatible with the three singleton and six doubleton revelations of the native's language, and they are the only compatible ones.

T3.3a Y, YB / Y → B, B → P

T3.3b Y, YP / Y → P, P → B

This pair of T3.3 alternatives formulates the same language that is formulated by the six T2 formulations that were compatible with the revelations.

This leaves one T3 theory to be considered, those formulations of the type indicated in T3.4. In considering T2.6, it was shown that neither $B \rightarrow Y$ nor $P \rightarrow Y$ could be part of a formulation of a language compatible with the revealed data. Since there cannot be any rules that will generate a singleton Y from some other singleton, the singleton basic sentence in T3.4 must be Y. With Y as the singleton basic sentence, there are just three ways for the other singletons to be generated by a pair of 1-to-1 rules.

T3.4a Y / Y → B, Y → P, ...

T3.4b Y / Y → B, B → P, ...

T3.4c Y / Y → P, P → B, ...

There are 27 different 1-to-2 rules to be considered as the other rule in T3.4a, but each of the 27 languages so formulated fails to match the revelations in some respect. Each of the languages formulated by adding $B \rightarrow YY$, $P \rightarrow YY$, or $Y \rightarrow YY$ generates the doubleton YY as a sentence, but it is not a sentence of the native's language according to the revealed data. Each of the languages formulated by adding $B \rightarrow YP$, $P \rightarrow YP$, or $Y \rightarrow YP$ fails to generate the doubleton PB as a sentence, and it is one of the revealed sentences. Each of the languages formulated by adding each of the other 21 possible rules fails to generate the doubleton YP as a sentence, and it is one of the revealed sentences. Thus, T3.4a is disconfirmed, leaving only T3.4b and T3.4c.

If $Y \rightarrow YB$ is the 1-to-2 rule of T3.4b, then the doubletons revealed to be sentences of the native's language will all be sentences in the language thus formulated, and only those doubletons will be sentences. Only two other rules considered as the 1-to-2 rule of T3.4b would produce these same results; these rules are $B \rightarrow YB$ and $P \rightarrow YB$. All other possible rules considered as the 1-to-2 rule of T3.4b formulate languages that either do not have all the revealed doubletons as sentences or have as sentences some doubletons that are not revealed as sentences.

Therefore, these three are all the possible T3.4b alternatives:

T3.4b1 $Y / Y \rightarrow B, B \rightarrow P, Y \rightarrow YB$

T3.4b2 $Y / Y \rightarrow B, B \rightarrow P, B \rightarrow YB$

T3.4b3 $Y / Y \rightarrow B, B \rightarrow P, P \rightarrow YB$

In the language formulated by T3.4b1 the following three-chip strings are sentences:

BBB	PBB	YBB
BBP	PBP	YBP
BPB	PPB	YPB
BPP	PPP	YPP

In general, any string of chips that has as its first chip a blue, a purple, or a yellow chip which is followed by any combination of blue or purple chips is a sentence in the language formulated by T3.4b1.

The language formulated by T3.4b2 and the language formulated by T3.4b3 have as sentences all those that are sentences of T3.4b1 plus some additional ones. Both formulate a language that has exactly the same additional sentences, so they both formulate the same language. The following three-chip strings, in addition to those in the language formulated by T3.4b1, are sentences in the language formulated by T3.4b2 and by T3.4b3.

BYB PYB YYB
 BYP PYP YYP

In general, any string of chips that is two or more chips long, that is comprised of any combination of blue, purple, or yellow chips, and that has as its last chip a blue or purple chip – any such string is a sentence in the language formulated by T3.4b2 and by T3.4b3.

Turning now to the languages formulated by T3.4c, there are just three 1-to-2 rules that result in languages that have exactly the right doubletons as sentences to match those that are revealed as sentences in the native's language. They are $Y \rightarrow YP$, $B \rightarrow YP$, and $P \rightarrow YP$; and the three accompanying T3.4c alternatives parallel those of T3.4b:

T3.4c1 $Y / Y \rightarrow P, P \rightarrow B, Y \rightarrow YP$

T3.4c2 $Y / Y \rightarrow P, P \rightarrow B, B \rightarrow YP$

T3.4c3 $Y / Y \rightarrow P, P \rightarrow B, P \rightarrow YP$

It turns out that all the strings of chips that are sentences in the language formulated by T3.4b1 are also sentences in the language formulated by T3.4c1, and vice-versa. So T3.4b1 and T3.4c1 formulate the same language. Similarly, all the strings of chips that are sentences in the language formulated by T3.4b2 and by T3.4b3 are also sentences in the language formulated by T3.4c2 and are also sentences in the language formulated by T3.4c3, and vice-versa in each case. So T3.4b2, T3.4b3, T3.4c2, and T3.4c3 all formulate the same language.

The analysis above covers all the possibilities, and the revealed data has narrowed the alternatives down to the following three languages:

- L1 the language formulated by T2.6c3A1, by T2.6c3A2, by T2.6c3B1, by T2.6c3B2, by T2.7a, by T2.7b, by T3.3a, and by T3.3b
- L2 the language formulated by T3.4b1 and by T3.4c1
- L3 the language formulated by T3.4b2, by T3.4b3, by T3.4c2, and by T3.4c3.

Notice that the string YBB is a sentence in L2 and L3, but not in L1; and that string YYB is a sentence in L3, but not in L2. Therefore, queries about these two strings will be effective in determining the native's language regardless of how he responds to the queries. If he indicates green to the query

Y Y B _____ 1 _____

then both L1 and L2 are disconfirmed and the native's language is determined to be L3. On the other hand, if he indicates red, then L3 is

disconfirmed and the second query must be posed to determine the native's language:

Y B B _____ 2 _____

If the response to this second query is green, then L1 is disconfirmed and the native's language is determined to be L2. If the response is red, then L2 is disconfirmed and the native's language is determined to be L1.

If it is not already apparent that players of this game have the opportunity to learn a powerful analytic skill by considering a programmed series of problems that gradually and steadily increase in complexity, and if the above 8-3 level example is not already persuasive of the high level thought that is necessary in order to play QUERIES 'N THEORIES effectively and the intimate relationship of that kind of thought with the strong inference approach to scientific method, then perhaps a final example will be the ultimate clincher. The example is drawn from a 10-5 level game; the querist who did the analysis is an experienced high school player. Knowing that the five colors of chips

Table I

<u>B</u>	<u>W</u>	---	---	---	---	1	<u>G</u>
<u>G</u>	<u>W</u>	---	---	---	---	2	<u>G</u>
<u>R</u>	<u>W</u>	---	---	---	---	3	<u>R</u>
<u>P</u>	<u>W</u>	---	---	---	---	4	<u>R</u>
<u>Y</u>	---	<u>W</u>	---	---	---	5	<u>R</u>
---	---	---	---	<u>W</u>	<u>B</u>	6	<u>G</u>
---	---	---	---	<u>W</u>	<u>Y</u>	7	<u>G</u>
---	---	---	---	<u>W</u>	<u>G</u>	8	<u>R</u>
---	---	---	---	<u>W</u>	<u>R</u>	9	<u>G</u>
---	---	---	---	<u>W</u>	<u>P</u>	10	<u>R</u>
---	<u>W</u>	<u>P</u>	<u>W</u>	---	---	11	<u>G</u>
---	<u>P</u>	<u>W</u>	---	---	---	12	<u>G</u>
---	<u>G</u>	<u>W</u>	---	---	---	13	<u>G</u>
---	<u>Y</u>	<u>W</u>	---	---	---	14	<u>G</u>
---	<u>B</u>	<u>W</u>	---	---	---	15	<u>G</u>
---	<u>R</u>	<u>W</u>	---	---	---	16	<u>R</u>
---	---	---	---	<u>W</u>	<u>P</u>	17	<u>G</u>
---	---	---	---	<u>W</u>	<u>G</u>	18	<u>G</u>
---	---	---	---	<u>W</u>	<u>Y</u>	19	<u>R</u>
---	---	---	---	<u>W</u>	<u>B</u>	20	<u>R</u>
---	---	---	---	<u>W</u>	<u>R</u>	21	<u>R</u>
<u>G</u>	<u>P</u>	<u>W</u>	---	---	---	22	<u>R</u>
<u>B</u>	<u>G</u>	<u>W</u>	---	---	---	23	<u>R</u>

AHA!

available in the game were B (blue), G (green), P (purple), R (red), and Y (yellow), and that the native's only revelation was Y, he constructed the sequence of queries in Table I.

As a test of your own skill in the use of the strong inference approach to scientific method, you might try your hand at determining from the data produced in the responses to these 23 queries what the native's language must be and why the data unerringly indicate it to be so.

As a further test, you might consider whether the language might have been determined in fewer queries, starting out the inquiry as this querist did. (Hint: With somewhat stronger theorizing after the response to query 9, the language could have been determined with just five more queries.)

I conclude that these examples are sufficiently persuasive to indicate the value of some rigorous empirical investigation of a series of hypotheses about the merits of QUERIES 'N THEORIES as a vehicle for practicing and improving skill in the use of the strong inference approach to scientific method. That, of course, is the task ahead. My hope, and hypothesis, should be unmistakably clear: that the results will convincingly show the unjustifiability of uniformly and uncritically associating games with connotations of triviality. May QUERIES 'N THEORIES prove to be a *good* instructional game – one that enables players to rehearse varieties of behavior that will be useful to them in other contexts. My wildest daydream is that QUERIES 'N THEORIES will become *the* counterexample that leaps to mind to help resolve an important controversy: for example, when somebody is deploring the chances of bringing Lord Snow's two cultures (Snow, 1959) closer together by upgrading the understanding of science among those who do not comprehend its processes, and a strong inference thinker inquires, "But sir, what experiment could *disprove* your hypothesis?"

References

- Allen, Layman E., Kugel, P. and Ross, Joan (1970), *Queries 'N Theories: The Game of Science and Language*. New Haven: Autotelic Instructional Materials Publishers.
- Platt, John R. (1964). "Strong inference." *Science*, 144:347–353.
- Polya, G. (1954), *Mathematics and Plausible Reasoning*. Vol. I: *Induction and Analogy in Mathematics*. Princeton: Princeton University Press.
- Popper, K. R. (1959). *The Logic of Scientific Discovery*. New York: Basic Books.
- Snow, C. P. (1959). *The Two Cultures and the Scientific Revolution*. London: Macmillan.
- Von Neumann, J. and Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.