UNDISCOVERED PUBLIC KNOWLEDGE

Don R. Swanson[1]

Knowledge can be public, yet undiscovered, if independently created fragments are logically related but never retrieved, brought together, and interpreted. Information retrieval, although essential for assembling such fragments, is always problematic. The search process, like a scientific theory, can be criticized and improved, but can never be verified as capable of retrieving all information relevant to a problem or theory. This essential incompleteness of search and retrieval therefore makes possible, and plausible, the existence of undiscovered public knowledge. Three examples intended to throw light on the logic of undiscovered knowledge are constructed and analyzed. The argument is developed within the framework of a Popperian or critical approach within science and on Popper's distinction between subjective and objective knowledge—the distinction between World 2 and World 3.

Imagine that the pieces of a puzzle are independently designed and created, and that, when retrieved and assembled, they then reveal a pattern—undesigned, unintended, and never before seen, yet a pattern that commands interest and invites interpretation. So it is, I claim, that independently created pieces of knowledge can harbor an unseen, unknown, and unintended pattern. And so it is that the world of recorded knowledge can yield genuinely new discoveries. In this paper I shall try,

1. University of Chicago, Graduate Library School, 1100 East 57th Street, Chicago, Illinois 60637.

[*Library Quarterly*, vol. 56, no. 2, pp. 103–118]

0024-2519/86/5602-0003$01.00

with the help of examples, to make such a claim at least plausible. The significance of these examples can best be understood if I first set the stage by providing a philosophical context.

Two important aspects of the Popperian philosophy of knowledge underlie my argument—first, the critical approach in science, and second, the distinction between objective and subjective knowledge. Although science serves as an example, the philosophy of knowledge that I describe embraces humanistic studies as well.

Criticism and the Growth of Knowledge[2]

The critical approach in science can be illuminated by first drawing a contrast. During the nineteenth and early twentieth centuries, science was exalted to a complete system of philosophical thought called "positivism." In that system, science was seen as the only valid and certain form of knowledge—certain because it was based on facts and on laws derived as generalizations from those facts. "Observed facts" and "direct observation" were enshrined as tokens of reality and as the building blocks of knowledge. Scientists, because they built their theories out of pure facts or observations, were presumed to be objective or "value free," as though they simply read the book of nature.

Volumes have been written about positivism, and I do not pretend that the above description is much more than an allusion, but it is enough to help me make clear some important contrasts with the critical method or critical approach.

Positivism, although not without influence even today, in every respect described above is a profoundly mistaken view of science. A key event in the assault on positivism was the publication in 1934 of *Logik der Forschung* by Karl Popper, translated into English in 1959 as *The Logic of Scientific Discovery* [1]. The positivist claim to "valid" or "verified" or "true" knowledge was overturned. Popper showed that scientific laws or theories cannot be induced or derived from observed facts or data; theories start from conjectures or free inventions. But these inventions confront, and may clash with, the real world. Such confrontation may therefore eliminate false or mistaken theories, and so by systematically testing and criticizing our theories we can learn from our mistakes.

2. This section is intended as a brief sketch of those aspects of Popperian philosophy, as I interpret it, that are central to this paper. My comments are influenced by various works by Popper; see esp. [1, pp. 13–61; 2, pp. 3–119, 215–48; 3, pp. 1–205; 4, "Introduction, 1982," and chap. 1; 5, pp. 46–181; 6, pp. 62–71, 143–49]. On positivism, see also [7, vol. 6, pp. 414–19].

But a theory that passes one or more tests cannot be presumed to be true, for it may contain hidden errors that eventually will come to light when other and more severe tests are designed. Because conjectures can go without limit beyond any previously observed data, and so could take forever to test under all possible circumstances, theories must remain forever conjectural. There is a crucially important duality then between the putative "real world" and our knowledge about that world. We are of course part of the world, and continually interact with it, but we cannot gain certain or true knowledge about it.

The heaviest burden that positivism laid upon the scientist was "objectivity." If the scientist indeed built true theories out of observed facts, then objectivity in observation must be the sine qua non of science. When it became obvious that scientists did not, and could not, live up to such an unrealistic expectation, science itself came under attack by those who failed to see how it differed from positivism. The mistaken idea was put forward that science was not a suitable model for the process of gaining knowledge in other disciplines, especially the humanities. Positivism is indeed an unsuitable model, and its influence should be attacked, but it is not science.

Scientists are of course neither objective nor value free. They choose and attack problems according to their interests, values, prejudices, and passions. Such an admission would demolish any illusion of science as "true" knowledge but is not damaging to science as a system of conjecture. For we must ask in any event, If theories are guesswork, what makes one guess better than another?

That question has a number of answers, and I cannot here do more than hint at their general nature. First of all, a guess that can be tested against the real world is better than one that cannot be tested. To be testable, a theory must have the potential for clashing with reality. Second, a guess that has been tested and passed the test is better than a guess that has failed the same test—other things being equal. More generally, a guess that has withstood critical argument by many scholars and scientists is better than one that has not. The role of criticism is therefore central to the process by which knowledge grows and becomes accepted.

Science is perhaps more systematically and self-consciously committed to a critical approach than other disciplines, but the critical approach is equally important for and applicable to theories about the world of social reality and the study of human action.

Scientists argue about reality even within the physical world because the "facts" of that world are not immediately available to our perception as the positivists believed. Rather, our perceptions are influenced by expectation. In a sense, we try to see what we look for—but we do not necessarily succeed. Expectation, theory, and pre-understanding can illuminate but cannot override reality. Reality wins out by eliminating—or refuting—erroneous expectations. The researcher can of course deny or evade refuting evidence, or even can hallucinate, but that is why criticism and corroboration by others are required.

An elegant analysis of how perception must be preceded by expectation is given by Popper, who argues that the human mind is not a bucket into which the data of reality are poured. It is more like a searchlight; theories and perceptions are freely invented and used to illuminate our experience with the world. The world may or may not accept the theory. If not, what is then important is that we can learn from our mistakes. A "mistake" is simply a clash between expectation and reality [3, pp. 341–61].

Scholars whose work is interpretive, such as in the humanities, may face especially difficult problems in evoking a clear confrontation with reality. But, ultimately, no one is free to ignore the real world. Neither physical reality nor social reality can be dismissed as purely a matter of opinion. (If there were no social reality we would not think of psychotics, for example, as being out of touch with it.) The question of what is or is not real can in part be fought out on the battleground of critical argument, with various scholars corroborating or correcting the perceptions of others.

Knowledge begins therefore with conjecture, hypothesis, or theory, all of which mean about the same thing. Scientific knowledge grows through testing and criticizing theories and through replacing theories with better ones that can withstand more severe tests and criticism. Thus knowledge is constructed of conjecture, and, though filtered through reality, remains forever conjectural. The sine qua non of science is not objectivity or even "truth," as is often thought, but a systematically self-critical attitude. Scientists are expected to propose testable theories and to be diligent in seeking evidence that is *unfavorable* to those theories. If they do not criticize their own work, someone else will.

Public criticism and published argument are crucial in helping the scientist create a product that can rise above his own prejudices and presuppositions. Although scientists and scholars may never be objective, the published products they create, shaped and weeded by criticism, can move ever closer to objectivity and truth.

Objective Knowledge—World 3

The word "knowledge" can be used either in an objective sense, as in the phrase "recorded knowledge" or "public knowledge," or in a subjective sense, in referring to knowledge in the mind of some person. It is the former, objective, meaning that I adopt unless otherwise indicated. The words "know," "known," and "knowable" I shall use only in a subjective sense.

The world of published knowledge certainly contains more than any one person can know and indeed contains more than the aggregate of what all persons know. The distinction between subjective and objective knowledge is most clearly brought out by Popper, who divides the world of reality, like all of Gaul, into three parts. World 1 is the physical world, World 2 is the world of subjective knowledge or experience, of mental states and mental processes, and World 3 is the world of objective knowledge—the world of problems, theories, and other products of the human mind. World 3 is real in that, through interacting with World 2, it can influence World 1, the ultimate standard of reality. That is, man can use theories and plans, grasped and understood subjectively, to modify the physical world [3].

World 3 is created by man, but it can give rise to problems not foreseen by its creators. Popper gives various examples of the idea of undiscovered World 3 knowledge; he points out that prime numbers, among other things, must have existed prior to their discovery. Man created the number system, but once it was invented, an infinity of unintended and unforeseen consequences, including prime numbers, then followed—and awaited discovery [3, p. 138]. World 3, while created by man, contains far more than man has ever thought of or dreamed about. World 3 indeed must contain ever increasing quantities of undiscovered knowledge.

The objective state of World 3, and what we subjectively know about World 3, are quite different concepts. So far as either certainty or "truth" is concerned, I shall show that World 3 must be in principle unknowable in the same sense that World 1 is unknowable. In either case, we cannot know, we can only guess. Only a small portion of World 3 is known to any one person. Some things we can perhaps know reasonably well, such as a theory that we ourselves have invented. But, once our invention becomes public knowledge—a bona fide resident of World 3—it takes on a life of its own. Someone else, somewhere else, without our knowing it might criticize that theory—and by so doing change it. In that sense we cannot claim to know with certainty even the objective knowledge that we ourselves create.

My aim in this paper is to explore the specific role that information retrieval might play in facilitating World 3 discoveries. With the aid of examples I shall argue that there must be vast areas of World 3 not yet discovered solely because of our limited ability to index, organize, and retrieve information. In that sense the undiscovered World 3 knowledge that I consider here is also public knowledge. At the same time, I shall show that information retrieval as a creative trial-and-error process, its limitations notwithstanding, must be a vital span in the bridge between World 2 and World 3 [8, 9].

Bertram Brookes has called our attention to the opportunity for exploration that World 3 offers; he stressed also the point that instant and total access to World 3 cannot be taken for granted [10, p. 130].

The Role of Information Retrieval in the Growth of Scientific Knowledge

The validity of a scientific theory can be understood only in terms of how it has been tested, criticized, and argued. From that state of argument we can learn whether the theory has been corroborated, refuted, or superseded, and how well it has stood up to critical attack. In testing and criticizing theories, and so in creating new knowledge, scientists and scholars interact with published knowledge and require therefore effective means for identifying and gaining access to that knowledge—that is, they require effective information retrieval systems and techniques.

I take here as the central problem of information retrieval, so far as the growth of scientific knowledge is concerned, the task of finding all published information that has significant bearing on the state of testing and of criticism of a given scientific theory or problem.

If one could by some means quickly and assuredly find all published information relevant to any given problem or theory, then the role of information retrieval in the growth of knowledge, even if vital, would be relatively uninteresting. I shall argue more fully at a later point that information retrieval is necessarily incomplete, problematic, and therefore of great interest—for it is just this incompleteness that implies the existence of undiscovered public knowledge. How knowledge can be public yet undiscovered will, I think, be made plausible by the following examples.

Example 1—Black Swans: A Hidden Refutation

The statement "all swans are white" can be thought of as the prototype of a simple scientific universal hypothesis or theory about the physical world—universal because of the unqualified "all." It is hypothetical and so can be no more than a conjecture. Though it can be put to the test by tracking down reports and rumors of nonwhite swans, it can never be verified—that is, established as true. One can verify that *some* swans are white, but not that *all* swans are white, for one cannot examine all swans at all possible locations in space and time. Although the swan hypothesis can never be verified, it can in principle be refuted by one good example (accepted, agreed upon, and reported by various competent observers) of a nonwhite swan.

To understand the role of information retrieval in the process of testing this hypothesis, our focus of interest will now shift from World 1, inhabited by swans, to World 3, inhabited by reports of swans. Instead of searching the far corners of World 1 for a nonwhite swan, we ask whether someone else has seen such a swan and described that observation in print. But short of examining every piece of information that has ever been published there is no way we can be sure of obtaining all reports that might be relevant.

Suppose for the sake of argument that scientists living in a remote part of the world were to publish, in a local wildlife journal, some observations about a family of black swans living on a nearby lake. We suppose further that the report comes from a half-dozen people who are reliable observers, and that they are unaware that other people in the world think that all swans are white. That is, the hypothesis about white swans, although published and perhaps even well known, is unknown to them. Their recorded comments about black swans were not made therefore as an intentional test of the hypothesis that all swans are white. Let us assume that those published comments were incidental to an article on some other topic, and, moreover, escaped the notice of indexers.

I have now described an imaginary situation that I think is of interest with respect to the role of information retrieval in the growth of knowledge. That is, a report of a family of black swans is presumed to be an acceptable and well-corroborated fact in the published literature, a fact that refutes a published hypothesis that all swans are white. But these two elements of information—the hypothesis and its refutation—may not both be known to any one person, at least not for some period of time, owing to the difficulties of identifying, locating, and gaining access to the relevant published information. Thus I have shown that it is possible for the refutation of a theory to stand, even if temporarily, as undiscovered public knowledge.

One cannot necessarily know when or whether someone, somewhere, has made the discovery in question—that is, has acquired it as subjective knowledge. But that question is cogent only in the context of influencing the future growth of objective knowledge. Subjective discovery is of course prerequisite to any further publication that takes account of the refutation in question. That is, to influence the future growth of knowledge, the hypothesis and its refutation would have to be brought together by a person who could understand and interact with both, a person knowledgeable enough to then publish a revision or retraction of the white-swan theory, or some logical consequence of its refutation.

Until that happens, the apparent refutation of the theory that all swans are white can have no influence on the future growth of knowledge. Here then is an example, albeit imaginary, of the vital role that information retrieval might conceivably play in the overthrow of an established theory—in particular in the ensuing process by which such overthrow might influence the creation of new objective knowledge.

Example 2: A Missing Link in the Logic of Discovery

A more complex, and perhaps more interesting, example will further illuminate the idea of undiscovered public knowledge. Suppose the following two reports are published separately and independently, the authors of each report being unaware of the other report: (i) a report that process A causes the result B, and (ii) a separate report that B causes the result C. It follows of course that A leads to, causes, or implies C. That is, the proposition that A causes C objectively exists, at least as a hypothesis. Whether it does or does not clash with reality depends in part on the state of criticism and testing of i and ii, which themselves are hypotheses. We can think of i and ii as indirect tests of the hidden hypothesis "A causes C."

In order for the objective knowledge "A causes C" to affect the future growth of knowledge—that is, in order for it to be tested directly, or to reveal new problems, solutions, and perhaps new conjectures, the premises i and ii must be known simultaneously to the same person, a person capable of perceiving the logical implications of i and ii. If the two reports, i and ii, have never together become known to anyone, then we must regard "A causes C" as an objectively existing but as yet undiscovered piece of knowledge—a missing link. Its discovery depends on the effectiveness with which information can be found in the body of recorded knowledge. Notice that "A causes C" is not something once known but forgotten; it is genuinely new knowledge that awaits discovery by explorers of World 3.

This abstract example can perhaps be made more convincing by means of a "live" example in the current scientific literature. Clinical tests reported in the biomedical literature have demonstrated a number of favorable effects on blood of dietary fish oil, effects generally attributed to eicosapentaenoic acid. Benefits that have been claimed include reduced platelet clumping, reduced levels of low-density blood fats, and reduced blood viscosity. All of these changes are thought to lower the risk of heart and artery disease. Letting "A" represent fish oil and "B" the reduction of platelet aggregability and blood viscosity, we can describe these publications in short as claiming "A causes B."[3]

3. The following five journal articles, selected from a larger number, offer evidence that fish oil reduces platelet aggregability and/or blood viscosity: [11–15].

There are also many articles on Raynaud's disease, a peripheral circulatory disorder of unknown cause and generally resistant to treatment. (The symptoms include episodic blanching of the fingertips and toes.) Several research groups have reported abnormally high blood viscosity and high platelet aggregability in Raynaud patients. It is plausible therefore to think that reducing blood viscosity would improve circulation and so at least symptomatically improve these patients. Moreover, several clinical trials have shown short-term beneficial effects from injections of prostacyclin, a powerful inhibitor of platelet aggregation. Thus we can again identify the effects "B" just as defined above—reduction of platelet aggregability and blood viscosity. Only now we can infer that "B" might cause improvement in Raynaud patients. Even if "B" only ameliorates the symptoms, or even if the effect is only temporary, we can still describe the logic of the situation by saying that "B causes C"—where "C" represents improvement for the patient.[4]

It follows, from the above two premises, that A causes C. Therefore the hypothesis implicitly exists that dietary fish oil might, at least in some cases, benefit patients with Raynaud's disease. But that hypothesis does not appear to have been published explicitly. None of the articles claiming that fish oil reduces blood viscosity and/or platelet aggregability mentions or cites in that connection any articles on Raynaud's disease, and none of the Raynaud articles mentions or cites the work on dietary fish oil, at least none that I have found so far. The two literatures appear to be remarkably isolated from one another with respect to common authors and mutual citations, among other things. We cannot know of course whether subjectively the hypothesis that dietary fish oil might benefit Raynaud patients has been formulated by one or more persons, but apparently no one has yet said so in print, so it is plausible to think that it might qualify as an undiscovered, but implicitly existing, hypothesis supported by indirect tests involving platelet clumping and blood viscosity.[5]

Example 3: The Hidden Cumulative Strength of Individually Weak Tests

It may so happen that the critical state of a theory is characterized, as is often the case in the health sciences, for example, by a series of indepen-

4. The following five journal articles, selected from a larger number, connect Raynaud's disease or syndrome with abnormally high blood viscosity and/or platelet aggregability: [16–20]. The role of platelets is indirect; see [21] for a more complete discussion.
5. My search for explicit connections between Raynaud articles and fish oil articles was reasonably diligent but far from exhaustive. I believe, tentatively, that there are no such connections, but I shall be glad to learn of any that I missed. A more complete account of example 2 is to be published in a biomedical journal [21] and so will reach an audience in a position to criticize my argument.

dently conducted tests, none of which is strong or convincing. Each alone may be so weak as to forgo any claim to surviving criticism on its own merits. It is possible however that, taken together, the aggregate of tests is stronger than any one of the tests. Any one test of a given theory may be weak if, for example, it admits of other theories or explanations that lead to the same evidence.

The degree to which a theory is corroborated depends on the severity of tests to which it is subjected. Popper defines the severity of a test in terms of the difference between the probability of the evidence, given the hypothesis, p(e,h). and the probability of the evidence without the hypothesis, p(e). That difference, p(e,h) − p(e), may also be called the weight of the evidence e in favor of the theory h.[6]

If the observed evidence follows logically from the hypothesis, then p(e,h) = 1, and the problem becomes that of estimating p(e)—the probability that the same evidence would be observed even if the hypothesis were false. Such an estimate would depend on the possibility, and the plausibility, of accounting for the observed evidence by explanations other than the hypothesis under test. The smaller the value of p(e), the stronger the test, because strength is computed by subtracting p(e) from p(e,h).

When the evidence is compounded from several independent experiments, the Popper formalism may be of particular interest in showing to what degree the joint probability of many different types of evidence is smaller than each of the separate probabilities. That is, it may permit a numerical estimate of the degree to which many individually weak tests of a theory can be combined into the equivalent of a much stronger test.

For the combined strength of several independently conducted weak tests to be recognized, the various tests must be brought together and understood by at least one person. Thus information retrieval can play a central role in enabling someone to know whether a scientific theory is or is not well tested. Such World 2 (subjective) knowledge is required as a point of departure for the further growth of objective knowledge in World 3.

We can presume that review articles often serve the purpose of assessing the significance of independent lines of evidence, each alone relatively unconvincing, that nonetheless point to similar and so more persuasive conclusions. The literature of the tobacco-cancer connection, for example, provides material for an interesting case study. There was a period many decades ago when fragments of evidence existed but there

6. Arguments in support of this idea of test severity, together with an improved, normalized, formulation of it, are given by Popper in a number of published papers, as well as in the following three monographs: [1, pp. 400, 410; 2, p. 391; 4, p. 240].

were no review articles, or any other scientific articles for that matter, that assembled these pieces. That fragmentary evidence came from case-control studies, cancer mortality statistics, animal experiments, and other data that can be considered as tests of the hypothesis that smoking may cause lung cancer. Each test taken alone is relatively weak in that other explanations can readily be adduced to account for the observed data. Yet, taken together, the combined evidence offers much stronger support for the hypothesis than was suggested or claimed in any scientific publication of that period—and so we may infer that the degree to which the lung cancer hypothesis had withstood tests was undiscovered.[7]

The Essential Uncertainty of Information Retrieval

Information retrieval is problematic solely because the quantity of published information is far larger than one person can read in any reasonable span of time. The ultimate, though unattainable, standard for insuring that no relevant information is missed is the direct inspection of all pieces of recorded information—a total exploration of World 3. This idea seems to be worth further reflection in the context of what I posed earlier as the central problem of information retrieval—the problem of finding all information related to the state of testing and criticism of a given scientific theory.

We circumvent total exploration by assigning or recognizing, for each piece of recorded information or "document," what we may call "points of access" or "searchable attributes" such as title words, index terms, descriptors, subject headings, or classification symbols. These attributes are no more than "handles" for selecting that document from a collection of documents, a collection too large for a total search. It is illusory to think that such handles can encode either the meaning or the relevance of a document with respect to all problems or theories to which it is logically related, especially to problems and theories not recognized or formulated at the time the document is created. No one has yet succeeded in demonstrating any formal relationship between relevance of this kind and search attributes. Retrieval of all information relevant to some given theory, or to the state of criticism of that theory, poses problems that are no less profound than trying to formalize human language, creativity, or inventiveness.

Any specific information retrieval request for all information related to some theory or problem entails an optimistic hope that may be cast as

7. This argument will be developed in a forthcoming paper (in preparation).

a universal hypothesis of the following kind: "All pieces of recorded information relevant to a given theory can be described and found by constructing some specific function of searchable attributes (the 'search function') associated with that theory." We now ask whether that search-function hypothesis can or cannot be established as true.

To verify that all relevant pieces of recorded information do in fact fit the description specified by a given search function, one would have to examine directly every piece of information that has ever been published. Moreover, such a task would never end, for, during the time it would take to examine even a small fraction of the material, more information would have been created. The above-stated hypothesis about a search function, in short, can never be verified. In that sense, an information search is essentially incomplete, or, if it were complete, we could never know it. Information retrieval therefore is necessarily uncertain and forever open-ended.

It follows that it is not possible to demonstrate the validity of any process for constructing search functions. There is no formal method of identifying searchable attributes, or any function of those attributes, that could guarantee retrieval of all relevant documents. Any search function is necessarily no more than a conjecture and must remain so forever.

The impossibility of verification amounts to a fundamental limitation that may be worth stressing, for the mistaken idea appears to be widespread that one can somehow, perhaps by constructing a search function thoroughly and carefully, insure that all relevant information will be found.

Although the hypothesis that a search function can retrieve all relevant information can never be verified, it can, in principle, be refuted or falsified. To falsify it, one need only produce one relevant piece of recorded information that does not have the specified combination of searchable attributes. Moreover, a search function can be criticized. One could, for example, invent a hypothetical relevant document with search attributes different from those on which the search function is based and so demonstrate that the search might fail.

The essential incompleteness of information retrieval perhaps becomes clearer if we think of starting with a scientific problem and then try to construct a search function for all solutions to that problem. This task is similar to a search for all criticisms and tests of a theory. Problems are solved by a trial-and-error process; there is no formal method for problem solving. Solutions are conjectures; whether they are good guesses or not depends on whether they work when put to the test. The same can be said of search functions.

No matter how adequate the results of a search may appear to be, there is no way to know whether additional solutions, or better solutions, remain undiscovered in the published literature. The outcome of any literature search for the solution to an empirical scientific problem, or for the state of criticism of that problem, is necessarily uncertain. In attempting to locate all information that may have some bearing on the solution to a problem, we are searching for what we do not understand rather than for what we do understand. A new solution to a problem may be based on a wholly new and different point of view. That is, not only do we seek what we do not understand, we often do not even know at what level an understanding might be achieved.

To better appreciate some of the difficulties in a problem-oriented search for information, consider a search request by an engineer who wants all available information on the mathematical analysis of how a child pumps a swing. I believe that even the well-informed layman, librarian, or experienced database searcher would probably not come very close to a good search function on the first try. Very little can be found by combining words related to pumping and to swings. A far greater quantity of relevant literature can be found by searching for articles and books on parametric amplifiers.

That example is far from unique and illuminates the difficulty of trying to construct any retrieval aid such as a thesaurus that would include a connection between parametric amplifiers and the pumping of swings. Of course an ad hoc or a post hoc thesaurus entry can solve any specific problem; the real problem is to build into the thesaurus, in advance, all relationships that people know about and then determine, for each search, which among those relationships are actually the ones needed. To build such a universal thesaurus entails no less than modeling all of human knowledge. Indeed, the thesaurus itself would have to contain a complete representation of World 3—the world it is designed to help search. To use such a thesaurus, one would have to retrieve relevant information from it, so a second universal thesaurus would be needed as a retrieval aid to the first, and so on ad infinitum. The builder of a thesaurus is, in principle, lost in an infinite regress.

Search functions and scientific theories have much in common. A search function is a conjecture or a theory about the contents of World 3, whereas a scientific theory is a conjecture about World 1, the physical world. Like universal scientific theories, search functions are falsifiable but not verifiable. There is no method for specifying how they can be invented, but that is less of interest than the question of how well they work once they are invented—that is, how effectively they can retrieve information. Most importantly, search functions and scientific theories are systems of conjectures, forever criticizable and forever improvable. In that sense there are no limits to either science or information retrieval. But then, too, there are no final answers.

Interactive Searching

It often happens, in science, that the most important consequence of solving one problem is that new problems then come to light. So it is with information retrieval. Because a search function can be criticized, it usually can be improved. The result of a literature search might then be of particular importance as a base for constructing a new and better search function. New search strategies may then emerge that could not have been foreseen, for they depend on having learned something about the current state of knowledge. The question itself may shift as new relationships are illuminated. Successful subject or problem-oriented information retrieval depends on a trial-and-error process, a process that, in principle, is interactive. The search function for each stage depends on the results of the preceding stage [8, 9, 22]. Here lies the significance of the online services—not in what the computer does in any single search, but rather in its facilitation of multistage interactive searching. Ironically, that multistage capability is not well exploited; most searching is on a "one-shot" basis [22, 23].

The Logic of Undiscovered Public Knowledge

I have argued that public knowledge may remain undiscovered solely because, like scattered pieces of a puzzle, the logically related parts that entail such knowledge have never all become known to any one person. The difficulties of information retrieval have delayed or prevented assembling the components.

My argument has depended heavily on three examples constructed to show what I meant by logically related but independent pieces of knowledge. These examples embody the logical relationships entailed in testing scientific theories. Examples 1 and 3 obviously do and are clearly linked to what I identified earlier as the "central problem of information retrieval"—finding everything related to the testing and criticism of a given theory. Example 2 is more complex; it is a syllogism consisting of three hypotheses—i: "A causes B"; ii: "B causes C"; therefore, iii: "A causes C." Now if we start with the concluding hypothesis iii, then i and ii can be thought of as indirect tests of that hypothesis. If iii is retrieved, but i and ii remain hidden, then example 2 resembles example 3. But in fact example 2 was designed to show something perhaps more interesting—namely how the hidden hypothesis itself, iii, could be revealed by retrieving and assembling the pieces i and ii.

The specific form that the "central problem of information retrieval" takes in any particular case depends then on which pieces of the puzzle, and therefore which logical links, are already retrieved and which are missing. In that sense, the central problem depends on what we might call "the logic of undiscovered public knowledge."

My three examples suggest that both information retrieval and the world of undiscovered public knowledge, while open-ended and limitless in their variety and complexity, yet have a certain order, form, and structure that may be worth systematic study. The growth of scientific knowledge is usually thought of in terms of wresting new discoveries from the physical world—World 1—admittedly a world that offers unlimited opportunity for discovery. But it should be of interest to librarians to notice that World 3 also qualifies as an endless frontier and to understand how and why this is so.[8]

REFERENCES

1. Popper, Karl R. *The Logic of Scientific Discovery.* New York: Harper & Row, 1968.
2. Popper, Karl R. *Conjectures and Refutations.* New York: Harper & Row, 1965.
3. Popper, Karl R. *Objective Knowledge: An Evolutionary Approach.* Oxford: Clarendon Press, 1975.
4. Popper, Karl R. *Realism and the Aim of Science.* Edited by W. W. Bartley III. Totowa, N.J.: Rowman & Littlefield, 1983.
5. Popper, Karl R. *Popper Selections.* Edited by David Miller. Princeton, N.J.: Princeton University Press, 1985.
6. Schilpp, Paul Arthur, ed. *The Philosophy of Karl Popper.* Library of Living Philosophers, vol. 14. La Salle, Ill.: Open Court, 1974.
7. *The Encyclopedia of Philosophy.* New York: Macmillan Publishing Co., 1967.
8. Swanson, Don R. "Information Retrieval as a Trial-and-Error Process." *Library Quarterly* 47 (April 1977): 128–48.
9. Swanson, Don R. "Libraries and the Growth of Knowledge." *Library Quarterly* 49 (January 1979): 3–29.
10. Brookes, Bertram C. "The Foundations of Information Science. Part I. Philosophical Aspects." *Journal of Information Science* 2 (October 1980): 125–33.
11. Terano, Takashi, et al. "Effect of Oral Administration of Highly Purified Eicosapentaenoic Acid on Platelet Function, Blood Viscosity and Red Cell Deformability in Healthy Human Subjects." *Atherosclerosis* 46 (1983): 321–31.
12. Goodnight, S. H., Jr., et al. "The Effects of Dietary ω3 Fatty Acids on Platelet Composition and Function in Man: A Prospective, Controlled Study." *Blood* 58 (1981): 880–85.
13. Nakagawa, Y., et al. "Effect of Eicosapentaenoic Acid on the Platelet Aggregation and

8. I do not mean to suggest that the discovery of new knowledge within the scientific literature rather than the laboratory is unprecedented or even unusual. My point is rather that the logical structure of unintended connections and the process of discovering such connections should be examined. Moreover, all scientific argument depends heavily on background knowledge that is grounded in the literature, and so any attempt to separate cleanly new laboratory-based knowledge from new literature-based knowledge is problematic. But the problems seem worth attacking.

Composition of Fatty Acid in Man: A Double Blind Study." *Atherosclerosis* 47 (1983): 71–75.

14. Cartwright, I. J., et al. "The Effects of Dietary ω-3 Polyunsaturated Fatty Acids on Erythrocyte Membrane Phospholipids, Erythrocyte Deformability and Blood Viscosity in Healthy Volunteers." *Atherosclerosis* 55 (1985): 267–81.
15. Woodcock, B. E., et al. "Beneficial Effect of Fish Oil on Blood Viscosity in Peripheral Vascular Disease." *British Medical Journal* 288 (February 1984): 592–94.
16. Blunt, R. J., et al. "Hyperviscosity and Thrombotic Changes in Idiopathic and Secondary Raynaud's Syndrome." *British Journal of Haematology* 45 (1980): 651–58.
17. Goyle, K. B., et al. "Abnormal Blood Viscosity in Raynaud's Phenomenon." *Lancet* 1 (1976): 1317–18.
18. Hutton, R. A., et al. "Assessment of Platelet Function in Patients with Raynaud's Syndrome." *Journal of Clinical Pathology* 37 (1984): 182–87.
19. Dowd, Pauline M., et al. "Treatment of Raynaud's Phenomenon by Intravenous Infusion of Prostacyclin (PGI_2)." *British Journal of Dermatology* 106 (1982): 81–89.
20. Belch, Jill J. F., et al. "Intermittent Epoprostenol (Prostacyclin) Infusion in Patients with Raynaud's Syndrome." *Lancet* 1 (1983): 313–15.
21. Swanson, Don R. "Fish Oil, Raynaud's Syndrome, and Undiscovered Public Knowledge." *Perspectives in Biology and Medicine*, vol. 30 (Autumn 1986), in press.
22. Harter, Stephen P. "Scientific Inquiry: A Model for Online Searching." *Journal of the American Society for Information Science* 35 (March 1984): 10–17.
23. Harter, Stephen P. "Online Searching Styles: An Exploratory Study." *College & Research Libraries* 45 (July 1984): 249–58.