Seek and Ye Shall Find: How Curiosity Engenders Discovery

Herbert A. Simon

Department of Psychology
Carnegie Mellon University
Pittsburgh, PA 15213

30 June 1998

Complex Information Processing
Working Paper #542

Seek and Ye Shall Find: How Curiosity Engenders Discovery

Herbert A. Simon
Departments of Psychology and Computer Science
Carnegie Mellon University

In this chapter I will address a crucial aspect of scientific discovery that has to be kept in the center of attention if the children in our schools are to achieve a better understanding of science and of the ways of doing science, and — at least as important — if they are to acquire and retain a strong desire to use these ways to cope with, to understand, and to enjoy the world around them, physical, biological, and social.

Philosophers of science have been prone to divide scientific activity into two parts: the processes of discovery and the processes of verification. In fact, the literature of philosophy of science and scientific methodology has generally given far more emphasis to the verification than to discovery. "Somehow," theories and hypotheses appear on the scene, and these are then tested to determine whether they should be accepted or rejected. Courses on research method focus almost exclusively upon how experiments should be designed to test given hypotheses, and how statistical tests should be applied to evaluate the outcomes of such experiments.

On the other hand, processes of discovery have engendered much cognitive and developmental research, as witness the many interests in these processes that are represented in this conference. One important theme in the developmental research is that children are born with curiosity, which, if maintained and even stimulated, leads them to enjoy and understand science (and sometimes become scientists). I will explore this theme and use it to critique some widely held views about research methodology that are sometimes taught in science curricula, but that might actually act to stifle curiosity, hence impede science education. My purpose, of course, is not just to remove from our teaching what is harmful, but to review some of the things we know about discovery that are perhaps both teachable and worth teaching.

Curiosity

We start with the widely accepted generalizations that learning depends on time on task, that time on task requires sustained attention, and that attention can only be sustained if the stimuli are interesting, or, in causal terms: *Interesting stimuli —> attention —> time on task —> learning.*

Daniel Berlyne (1960) showed, with the help of extensive experiments with children, adults and rats, that attention is sustained longest if the stimuli are neither too simple (boring) nor too complex (inscrutable). Our attention can be sustained as long as we can continue to find interesting pattern in the stimulus. What is just complex enough depends on what we already know. After we have heard lots of music, the simple tunes that initially fascinated us no longer retain our attention very long; whereas music that was initially experienced as a jumble of raucous sound can now hold our attention for hours as we begin to discover the pattern buried in it. We pass gradually, one might say, from nursery songs to Bartok, if not, perhaps, to John Cage.

Children, left to themselves in a rich environment find, and attend to, stimuli that are at the right level of complexity for them — in which they can find interesting pattern. With experience, they learn to discover and enjoy more and more complex patterns. We say that they have curiosity, and we are concerned that this curiosity seems often to be burned out of them in the process of growing up and being schooled.

Although I know of only a little research that supports (and none that refutes) my conjecture, I would guess that curiosity — the habit of examining the environment for interesting pattern — can be learned. I would venture the further guess that a reasonably rich environment, but one that does not continually force new stimuli on children, instead leaving to them the initiative in seeking pattern, is most conducive to encouraging active curiosity. I would venture a third guess that the best environments for this purpose react to the child's exploration of them (provide feedback that helps reveal pattern) and do not simply present unmotivated change and variety (Qin & Simon, 1990). (These guesses would be at the core of my criticism of TV fare, which egregiously violates the second and third conditions.)

It has often been remarked (perhaps even shown empirically) that curiosity — wonder at all sorts of surrounding phenomena — is a common characteristic not only of young children, but also of good scientists. It is not unreasonable to suppose that this has something to do with their capability for doing good science, but it's not easy to run an experiment in standard form to prove this.

Experimental Method

Which gets me to the other half of my thesis: What is an experiment in "standard form"? And does skill in running such experiments produce good science?

Standard textbooks on scientific methodology, and on experimental method in particular, place controlled experiments (and especially critical experiments) and the center of the scientific stage. To design a controlled experiment, one needs a hypothesis: say, the hypothesis that the presence or absence of a particular condition causes a particular effect to occur or not. Following Mill's canon of difference, the underlying logic is that: "If an instance in which a phenomenon occurs, and one in which it does not, differ in only one other circumstance, this circumstance is the cause, or the effect, or an indispensable part of the cause, of the phenomenon." What you get out of a controlled experiment is some evidence that influences you (or should) to have somewhat more or less credence in your particular hypothesis than you had before.

Of course, to make this particular rabbit stew, you have to catch the rabbit — to produce an effect. So you set up an "experimental" condition, in which the presumptively causal condition is present and a "control" condition in which it is not, and you measure the level of the phenomenon of interest in each condition. If the occurrence of the phenomenon is more frequent or more intense in the experimental condition than in the control condition, then it is a causal factor, otherwise it is not.

But how much more frequent or intense must it be? Nowadays, we are all sophisticated about random variables in the air and the wicked way in which they infect data. We require "significance" — that is "statistical significance" of our differences, which brings us into the whole tangled and badly understood morass of modern statistics. I'll say something later about what scientists do when they don't know statistics. For the moment, let's ignore this particular complexity, for there is another even more important one that hides in the usual notion of experiment.

To run the orthodox experiment, you have to have a hypothesis: that some specified variable is the causal agent for some specified phenomenon. You have to have a conjectured law of nature. It is generally regarded as bad form (often referred to as "counting the bricks in a wall") to run an experiment without a hypothesis. And that leads to the question of where hypotheses come from. A theory of the design of experiments and of statistical tests of the validity of hypotheses obviously does not fill this bill, for, according to orthodox doctrine, the design, execution and evaluation of experiments can only occur after the hypotheses are already in place. If running experiments to test hypotheses lies at the core of science, then a theory of how science works must include a theory of the origin of hypotheses. "Discovery" is the usual term for the crucial missing piece: we need a theory of scientific discovery as well as a theory of verification.

(A "critical" experiment is simply a controlled experiment in which *two* hypotheses are already proposed, one of which is consistent with one outcome of the experiment, the other with another outcome. Thus, in the famous Michelson-Morley experiment, if the velocity of light were independent of the direction of the reflected beam relative to the Earth's motion, then the hypothesis that the Earth moved through an ether was refuted; if the velocity depended on the direction of the beam *(to a specific predicted extent)* then the existence of an ether was supported. If the velocity depended on the direction, but to a different extent, then all bets were off — or were they?)

Scientific Discovery: The Origins of Hypotheses

We all "know" where hypotheses come from: We read last month's journals, find reports of some experiments, decide that if we did the experiments with better controls or with more subjects or with different instructions they would come out differently. Of course that explains some circumstances (probably not very productive ones) in which we think up hypotheses, but not what the process is of thinking them up. Perhaps they simply well up by intuition, insight or creativity from that mysterious region known as the subconscious — itself not a very interesting hypothesis, or one easy to test.

Is there a better answer to the question of the origins of hypotheses? Let's look at some evidence from Nobel-level science.

There's the notorious case of Fleming, returning from vacation to be greeted by a stack of dirty Petri dishes, which he had left unwashed in his laboratory sink. On one, he notices some very sick (lysed) bacteria, and just next to them some mold, which he recognizes as belonging to the genus Penicillium. Now he forms a hypothesis: the bacteria are dying because of something the mold is excreting.

Where does the hypothesis come from? Notice and recognize two words in the second sentence of the last paragraph: the words "notice" and "recognize." Forming an hypothesis, at least in this case, depends on noticing some phenomena, and recognizing some things about them. Both the noticing and the recognizing depend on prior knowledge: the noticing, because we mainly notice things that are unusual or surprising in their current surroundings; the recognizing, because there is very little we can say about things that we don't even recognize. (Although noticing something that is unrecognizable in a context where everything should be familiar is itself a source of hypotheses.)

Fleming is surprised, and surprise occurs only when we have expectations that are violated. Out of that violation of expectations, a hypothesis arises. In this case, Fleming's knowledge suggests no reason why the bacteria should be lysing, but they are lysing: hence, he hypothesizes that the presence of the mold is a cause of the lysis of the bacteria. (His knowledge also suggests to him that, if the hypothesis is true it is important, for lysing bacteria is something we humans would like to be able to do in the interest of our health.)

So hypotheses, at least in this case, come out of observations that occasion surprise; and to make such observations, there must be phenomena to observe. If so, perhaps our account of scientific discovery should begin, not with the experiment, but with the processes of observing phenomena (of a familiar kind) that may produce surprises.

Is Fleming's case an isolated curiosity? Decidedly not. I have identified more than a half dozen others that resulted in Nobel prizes, and I am certain that my list is very incomplete. I'll just mention Planck, the Curies, Krebs, Roentgen, Penzias and Wilson, and come to think of it, myself. As Nobel prizes have been awarded in the hundreds, but not the thousands, this is not an uninteresting phenomenon, although I do not claim it is the whole story.

More generally, hypotheses arise from exposure to phenomena, with or without surprise. I'll return to that in a moment.

Are there implications for education? Train children not to wash the dishes? That may not be the right lesson to draw from the example. Perhaps more important for Fleming is that he was knowledgeable: He had put in his ten years acquiring expert knowledge about molds and bacteria. In this domain, he was Pasteur's "prepared mind" to whom accidents happen. But "knowledgeable" is not enough either: he was also curious. He did not say, "Let's clean up that damned mess and get on with it." He asked, "What is causing that curious phenomenon?" Perhaps another observer, with a different background of knowledge, would observe, while counting the bricks in a wall, something unusual about their texture, or that alternate bricks in every third row were cracking. For the right person, that could be his or her path to the Nobel.

Observing Phenomena: The Case of Faraday

Letter of Michael Faraday to Richard Phillips, 29 November 1831
(on the occasion of his great discovery of magnetic induction of electricity)

It is quite comfortable to me to find that experiment need not quail before mathematics, but is quite competent to rival it in discovery.

There is no better example of observation driven by curiosity as a wellspring of science than the case of Michael Faraday, who was responsible for laying the empirical, and to a large extent, the theoretical, foundations of electromagnetism. Electricity and magnetism were subjects already steeped in mathematics by such imposing figures as Coulomb and Ampere at the time (about 1820) when Faraday began working in them. Faraday had no mathematics at all, nor ever acquired any. Such theories as he created were verbal and visual theories, particularly the latter; yet they included the basic laws of magnetic and electrical fields, later mathematized by Maxwell.

To emphasize, as I do here, Faraday's observations rather than his theories is not to discount his interest in theory and capabilities for generating it. Rather, it is to show how, in making his most important discovery, of magnetic induction of electricity, observation led the way to theory, rather than deriving from it. A second theme is to show that "theory" means many things, and that the "theories" that guide experimentation may be of a much vaguer and looser form than those that are discussed in the literature on experimental method and that are "tested" by critical experiments.

In its simplest terms, the story is this (Williams, 1965; Magnani, 1996). The world of science was electrified (literally) in 1820 by Ørsted's demonstration that an electric current induced a magnetic field around it. Until that time, electricity and magnetism were completely distinct phenomena, although similarities in their laws of attraction and repulsion had, of course, attracted attention. Ørsted, starting out with no hypothesis more definition than that all forces in nature should be related, simply put a magnet near a live electric wire, and noticed that the needle was deflected. The experiment was immediately repeated and extended all over Europe.

Faraday, upon learning of Ørsted's finding, also formed a hypothesis, based on a principle as vague as Ørsted's: that if electricity could induce magnetism, then, by symmetry, magnetism should be able to induce electricity. Beginning in 1821, he periodically conjured up observational situations that he thought would produce such induction. They generally involved bringing magnets and electricity into some kind of spatial proximity. They were also generally unsuccessful, as were those of other investigators playing a similar game. (The story is a little more complicated, Arago produced a partial effect that later proved important, but no one at first could make heads or tails of it.)

In 1831, on learning of the great progress that had been made by Joseph Henry in America and by Moll in the Netherlands, in producing powerful electromagnets, Faraday decided to try with one of the new magnets essentially the same observations he had tried with weaker magnets, and without success, in 1825. Without specific reasons for acting in this order, on this occasion he closed the circuit, B, in which it was hoped a current might be induced, before he closed the magnet circuit, A, that was connected to the battery; and when the latter was closed, a momentary current was detected in circuit B. In all his previous experiments, it had happened that circuit A was closed before circuit B, and no current was detected.

Faraday quickly constructed a "theory" of the phenomenon he had produced: the battery current and its induced magnetic field created both an induced current and an "electrotonic state" in circuit B. The latter resisted the (steady) current that would otherwise have been induced in B, and permitted only a momentary transient. When the battery circuit, A, was opened, the pressure that had created the "electrotonic state" in B was removed, and a transient in the other direction occurred. Although Faraday tried to find other evidences, physical and chemical, of the hypothesized "electrotonic state," he did not succeed; but he only gave up the hypothesis, when he arrived at a more satisfactory (and operational) one some months later.

Within a month of this first induction of a surprising transient current, Faraday had extended his initial surprise to new observations in which he produced a continuous current by spinning a copper disk between the poles of a powerful magnet (an idea stimulated by his knowledge of Arago's experiment, mentioned earlier). In the course of his successive observations, he had to modify his theory of the phenomena he was producing, and by the end of the year, he had created his famous concept that current was produced when a wire cut the "lines of magnetic force" of a "magnetic field." The first of these quoted phrases had been occasionally used earlier by Faraday, on the basis of his observation of iron filings scattered on paper over a magnet, but these concepts essentially grew out of observations, and did not precede them. They gradually replaced the earlier "electrotonic state."

Did Faraday Perform an Experiment?

Let me raise next the question of whether what I have been calling Faraday's "observations," which led him to conceive his theory of magnetic fields, were or were not experiments in the modern sense. What was the experimental condition and what was the control in the crucial first experiment of 1831? One might suppose that the experimental condition was that in which the battery circuit, A, was closed and the control condition that in which it was open. But the effect (the momentary current in B) did not occur during the former condition, but only at the moment of closing (or opening) circuit A — a "condition" that, as far as we know, Faraday had not even conceived when he set up the experiment.

If we insist on calling this an experiment, then the phenomena observed created its design rather than the design creating the phenomena. Why not simply say that Faraday set up, over his ten years of intermittent curiosity, a whole series of situations in which he could observe various phenomena, and on the last occasion (through the accident of the order in which the two circuits were closed) he obtained an interesting effect — his curiosity was rewarded? Does this demean Faraday? Does it mean he was a mere brick counter?

Something even worse must be said about the crucial experiment. Faraday did have a galvanometer mounted in the second circuit, in order to measure any current that might be produced in it, but in the crucial event he provides no numerical reading of the galvanometer. The entry in his lab diary simply records: "immediately a sensible effect on needle. It oscillated and settled at last in original position. On *breaking* connection of A side with Battery again a disturbance of the needle." He uses essentially the same statement in his published report of the event — the event that paved the way for the dynamo — and the referees of the Proceedings of the Royal Society did not blink.

If we see almost no numbers in this earth-shaking first paper, then of course we see no tests of statistical significance. Does that mean that scientists of that time were not aware that spurious effects might be produced by currents of air in the laboratory or dirty circuit connections or a thousand other unintentioned irrelevancies? On the contrary, Faraday's diary is full of comments about precautions he took and possible causes of error in his interpretations. He simply felt no urge to quantify these disturbances. Instead, his reaction, and that of his contemporaries (and of most scientists today), was, when he obtained an effect he thought interesting, to consider how he could change the situation to magnify it. If you can magnify an effect enough it becomes significant — by any criterion. Before he communicated the first exciting finding, he worked for more than four months, successfully, to greatly magnify and clarify the original phenomena.

Hence, that first event was not all that he reported in his initial paper (which appeared in print, by the way, within about three months of submission). He reported the whole sequence of subsequent observations made over about four months. The success of his first act of curiosity suggested variants of the situation that gradually enabled him to magnify the effect in which he was interested and to transform a transient effect into a continuous one. Each change produced new phenomena and each phenomenon suggested new changes. The reports of these experiments occupy nearly fifty pages of the printed edition of his laboratory diary.

This does not sound like an inexorable logic of experimentation that begins with hypotheses, created in some unexplained way; leads to logically deduced predictions from these hypotheses; thence to controlled, and even critical, experiments; then to the torture chamber of significance tests; thence to acceptance of the hypotheses or oblivion. It sounds like a much more rambling (if highly exciting) walk through the woods in search of mushrooms, or rare flowers, or phenomena of any kind that excite surprise, or the sense of beauty, or wonder, or puzzlement — and motivate the search for more.

So perhaps we should say that Faraday did not perform experiments; he followed his curiosity. On the basis of very vague hypotheses (e.g., if electricity produces magnetism, then magnetism will produce electricity), he set up situations that he thought might reveal the contemplated phenomena, tried to build simple theoretical constructs that would explain (or, at worst, name) the phenomena actually observed, used these to suggest new situations that would produce new phenomena, etcetera. Perhaps this is the behavior that we should be studying when we study scientific discovery.

Perhaps this is the behavior that we should be teaching when we try to preserve and enhance the curiosity of children. Perhaps this is the behavior we should be teaching our graduate students. Perhaps this is even the behavior we should be capturing when we rewrite our textbooks on experimental method.

Is Faraday Unique?

Before resting my case, I should raise again, as I did with Fleming, the question of whether Faraday is unique. There is ample evidence that he is not. Another case providing an equally full record of the course of discovery is Hans Krebs' discovery of the reaction path for the synthesis of urea in vivo, a case examined in careful detail by the science historian, C. L. Holmes (1991), and simulated with some success by two distinct computer programs (Kulkarni and Simon, 1988; Grasshoff and May, 1995). Here again, the initial experimental or "observational" inquiry was conducted on the vaguest hypotheses (that the synthesis must most likely begin with amino acids and/or ammonia as the source of the urea nitrogen). Again, the key observation was an "accident" that revealed that a particular amino acid, ornithine, was implicated in the reaction path; and this accident led to the gradual discovery that ornithine served as a catalyst, not as the initial source of the nitrogen. (The key observation was an accident in that ornithine was selected for testing for the wrong reason.)

A similar example of curious attention to phenomena is provided by the acceptance of the theory of continental drift as a consequence of new evidence discovered in the mid-Atlantic Ridge and a trench in the Pacific outside Puget Sound. I could go on with many other examples, but perhaps these are enough to create a certain amount of curiosity about my hypothesis that will motivate others to pursue it.

In the first sole-authored monograph that I published, in 1943, with the formidable title of *Fiscal Aspects of Metropolitan Consolidation*, the dedication read:

> To my mother and father,
>
> who taught me that curiosity is the beginning of all science.

References

Berlyne, D. E. (1960) *Conflict, arousal, and curiosity.* New York, NY: McGraw-Hill.

Grasshoff, G., and May, M. (1995) Methodische Analyse wissenschaftlichen Entdeckens. *Kognitionswissenschaft, 5,* 51-67.

Holmes, F. L. (1991) *Hans Krebs.* Oxford, UK: Oxford University Press.

Kulkarni, D., and Simon, H. A. (1988) The process of scientific discovery: The strategy of experimentation. *Cognitive Science , 12,* 139-175.

Magnani, G. (1996), Visual representation and scientific discovery: The historical case of the discovery of electromagnetic induction. M.S. Dissertation, Department of Philosophy, Carnegie Mellon University, Pittsburgh, PA.

Qin, Y., and Simon, H. A. (1990) Laboratory replication of scientific discovery processes. *Cognitive Science, 14,* 281-312.

Williams, L. P. (1965) *Michael Faraday.* New York, NY: Basic Books.