POLYA'S "PATTERNS OF PLAUSIBLE INFERENCE" AND THE QUEST FOR MODULARITY

Judea Pearl

December 1987
CSD-870064

TECHNICAL REPORT
CSD 8600##
R-71
November 1986

POLYA'S "PATTERNS OF PLAUSIBLE INFERENCE" AND THE QUEST FOR MODULARITY *

Judea Pearl

Cognitive Systems Laboratory
Computer Science Department
University of California
Los Angeles, CA 90024

* This work was supported in part by the National Science Foundation Grants #DCR 83-13875 & 86-44931.

POLYA'S "PATTERNS OF PLAUSIBLE INFERENCE" AND THE QUEST FOR MODULARITY

1. INTRODUCTION

The mathematician George Polya (b.1887) was one of the first to attempt a formal characterization of qualitative human reasoning. In his book *Patterns of Plausible Inference*, [Polya, 1954], he argues that the process of discovery, in even as formal a field as mathematics, is guided by non-deductive inference mechanisms, entailing a large amount of guesswork. He termed the principles governing this guesswork *patterns of plausible inference*.

Among the conspicuous patterns listed by Polya, we find the following:

1. ***inductive patterns*** - "The verification of a consequence renders a conjecture more credible."

   For example, the conjecture "It rained last night" becomes more credible by verifying that "The ground is wet."

2. ***successive verification of several consequences*** - "The verification of a new consequence counts more or less if the new consequence differs more or less from the former, verified consequences."

   For example, if in trying to substantiate the conjecture "All ravens are black," we already observed n Australian ravens, and all turned out black, our subsequent confidence in the conjecture would be increased substantially if the $(n + 1)$th were a black Brazilian raven rather than another black Australian raven.

3. ***verification of improbable consequences*** - "The verification of a consequence counts more or less according as the consequence is more or less improbable in itself."

   For example, the conjecture "It rained last night" obtains more support from "The roof is leaking" than from the more common observation "The grass is wet."

4. ***inference from analogy*** - "A conjecture becomes more credible when an analogous conjecture turns out to be true."

   For example, the conjecture "Of all objects occupying the same volume, the sphere has the smallest surface" becomes more credible when we prove that "Of all curves entailing the same area, the circle has the shortest perimeter."

Polya further identified three main sub-patterns of inductive reasoning:

1.1 ***examining a consequence*** - same as (1) above.

1.2 ***examining a possible ground*** - "Our confidence in a conjecture can only diminish when a possible ground for the conjecture is exploded."

1.3 ***examining a conflicting conjecture*** - "Our confidence in a conjecture can only increase when an incompatible rival conjecture is exploded."

These patterns can be further refined depending on whether propositions are verified categorically, or just become more credible (which Polya calls "shaded" verification). The sum total of these patterns and subpatterns are summarized in Table 1.

POLYA'S PATTERNS OF PLAUSIBLE INFERENCE

	(1) Demonstrative	(2) Shaded Demonstrative	(3) Shaded Inductive	(4) Inductive
1. examining a consequence	$A \to B$ B false ——— A false	$A \to B$ B less cr. ——— A less cr.	$A \to B$ B more cr. ——— A s. more cr.	$A \to B$ B true ——— A more cr.
2. examining a possible ground	$A \leftarrow B$ B true ——— A true	$A \leftarrow B$ B more cr. ——— A more cr.	$A \leftarrow B$ B less cr. ——— A s. less cr.	$A \leftarrow B$ B false ——— A less cr.
3. examining a conflicting conjecture	$A \mid B$ B true ——— A false	$A \mid B$ B more cr. ——— A less cr.	$A \mid B$ B less cr. ——— A s. more cr.	$A \mid B$ B false ——— A more cr.

Table 1

In this table, $A \to B$ stands for "A implies B," *cr.* is short for "credible," *s.* is short for "somewhat" and $A \mid B$ stands for "A is incompatible with B," i.e., A and B could not both be true at the same time.

The patterns in the second row, "examining a possible ground" are logically equivalent to those in the first row, "examining a consequence." For example, entry 2 (2) follows from 2 (1) because $A \to B$ is logically equivalent to $(\neg B) \to (\neg A)$ and, calling $B' = \neg B$, $A' = \neg A$ we get $B' \to A'$ which yields entry 2 (2) (assuming, of course, B more cr. $\Longleftrightarrow$ $\neg B$ less cr.). However, restating row 2 (separately) has its own merit since people do not readily perceive logical identities as psychological necessities; so, redundant inference rules are useful for dealing with logically equivalent but syntactically different situations.

2. DEMONSTRATIVE vs. PLAUSIBLE INFERENCES

When stated individually, each pattern in Table 1 appears plausible and is supported by many examples. However, after extracting a sizable number of such ''conspicuous'' primitive patterns, Polya stops short of proposing them as syllogistic axioms (or inference rules) for a new logic, capable of manipulating concepts such as ''credible,'' ''more credible,'' ''somewhat credible,'' etc. Rather, Polya shelves this promising prospect and retreats into the safety blanket of probability calculus -- from which all the qualitative patterns of plausible inference should follow naturally and automatically, thus circumventing the need to express them in symbolic terms. It is interesting to trace the path traversed by Polya, who was one of the most eminent mathematicians of his time, in his quest for formal principles of plausible reasoning.

The reason for Polya's sharp retreat is explained in Chapter XV of his book and is based on the the realization that primitive patterns of plausible reasoning, as ''reasonable'' as they appear and, despite their syntactic similarity to logical syllogisms, are of basically different character than those syllogisms. Polya identified four basic differences between the two modes of reasoning and, since his analysis is still relevant to modern-day thinking (especially to on-going debates regarding the role of Logic in Artificial Intelligence), it is timely and instructive to examine the differences afresh.

According to Polya, demonstrative reasoning is *impersonal, universal, self-sufficient* and *definite*, while plausible reasoning shares only parts of these qualities.

1. *impersonal* - the validity of logical syllogisms does not depend on the personality, mood or taste of the reasoner.

   In plausible reasoning, personal differences *do* affect "how much" credibility changes in each inferential step, though it does not affect the direction of change.

2. *universal* - the validity of syllogistic conclusions does not vary from one subject matter to another.

   By comparison, in plausible inferences, the magnitude by which credibility changes *does* depend on the domain of discourse. The direction of change, though, remains universal.

3. *self-sufficient* - new information, as long as it does not conflict with the premises, will never change the conclusions reached by demonstrative inferences. "Nothing is needed beyond the premises to validate the conclusion and nothing can invalidate it if the premises remain solid". These qualities have been recognized as dis-

tinct merits of logical formalisms since Aristotle, and have been called "monotonicity" and "modularity" in modern terms.

By contrast, credibility levels established by plausible inferences are not "durable", as they may change with new information. In each inferential step, though, the direction of change depends only on the premises considered at that step. (Here Polya erred, as we shall next demonstrate; the direction, too, depends on things outside the premises.)

4. *definitive* - In logic, once a conclusion is drawn, we can forget which premise led to it; derived truths are no less "true" than observed truths.

The premises cannot be ignored in plausible reasoning because they define the reference conditions relative to which propositions become more or less credible. The conclusions are "provisional" in the sense that the credibility measures which they establish may undergo additional changes in the future. (Polya did not elaborate on this feature, but it is important to note that both the magnitude and direction of future changes may depend on the original premises; so, one cannot truly dispose of the premises and retain only the conclusions.)

3. WHY POLYA'S PATTERNS CANNOT BE USED AS PRODUCTION RULES

The most striking aspect of Polya's analysis is the emphasis on "change" or "update," involving transitions between states of knowledge where the credibility of a proposition undergoes adjustment. The very reliance on "change" already constitutes a marked departure from traditional logics. Logical syllogisms normally respond to static conditions not do dynamic comparisons. In other words, to assert the validity of the conclusion of a certain inference rule, we need only ascertain the current truth value of its antecedent part; *when* this truth has been established does not matter.

But the difficulty with the patterns assembled by Polya does not stem from the dynamics associated with the terms "more credible" or "less credible," neither from the *imprecise* and *provisional* character of the inferences drawn. Quite the contrary, any logic that can manage provisional beliefs without specifying their exact numerical strengths would be exactly what modern non-monotonic logics aspire to achieve. The difficulty with Polya's patterns is that they are plainly *wrong*, namely, if used as production rules by an inference engine, they will invariably produce counter-intuitive conclusions, unanticipated by the rule provider.

To demonstrate this claim, we first need to give clear definitions to the notions of "credible" and "more credible," reflecting their use in ordinary, everyday discourse. These terms clearly refer to some feature of one's *state of knowledge* in which measures of credibility are assigned to propositions. Thus, we imagine that each state of knowledge S is characterized by a triple $(AX \;\; LC, PC)$, where AX is a set of axioms representing undisputable facts, LC a set of logical conclusions drawn from those axioms by logical inference rules and PC are conclusions drawn by plausible inference rules. Every well formed sentence p in S will be assigned a real valued truth function $cr(p)$ $(0 \leq cr(p) \leq 1)$, which measures its degree of credibility. If p is an axiom or a logical derivative of some axioms ($p \in LC \cup AX$), then its credibility will be 1. If, however, p was inferred by some plausible inference rules, its credibility may be lower than 1. Accordingly, the intended meaning of plausible rules such as Polya's patterns could be interpreted in terms of the constraints which they imposed on the assignment of cr to the sentences in S. In other words, the assignment of $cr(\bullet)$ to the various sentences is not entirely arbitrary but must comply with some internal consistency rules, such as the inductive pattern:

$$(A \rightarrow B) \;\&\; (B = true) \implies A \textit{ more credible} \tag{1}$$

which constraints the changes allowed in $cr(\bullet)$ as new information obtains.

Specifically, the meaning of "more credible" in Eq.(1) should be viewed as a relation between two typical states of knowledge, $S_1 = (AX_1, LC_1, PL_1)$ and $S_2 = (AX_2, LC_2, PL_2)$, both sharing a set of basic axioms (e.g., $A \rightarrow B$) but S_2 contains additional facts (e.g., B) so, $AX_1 \subseteq AX_2$ and $LC_1 \subseteq LC_2$. In addition, we imagine that A and B are uncertain in S_1 while B is true in S_2, so, $cr_1(A) < 1$, $cr_1(B) < 1$ and $cr_2(B) = 1$. Under these conditions (1) appears to claim that $cr_2(A)$ is not arbitrary but must obey

$$cr_2(A) > cr_1(A) \tag{2}$$

Formally summarizing this interpretation, (1) asserts

$$[(A \rightarrow B) \in AX_1 \subseteq AX_2] \;\&\; [cr_1(A) < 1, cr_1(B) < 1, B \in LC_2] \implies cr_2(A) > cr_1(A) \tag{3}$$

But this interpretation can easily be shown to be too permissive, leading to paradoxical conclusions unintended by (1). For example, (3) is violated if the facts added to S_2 entail both (B) and (Not-A) because $cr_2(A) = 0$, which clearly implies $cr_2(A) < cr_1(A)$. For example, if A stands for "it rained last night" and B for "my grass is wet" then observing B together with C – "my neighbor's grass is dry" surely precludes A, in clear violation of the conclusion of (1).

A natural attempt to prevent such paradoxical conclusions would be to modify the interpretation (3) by restricting the set of facts by which S_2 may differ from S_1. For example, we might claim that (3) holds only between two states of knowledge S_1 & S_2 such that the information added to S_2, $AX_2 - AX_1$, is B alone, nothing else. Unfortunately, this rescue of (3) will also preclude the application of (1) in all cases where the truth of B is accompanied by some other irrelevant facts, e.g., D – "There is a civil war in Mongolia" and, more seriously, it will preclude cases where the truth of B is infered from more direct observations e.g., "E – My grass looks shiny and feels cold." This would render (1) almost useless because truths are normally established by indirect observations, e.g., from the fact "fire $\rightarrow$ smoke" we would be unable to infer that the likelihood of fire (A) increases when a smoke-detector takes off. To apply (1) to this situation would require that we literally verify the truth of "smoke" prior to noticing the smoke-detector, in clear violation of both intuition and entry 1-(2) of Table 1.

A more lenient restriction on (3) would be to permit the added information in S_2 to include other facts beside B, but limit it to such facts only which do not *impinge directly* on A. This restriction, too, will fail the test of intuition because some facts (called "undercutting" defeaters [Pollock, 1983]) can defeat (1) without impinging directly on A; they undermine the very connection between B and A by offering an alternative explanation to B, such as F = "My sprinkler was on last night."

As another example, imagine that $A \rightarrow B$ stands, again, for the rule "fire $\rightarrow$ smoke" and that we now observe a new fact C = "my muffler is bad." This observation, together with the common knowledge: $(C \rightarrow B)$, surely implies "smoke" (B) without impinging directly on "fire" (A). Yet, contrary to (1), no person in his right mind will get panic of fire by merely observing or learning about a bad muffler.

This lenient interpretation of (1) would also invite contradictory examples where A actually becomes *less* credible after observing B. Suppose that in state S_1 we had some weak indication of "smoke" (e.g., by virtue of noticing a bad smell), naturally evoking an increase of belief in "fire." Once we ascertain the finding "bad muffler" in state S_2 (together with its compelling prediction "smoke") it is only natural that $cr\ (fire)$ should go down, not up as claimed by (1). Thus, the induction pattern (1) fails to predict not merely the *strength* of confidence change but also its *direction*, relative to which Polya's principles were thought to be impersonal and universal.

The fact that a mathematician of Polya's caliber has not detected these basic flaws is not so much a blemish of his own insight as it is an indication of the type of assumptions we all make when asked to judge the plausibility of an argument. The inductive pattern (1) appears innocently plausible to almost every person, because we tacitly assume that the truth of B is the *only* change that took place in the world. In other words, unless stated explicitly, all truth values, especially of events that preceed B, are presumed to persist unaltered. Since changes in the truth of other propositions (e.g., muffler condition) are not mentioned in (1), we habitually presume that in the transition from S_1 to S_2 the truth of B (e.g., smoke) was established by external means, e.g., direct observations or reliable testimonies, not as a consequence of other, unmentioned changes.

This presumption is related to the infamous *frame problem* in Artificial Intelligence [McCarthy & Hayes, 1969] and is also called "chronological persistence" [Kautz, 1986] or "chronological ignorance" [Shoham, 1986]. However, translating this assumption into a more refined restriction on the applicability of rule 1 (or (3)) is not an easy task; it requires that we examine the entire set of new propositions in S_2 and determine their relationships to A and B by *extra-logical* criteria such as chronological precedence and/or causal ordering.

4. MUST PLAUSIBLE REASONING FORFEIT MODULARITY?

It is perhaps in anticipation of such difficulties that Polya stopped short of treating his plausible patterns as rules of a calculus. One of the features that has made Logic so appealing to both philosophers and computer scientists is that logical syllogisms, by their very nature, are *modular*, i.e., verifying the truth of the premises anywhere in the database constitutes a sufficient license to add the conclusion to the database, regardless of where, when and how these premises were established and regardless of what other truths the database contains. This useful feature of classical logic must be abandoned when we move to plausible inferences. Before we can apply even a simple rule such as (1) we must check the entire database (of S_2) to verify that either, it does not contain a set of propositions $\{C_1, C_2, \cdots\}$ from which the truth of B can be derived or, if such a set exists, that it does not constitute an (alternative) *explanation* of B (e.g., bad muffler) but, merely an *evidence* confirming its truth (e.g., the smoke detector). The distinction between *explanation* and *evidence*, as does the whole notion of causality, has never been a favorite study among Logicians, as it involves considering not merely the truth value of propositions and sentences but also how they are indexed and organized in the intricate fabric of human knowledge.

Polya's patterns do appear plausible to us because, our culture, language and experience have developed effective indexing schemes for determining just what portions of our database need be examined before safely applying these inference rules. These indexing schemes revolve around "causality," and involve making local tests to determine if concepts such as "bad muffler" and "smoke" stand in causal or evidential relationship to one another. Having been endowed with the machinery of causal indexing, the wisdom to recognize it and the freedom and courage to use it, saves us from the need to examine the entire database, and protects us from improper usage of Polya's patterns. These same patterns, if treated as logical rules with the bluntness and permissiveness of traditional syllogisms (i.e., "truth is truth is truth") will invariably lead to paradoxical conclusions [Pearl, 1986a].

It was perhaps this realization that led Polya to conclude his Chapter XV with a pessimistic note:

> "From the outset it was clear that the two kinds of reasoning have different tasks. From the outset they appeared very different: demonstrative reasoning as definite, final, "machine-like"; and plausible reasoning as vague, provisional, specifically "human." Now we may see the difference a little more distinctly. In opposition to demonstrative inference, plausible inference leaves indeterminate a highly relevant point: the "strength" or the "weight" of the conclusion. This weight may depend not only on clarified grounds such as those expressed in the premises, but also on unclarified unexpressed grounds somewhere in the background of the person who draws the conclusion. A person has a background, a machine has not. Indeed, you can build a

> machine to draw demonstrative conclusions for you, but I think you can never build a machine that will draw plausible inferences.''

Ironically, we do not share Polya's pessimism, even after showing that the gap between demonstrative and plausible inferences is, in fact, wider than that identified by Polya, i.e., not only the ''strength'' of the conclusions but their ''direction,'' too, depends on ''unclarified unexpressed grounds somewhere in the background...''. Polya's pessimistic prophecies would apply if one insists on pursuing the logical tradition of pretending that inference rules are activated merely by the truth of their premises. An alternative would be to adopt a more flexible attitude and make reasoning steps respond not merely to the truth of propositions but also to the various shades of their beliefs and to the type of justifications that support them. This is what belief networks aspire to achieve[Pearl, 1986b; Heckerman & Horvitz, 1986].

REFERENCES

Polya, G., *Patterns of Plausible Inference*, Princeton University Press, 1954.

McCarthy, J. & Hayes, P., "Some Philosophical Problems fromthe Standpoint of Artificial Intelligence," *Machine Intelligence, Vol. 4, 1964.*

Pearl, J., "Toward Causal Reasoning in Default Logics" (note), UCLA Cognitive Systems Laboratory *Technical Report (R-69)*, September 1986a.

Pearl, J., "Fusion, Propagation and Structuring in Belief Networks, *Artificial Intelligence*, Vol.29, No. 3, pp. 241-288, September 1986b.

Pollock, J., "A Theory of Direct Inference," *Theory and Decision*, Vol. 16, 1983.

Kautz, "The Logic of Persistence," *Proceedings, AAAI-86*, Philadelphia, Pennsylvania, pp. 401-405.

Shoham, Y., "Chronological Ignorance," Proceedings, AAAI-86, Philadelphia, Pennsylvania, pp. 389-392.

Heckerman, D. E. & Horvitz, E. J., "The Myth of Modularity in Rule-Based Systems," *Proceedings, AAAI Workshop on Uncertainty in AI*, Philadelphia, Pennsylvania, August 8-10, 1986.