Philosophy of Science Association

Philosophy of Science

December, 1973

DOES SCIENTIFIC DISCOVERY HAVE A LOGIC?*

HERBERT A. SIMON†

Carnegie-Mellon University

It is often claimed that there can be no such thing as a logic of scientific discovery, but only a logic of verification. By 'logic of discovery' is usually meant a normative theory of discovery processes. The claim that such a normative theory is impossible is shown to be incorrect; and two examples are provided of domains where formal processes of varying efficacy for discovering lawfulness can be constructed and compared. The analysis shows how one can treat operationally and formally phenomena that have usually been dismissed with fuzzy labels like 'intuition' and 'creativity'.

1. Introduction. It is unusual for an author, less than one-tenth of the way through his work, to disclaim the existence of the subject matter that the title of his treatise announces. Yet that is exactly what Karl Popper does in his classic, *The Logic of Scientific Discovery*, [4], announcing in no uncertain terms on page 31 that scientific discovery has no logic. The disclaimer is so remarkable that it deserves to be quoted at length:

> I said above that the work of the scientist consists in putting forward and testing theories.
>
> The initial stage, the act of conceiving or inventing a theory, seems to me neither to call for logical analysis nor to be susceptible of it. The question how it happens that a new idea occurs to a man—whether it is a musical theme, a dramatic conflict, or a scientific theory—may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge. The latter is concerned not with *questions of fact* (Kant's *quid facti?*), but only with questions of *justification or validity* (Kant's *quid juris?*). . . .
>
> Accordingly, I shall distinguish sharply between the process of conceiving a new idea, and the methods and results of examining it logically. As to the task of the logic of knowledge—in contradistinction to the psychology of knowledge—I shall proceed on the assumption that it consists solely in investigating the methods employed in those systematic tests to which every new idea must be subjected if it is to be seriously entertained. . . .

... my view of the matter, for what it is worth, is that there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process. My view may be expressed by saying that every discovery contains 'an irrational element', or 'a creative intuition', in Bergson's sense. In a similar way, Einstein speaks of the 'search for those highly universal laws ... from which a picture of the world can be obtained by pure deduction. There is no logical path', he says, 'leading to these ... laws. They can only be reached by intuition, based upon something like an intellectual love ('*Einfühlung*') of the objects of experience'. ([4], pp. 31–32)

This mystical view towards discovery, while shared by most of the world, including most of its creative scientists and artists, has not gone without challenge. Peirce coined the term 'retroduction' as a label for the systematic processes leading to discovery; while Norwood Hanson, in his *Patterns of Discovery*, revived that term and gave us a careful account of the retroductive path that led Kepler to the elliptical orbits of the planets. It is instructive to confront Popper's view, just quoted, with Hanson's:

H-D [hypothetico-deductive] accounts all agree that physical laws explain data, but they obscure the initial connexion between data and laws; indeed, they suggest that the fundamental inference is from higher-order hypotheses to observation statements. This may be a way of setting out one's reasons for accepting an hypothesis after it is got, or for making a prediction, but it is not a way of setting out reasons for proposing or for trying an hypothesis in the first place. Yet the initial suggestion of an hypothesis is very often a reasonable affair. It is not so often affected by intuition, insight, hunches, or other imponderables as biographers or scientists suggest. Disciples of the H-D account often dismiss the dawning of an hypothesis as being of psychological interest only, or else claim it to be the province solely of genius and not of logic. They are wrong. If establishing an hypothesis through its predictions has a logic, so has the conceiving of an hypothesis. ([2], p. 71)

Hanson made his case for retroduction by examining historical examples of scientific discovery. He did not propose an explicit formal theory of the retroductive process, nor did he draw any sharp line between psychology and logic. Indeed, his analysis places great emphasis upon perceptual processes, upon the discovery of pattern—a pun upon the title of his book that, I am sure, had occurred to him. For this reason, it is easy for persons of an H-D persuasion to judge Hanson's work superficially as a contribution only to psychology and not to the logic of science. In this, they are wrong also.

It is the aim of this paper to clarify the nature of retroduction, and to explain in what sense one can speak of a "logic of discovery" or "logic of retroduction." Like Hanson, I shall proceed from examples of retroductive processes, but examples that are less impressive than his. Their modesty as instances of discovery will be compensated by their transparency in revealing underlying process. The argument of Popper and of the others who agree with his position is, after all, a general

argument. If "there is no such thing as a logical method of having new ideas," then there is no such thing as a logical method of having small new ideas.[1]

2. What is a Logic of Method? At the outset it is important to make clear what might be meant by the term 'logic' in this context. I suppose no one any longer imagines that there exists a process for deriving scientific laws deductively (much less, for validating them) either from particular facts or from any other kinds of prevalidated premises. Nor is that what Popper means by "logical analysis" in the passage quoted at the beginning of this paper. The distinction Popper draws is between psychology and logic—between a description of how people actually behave (e.g. judge, reason), and a prescription of norms of valid behavior (e.g. judging soundly, reasoning correctly and rigorously).

We commonly call a process "logical" when it satisfies norms we have established for it; and these norms derive from our concern that the process be efficacious or efficient for accomplishing the purpose for which it was established. A logic of scientific method, then, is a set of normative standards for judging the processes used to discover or test scientific theories, or the formal structure of the theories themselves (Simon [6], p. 443). The use of the term 'logic' suggests that the norms can be derived from the goals of the scientific activity. That is to say, a normative theory rests on contingent propositions like: 'If process X is to be efficacious for attaining goal Y, then it should have properties A, B, C'.

It is in this sense that we can speak of the "logic" of a chess strategy. The game of chess has a goal: to checkmate the opponent's King. The chessplayer uses a strategy to discover and evaluate moves directed toward that goal. Books on chess contain normative statements about these discovery and evaluation processes— e.g. 'In a position where the player has greater mobility than his opponent, he should examine moves that attack the position of the opponent's King directly'. The validity of this normative statement rests on a premise like: 'Where one player has the greater mobility, direct attacks on the position of the opponent's King are frequently the best paths toward checkmating the opponent'.

To generalize, suppose that we have a goal, G, a set of processes, $p \in P$, and a set of conditions, $c \in C$. The conditions can be attributed to processes, so that $c(p)$ is a function from $C \times P$ to the truth-values T and F. If, now, $\forall c(G \supset c)$, then we can regard C as a set of norms for P with respect to G. That is to say, if attainment of the goal, G, implies that the conditions, C, be satisfied, then we should employ a process, p, that satisfies C (i.e. such that $\forall c(c(p) = T)$).

If G is the goal of discovering valid scientific laws, and P is a class of discovery processes, then C provides a normative theory of scientific discovery. If G is the goal of testing the validity of proposed laws, and P is a class of test processes, then C provides a normative theory of testing laws. The premises, $G \supset c$, may themselves have either a logical or an empirical basis. In the game of tic-tac-toe, a move that puts a second cross at the intersection of two unblocked arrays, each of which has one cross already, is a winning move. A normative theory of

---

[1] For another essay taking essentially the same view as this one toward the possibility of a logic of discovery, and considering at length the philosophical objections to this view, see Buchanan, [1].

tic-tac-toe would recommend strategies that make such a move when possible, and condemn strategies that do not. The correctness of this condition can be deduced rigorously from the rules of tic-tac-toe. On the other hand, the norm of chess strategy mentioned earlier—that attacks on the King should be considered when superior mobility has been achieved—is an empirical rule based on the cumulated experiences of chess players. No formal deduction of this norm from the rules alone of chess is known.

Notice that we have formulated matters so that the use of a recommended strategy does not guarantee the achievement of the goal. We wrote $G \supset c$ and not $c \supset G$. In order to achieve the goal (e.g. to discover a scientific law), it is recommended that we use processes that satisfy the condition, c; it is not implied that using such processes will necessarily lead to discovery of a law. Hence, we have not smuggled an induction axiom into our formulation.

We have seen that the norms can have either a logical or an empirical basis. Perhaps the presence of an empirical component in norms has encouraged the notion that certain processes cannot be subject to "logical analysis" but only to description. But if the phrase 'logical analysis' is interpreted broadly, then, as we have just seen, one can undertake normative logical analysis of any goal-directed process. And if the phrase 'logical analysis' is interpreted narrowly, in order to exclude deductions from empirically based premises, then the dichotomy between logical analysis and description is a false one. For in this case, we must consider three possibilities, not two: (1) description, (2) normative analysis not dependent on empirical premises, and (3) normative analysis dependent on empirical premises. In any event, we see that we must reject Popper's assertion that the "question how it happens that a new idea occurs to a man . . . may be of great interest to empirical psychology; but it is irrelevant to the logical analysis of scientific knowledge" ([4], p. 31). The genuine normative questions about the efficacy of processes of scientific discovery cannot be dismissed in this way.

3. Law Discovery: an example. We might ask how the premature dismissal of the possibility of a normative theory of discovery occurred. I suspect, but only suspect, that it occurred because the possibility of such a theory was supposed, erroneously, to depend on the solution of the problem of induction. The discovery process runs from particular facts to general laws that are somehow induced from them; the process of testing discoveries runs from the laws to predictions of particular facts deduced from them. Hence, ordinary, garden-variety deductive logic provides the formal foundation for a normative theory of law testing (in particular, of law falsification); while a normative theory of law discovery has been thought to require a quite different inductive logic as its foundation.

If this were the case, the normative theory of discovery would then share the difficulties of the inductive logic on which it rested. And if there is anything we can reach agreement upon with respect to inductive logic, it is that any particular proposed system of inductive logic (unless it is our own system) is defective. In the light of the innumerable, unsuccessful attempts to untie the Gordian knot of induction, we shall be well advised to cut it—to construct a normative theory

for evaluating discovery processes that does not demand a deductive justification for the products of induction. I should now like to explain exactly how this can be done, beginning with a concrete example.

Consider the following sequence of letters:

ABMCDMEFMGHMIJMKLM
                 MNMOPMQRMSTMUVMWXMYZMABMC...

If we examine the sequence, we soon detect that it is patterned; that is to say, it is redundant, and can consequently be described more parsimoniously by defining the pattern than by exhibiting the sequence itself. In particular, it can be described as a sequence of triads. The first two letters of each triad progress through the alphabet, the third letter is an 'M'. Given an appropriate notation for patterns, we can represent the general triad of the sequence by some such pattern as:

(1) $n(\alpha)n(\alpha)s(\beta)$; $\alpha = Z, \beta = M,$

where '$n(\alpha)$' means replacing a symbol by the symbol next to it on the alphabet, α; '$s(\beta)$' means repeating the same symbol as β; while the expressions '$\alpha = Z$' and '$\beta = M$' set the initial values on the alphabets, at Z and M, respectively. (The alphabets are assumed "circular," so that A follows Z.)

Are we really certain that (1) is the law of the sequence? Yes, for we can apply it to the sequence and see that it matches the letters that actually appear there. But what about the continuation of the sequence—the omitted letters indicated by dots? The expression (1) predicts that the extrapolation will begin with DMEFM, and so on; and there is no guarantee that it will in fact continue in this particular way. To provide a logical justification for the extrapolation, we need to appeal to some principle of the uniformity of nature, or some other premise of induction. We appear not to have avoided at all the troublesome problem of induction.

However, the difficulty with which we are confronted here is illusory. It does not arise at all in connection with discovering a pattern—recoding parsimoniously the portion of the sequence that was presented explicitly. It arises only if we wish to predict and test whether this same pattern will continue to govern the sequence when it is extrapolated. Law discovery means only finding pattern in the data that have been observed; whether the pattern will continue to hold for new data that are observed subsequently will be decided in the course of testing the law, not discovering it.

We will banish the problem of induction from our discussion of law discovery once and for all by defining:

> A *law-discovery process* is a process for recoding, in parsimonious fashion, sets of empirical data.

> A *normative theory of scientific discovery* is a set of criteria for evaluating law-discovery processes.

4. Efficient and Inefficient Law-Discovery Processes. We have seen that the discovery process has nothing to do with the usual problem of justifying inductions, because the adequacy of the pattern to describe the sequence actually presented (as distinguished from its extrapolation) can be settled in purely finitary terms. There is no doubt at all that the particular pattern, (1), describes accurately the sequence of latters that was exhibited above. Of course no claim is made that this description is unique, or even that it is the most parsimonious possible.

But we have not explained how we discovered the pattern. To explain the discovery, do we have to postulate an "irrational element" or a "creative intuition?" Or, on the contrary, can we give a descriptive account of how people make such discoveries, and a normative account of relatively efficient ways of making them? If we can do the latter, then we will have constructed a logic of discovery—at least for simple situations like the one of the example.

Let us suppose that we have a pattern discovery process possessing means for representing the relations of *same* and *next*, so that it can compose formulas like 'n(α)n(α)s(β)'. Then it can discover the pattern in a sequence by generating such formulas in some order—starting, say, with the simplest and generating progressively longer ones—testing each until it finds one that fits the actual sequence. In another context, Allen Newell and I have called such a generative process a "British Museum Algorithm" (BMA) to honor the monkeys who were alleged to have used it to reproduce the volumes in the British Museum. For the simple case before us, the algorithm might actually succeed in a reasonable time. For example, it might generate a sequence of patterns like: s(α), n(α), s(α)s(β), n(α)n(β), s(α)n(α), s(α)n(β), and so on. The pattern n(α)n(α)s(β) would then occur among the first fifty or so generated.

It is easy, however, to construct a pattern discovery process that is far more efficient than the British Museum Algorithm. By drawing upon information that is explicit in the sequence presented, the algorithm can construct a suitable pattern directly, with little or no trial and error. It begins by scanning the sequence, and noting occurrences of the relations of *same* and *next* between symbols that are not too far separated. In the case before us, it would find that every third symbol in the sequence is an 'M'. From this fact it would conjecture that the sequence has a period of three symbols. It would then notice that the second symbol in each triad terminated by an 'M' is the successor in the alphabet of the first symbol, and that the first symbol is the successor in the alphabet of the second symbol of the previous triad. This information is sufficient to define the pattern 'n(α)n(α)-s(β)'. Finally, the initial conditions, $\alpha = Z$ and $\beta = M$, can be determined directly by matching with the appropriate symbols in the sequence.

If we are interested in reducing the amount of trial-and-error search required to find the formula that describes the sequence, then we will prefer this second algorithm (which we will call a Heuristic Search Algorithm—HSA) to the British Museum Algorithm. The BMA tries alternatives systematically until it finds one that works; the HSA extracts information from the sequence in order to generate directly an alternative that will work. The difference between the two algorithms is exactly parallel to the difference between solving an algebraic equation by try-

ing possible solutions in some systematic order, and solving it by eliminating constants from the left side, next eliminating variables from the right side, and then dividing through by the coefficient of the variable.

The normative theory of discovery processes can be viewed as a branch of the theory of computational complexity. Given a class of computational problems, we wish to discover algorithms that are most efficient on average; or, if we cannot find the most efficient, at least to discover some that make good use of the information available in the problem situation. In the case of symbolic sequences like the example, this information consists in the periodic recurrence of certain symbols, or the periodic recurrence of *successor* relations between neighboring symbols drawn from the same ordered set.

It is not hard to write computer programs that will detect relations of these kinds, and conduct heuristic searches for pattern in an orderly and efficient way. Several such programs have, in fact, been written and tested with success on such materials as the Thurstone Letter Series Completion Test. The programs are vastly more efficient than would be a BMA-like program for the same task. (See [5] and [8].)

If there is a pattern in the sequence, of one of the kinds that the program is capable of detecting, it will be detected, and usually with a relatively small amount of search. Of course, failure to detect a pattern does not guarantee that one does not exist—a pattern, say employing relations that are not within the program's competence. Given a particular class of potential patterns, however, some programs will regularly find patterns in this class much more rapidly and efficiently than will other programs. The normative theory of discovery is concerned with characterizing these efficient programs.

Moreover, the particular class of patterns we have been discussing in the example is not as specialized as might appear at first blush. The relation of *successor* in an ordered set (with the circularity convention mentioned above) defines a cyclic group on the set. This is just the capability we need to detect groups of symmetries; and most of the patterns we succeed in extracting from nature appear to be representable in terms of such groups. Consider, for example, the Periodic Law of Mendeleev. Discovering this law involves arranging the elements by the successor relation with respect to atomic weight, and then noting the periodic repetition of certain chemical properties (e.g. valence) in this list. Other examples of pattern discovery in science are discussed in Simon, [6].

5. Concept Attainment: a second example.

The notion of informational efficiency of law-discovery processes can also be illustrated with an example drawn from the domain of concept attainment. In the standard psychological concept-attainment task, a sequence of stimuli is presented, some of which are positive instances of a concept and others negative instances. The goal is to discover the concept—that is, a criterion that will allow the positive instances to be distinguished from the negative instances.

Suppose that the stimuli vary in color, size, and shape. Suppose further that only simple concepts—a particular value on a particular dimension—are

admitted. Thus, the concept in a given case might be "circles." A relatively ineffi-
cient process would examine the instances in sequence after selecting an initial
hypothesis. If the hypothesis led to a wrong classification in any single case, it
would be rejected and a new hypothesis tried. A much more efficient process would
examine the instances in sequence and eliminate from further consideration the
whole set of concepts inconsistent with each successive instance. In general, the
number of instances the latter process would have to examine before selecting a
concept that was consistent with the entire set would vary with the logarithm of
the number of possible hypotheses. For the former process, it would vary linearly
with the number of possible hypotheses. Clearly the expected number of trials
needed to find the correct hypothesis is smaller with the more sophisticated pro-
cess. Again, we can construct a normative theory that evaluates the relative power
and sophistication of different discovery processes in this domain.

6. Reply to an Objection. A perceptive reader of an earlier draft of this paper ob-
served that in arguing for the possibility of a logic of discovery I have used exam-
ples that prejudice the case in my favor. The examples envisage situations in which
the range of alternatives can be delimited in advance (the sorts of patterns at issue,
the range of possible hypotheses, or the like). These examples do not prove at all
that outside the range of limited and "artificial" problems one can, in science,
define the processes of genuine innovation that redefine the range of alternatives
that can be envisaged.

This objection rests on the commonly drawn distinction between well-structured
problems, which are amenable to orderly analysis, and ill-structured problems,
which are the exclusive domain of creativity. It is reminiscent also of the
Kuhnian distinction between normal and revolutionary science. The objection
depends for its force upon these distinctions being qualitative, and not merely
matters of degree.

The simplest reply to the objection is to admit it, but to observe that it does not
invalidate the thesis of this paper. As I observed in the introduction, those who
have argued against the possibility of a logic of discovery have drawn no such dis-
tinction between "normal" and "revolutionary" discoveries. If a logic of dis-
covery can be constructed for even one small realm—and my examples show that
it can for several—then their arguments must be faulty. If they wish now to argue
against the possibility only of a logic of revolutionary discovery, they will have
to find new arguments that apply to this realm alone.

But I am not persuaded that even this much ground must be yielded to the ob-
jectors. The notion of "delimiting in advance" the range of hypotheses to be con-
sidered is not at all a clear concept. We know that all that is required to generate
an infinite number of elements from a finite set of primitives is some kind of re-
cursively applicable rule of generation. Thus, *modus ponens* applied to a small
finite set of axioms may suffice to generate an infinity of theorems; and an ex-
tremely simple phrase-structure or transformational grammar can spawn an
infinite number of grammatical sentences. Have we any genuine reason to deny

that "revolutionary" hypotheses are the products of this kind of generation of much from little?

It was pointed out earlier that Mendeleev's Periodic Table does not involve a notion of pattern more complex than that required to handle patterned letter sequences. To be sure, Mendeleev had to conjecture that atomic weight was the attribute according to which the elements should be ordered. But it should be observed that he made his discovery just a few years after the notion of atomic weight had been clarified, its central importance had been accepted, and the atomic weights of most of the known elements had been determined. The space of possible (or plausible) patterns in which Mendeleev was searching was perhaps of rather modest size. And, indeed, at least a half dozen of Mendeleev's contemporaries had noticed the pattern independently of him, although they had not exploited it as systematically or vigorously as he did.

Before we accept the hypothesis, therefore, that "revolutionary science" is not subject to laws of effective search we would do well to await more microscopic studies than have generally been made to date of the histories of revolutionary discoveries. The case of Mendeleev may prove to be not at all exceptional. At present, one can argue this only as a possibility. But as long as it is a possibility, we must receive with skepticism arguments that would seek to prove the "impossibility" of a logic of scientific discovery—even revolutionary discovery.

7. Conclusion. The simple examples cited in this paper show that one can construct a normative theory—a "logic," if you will—of discovery processes. The greater efficacy of one process compared with another in discovering laws need not be attributed to chance, irrationality, or creative intuition. Rather, it is a matter of which process is the more capable of detecting the pattern information contained in the data, and using this information to recode the data in more parsimonious form.

At present, in the field of artificial intelligence, there exist the beginnings of a normative theory of problem solving—a theory of the design of effective problem-solving algorithms. (See, for example, Nilsson [3], and Simon [7].) The normative theory of pattern-discovery is less far advanced, partly because fewer successful examples have been constructed of powerful pattern discovery programs. Our analysis suggests, however, that this difference in stage of development is rather to be viewed as an historical accident than as an indication of some fundamental qualitative difference in the requirements for normative theory in these two domains.

The fact that a process can extract pattern from finite data sets says nothing about the predictive power of the patterns so extracted for new observations. As we move from pattern detection to prediction, we move from the theory of discovery processes to the theory of processes for testing laws. To explain why the patterns we extract from observations frequently lead to correct predictions (when they do) requires us to face again the problem of induction, and perhaps to make some hypothesis about the uniformity of nature. But that hypothesis is neither required for, nor relevant to, the theory of discovery processes. The latter theory

does not assert that data are patterned. Rather, it shows how pattern is to be detected if it is there. This is not a descriptive or psychological matter, it is normative and logical. By separating the question of pattern detection from the question of prediction, we can construct a true normative theory of discovery—a logic of discovery.

REFERENCES

[1] Buchanan, B. G. *Logics of Scientific Discovery*, A. I. Memo 47, Computer Science Department, Stanford University, 1966.
[2] Hanson, N. R. *Patterns of Discovery*. Cambridge: Cambridge University Press, 1958.
[3] Nilsson, N. J. *Problem-Solving Methods in Artificial Intelligence*. New York: McGraw-Hill, 1971.
[4] Popper, K. R. *The Logic of Scientific Discovery*. London: Hutchinson and Company, 1959.
[5] Simon, H. A. and Kotovsky, K. "Human Acquisition of Concepts for Serial Patterns." *Psychological Review* 70 (1963): 534–546.
[6] Simon, H. A. "On Judging the Plausibility of Theories." In *Logic, Methodology and Philosophy of Science*. Vol. III. Edited by B. Van Rootselar and J. F. Staal. Amsterdam: North-Holland Publishing Company, 1968.
[7] Simon, H. A. "The Theory of Problem Solving." *Information Processing* 71 (1972).
[8] Simon, H. A. and Lea, G. "Problem Solving and Rule Induction: a Unified View." (To be published in *Proceedings of the Ninth Carnegie-Mellon Symposium on Cognition*. Edited by L. W. Gregg.)