To appear in Proceedings, AAAI 1989 Spring Symposium Series, March 1989, Stanford University, Stanford, CA.

TECHNICAL REPORT
R-124
February 1989

LOCALITY-BOUNDED RATIONALITY*

Judea Pearl

Computer Science Department
University of California, Los Angeles

Abstract

We examine the thesis that the major forces shaping bounded rationality stem from the requirement of local computation, namely, from the need to consider only a few data items at any inference step and the need to avoid both the search for these items as well as the decision where to store intermediate results. We explore how this requirement constrains the kind of representations we can handle and the kind of inferences we can make. Additionally, we propose a paradigm whereby the requirement of locality, by virtue of inducing a modular perception of reality, can be used to extract approximate inference strategies that are both computationally attractive and rationally defensible.

Rationality and Its Origin

We take the position that rationality stands for a set of principles that a culture accepts as means for justifying actions and beliefs, these principles being plausible generalization of concrete familiar experiences. A generalization is plausible if it is not readily violated by some concrete familiar experience.

EXAMPLE 1: Modus Ponen in logic.

a) Familiar experience: I believe that whenever it rains the ground is wet. It rains today, and I conclude that the ground is wet today.

b) Generalization: for any two propositions A and B

$$\frac{A \Rightarrow B \qquad A \text{ is True}}{B \text{ is True}}$$

c) Test for plausibility: I am unable to recall any experience where the first two sentences are strongly believed and the third is doubted.

The process of scanning our experiences to test the plausibililty of a generalization is often fallible, because it is laden with hidden assumptions tending to limit the scope of experiences within which we search for violation.

EXAMPLE 2: Polya's pattern of induction (Polya 1954).

a) Familiar experience: I believe that whenever it rains the ground is wet, I notice that the ground is wet, and I feel that the conjecture "It rained last night" becomes more credible.

b) Generalization: For any two propositions A and B:

$$\frac{A \Rightarrow B \qquad B \text{ becomes True}}{A \text{ becomes more credible}}$$

c) Test for plausibility: It is hard to generate counterexamples; even Polya believed the universality of this pattern as long as it only concerns the direction, not the degree of the change in credibility. The hidden assumption made in this case is that B is the only new fact learned (similar to the close- world assumption). This assumption (or convention) causes us to miss the following type of violating experience: Initially, we have some inconclusive evidence that the ground is wet, then the truth of "The ground is wet" is firmly established (by prediction) from a new fact "The sprinkler is on". In this case, B becomes true and A becomes less credible, in violation of the inductive pattern.

*This work was supported in part by Naval Research Laboratory Grant #N00014-87-K-2029.

EXAMPLE 3: Transitivity of Preferences

a) Familiar experience: Yesterday I preferred coffee to tea, tea to milk, and coffee to milk.

b) Plausible generalization: Whenever I am offered three *items* A, B and C, if I prefer A to B and B to C, then I would also prefer A to C.

b') Further generalization: Whenever I am facing a choice between three *situations* A, B and C, if I prefer A to B and B to C, I would also prefer A to C.

c) Test for plausibility: It is generally hard to imagine experiences that violate transitivity, but such experiences nevertheless can be reconstructed. Examples are choices among lotteries that were contrived by experimental psychologists to specifically prove this point.

Limits to Rationality

Any set of generalizations that pass the plausibility criterion can be assembled as axioms and proclaimed a normative criterion for rational behavior. Normally, the select set also enjoys the features of consistency, compactness (non-redundancy) and power, in the sense that it captures a wide spectrum of reasoning patterns. Examples are the axioms of logic, of probability theory, of utility theory of relevance theory etc.

Once we accept such a normative criterion, it can further be used to generate, not merely justify, choice behavior. In their new role as an inference mechanism, the normative axioms lead to both epistemological and computational difficulties.

Epistemologically, the axioms are now in a position to expose the empirical exceptions that were ignored while the plausibility of the individual axioms was tested and accepted. For example, accepting Polya's inductive pattern as a syllogistic rule of inference would quickly reveal counter intuitive conclusions such as a systematic increase in the credibility of "Rain" as soon as one discovers the truth of "The sprinkler is on".

Computationally, two difficulties arise. First, the rationality axioms in themselves are normally too weak to generate intersting inferences, and the information required to unleash such inferences is often too voluminous. For example, in probabilistic reasoning, too many combinations of events need be considered and assessed before a complete probabilistic model is specified and reasoning can commence. In symbolic nonmonotonic reasoning, too many exceptions and exceptions to exceptions need to be enumerated before commonsensical conclusions can be derived. Had the axioms been used simply as guardians against gross violations of some normative principles, we could just fill the gaps with any arbitrary set of assumptions (or parameters, or exceptions) and still be protected from gross violations. However, as generators of rational behavior, the added assumptions must now reflect real-life experience, and assembling such a body of knowledge, in a format acceptable to the inference mechanism, requires an enormous labor and storage.

Second, even if we obtain the information required by the rationality axioms, the process of drawing rational conclusions is, in general, intractable. For example, even if we obtain all the probabilities necessary for constructing a probabilistic model of some phenomenon, the task of computing $P(x|y)$ is in most practical cases NP-hard. Simple decisions in the propositional logic are, likewise, NP-complete.

In addition to the usual limitations on resources such as time and memory, we now wish to focus on *locality* as another constraint that makes rational behavior limited. While it might be possible to relate (or reduce) locality to more basic computational restrictions, it nevertheless deserves consideration as an independent fundamental force that shapes human rationality.

Locality as an Architectural Constraint

All realistic models of human reasoning invoke the notion of locality in one form or another. For example, spreading activation in conceptual memories is grounded in the notion that activity spreads locally, among conceptually neighboring entities, but does not leap across neighbors toward some designated address. Communication takes place only along the pathways laid down through the initial organization of knowledge. Firmly embodied in the notion of locality is also that of autonomy, i.e., the absence of central supervision or control. We normally envision local processing steps to be triggered either by local events (i.e., a significant change in neighbors' activity) or totally at random -- timing information is not critical.

Whereas the principle of locality is a biological necessity in low level reasoning tasks such as perception, it also seems to dominate high level reasoning. Here the picture of parallel processors working autonomously and distributedly is only a useful metaphor, still it carries several advantages: There are only few data items participating in each inference step, and these items bear meaningful conceptual relationships to one another. Partial results are stored exactly where they will be useful. Computational steps can be performed in any order, and there is no need to remember which part of the knowledge has been processed and which part has not. (Note that many of these features are satisfied by the paradigm of logical deduction, for example, order invariance. However, in logical deduction partial results are stored in an unstructured database, thus rendering it difficult to identify those data items that are relevant to the next inference task).

Locality as a Principle for Bounding Rationality

Granted that we wish to conform to the architectural constraints imposed by locality, the question remains, what are the local procedures that we should adopt in our inference systems, knowing that the final outcome is bound to be merely an approximation to the inference that should be produced, were it not limited by locality considerations.

One way of going about choosing these procedures is, again, to consult our intuition and ask what are the basic inferential steps that seem to make up our everyday reasoning process. Once we assemble a reasonable collection of such procedures, they can be incorporated in software systems and used effectively for both inferencing and explanations. Indeed, this has been the predominant practice in most work in AI, especially in the area of expert systems. The Mycin experiment (Shortliffe, 1976) is a typical example of such practice. Adhering to the local policy of attributing to each consequent (of a knowledge rule) a degree of certainty that is function of the uncertainties of antecedents and the uncertainty tagging the rule itself, the combining functions were originally selected to match familiar experiences in medical diagnosis. They were later tuned to produce as reasonable results as possible, subject to the prevailing strategy of considering each rule in isolation. This strategy (often called rule-based, syntactical, extensional, truth-functional or componential) also governs the calculus of fuzzy logic, and in fact every commercially available uncertainty management system in existence. Similar strategies have also been guiding works in the area of truth maintenance systems (sacrificing completeness for locality) and inheritance networks (generalizing from resolving local conflicts among defaults to resolving global conflicts among arguments).

There is a more disciplined strategy of generating and positing rational local approximations to a given domain of problems. Rather than attempting to find a solution that approximates the entire domain, we adopt an exact solution to a simplified model of the domain. In other words, we can reason backward and ask what idealized models of reality lend themselves to exact solution, given the architectural constraints we wish to satisfy. If we find such a model, we then identify the computational procedures that make up an exact solution to that model, and posit them as the basic building blocks of rational (albeit approximate) reasoning. The advantage of following this strategy are that the procedures so discovered are guaranteed to be consistent, that they are known to produce exact results on at least a subset of problem domains, and that it is possible to determine in advance when the environment lies outside their range of applicability.

Returning to the Mycin example, we can ask what models of reality are solved coherently by the local, rule-based computations proposed by the Mycin architecture. The answer, phrased in probabilistic terms , is that the dependencies among the variables in any such reality must form a tree structure. Additionally, the updating functions in any such reality must form an ordered Abelian group, i.e., any monotone transformation on the likelihood ratio update (Hajek 1985, Heckerman 1986). Consequently, we conclude that the appropriate updating procedure to adopt is any monotone variant of the message passing technique developed for Bayesian trees (Pearl 1982, 1986).

We can go further and ask: Suppose we permit belief updating functions to have a slightly broader scope than those of Mycin; for example, suppose we admit combining functions that consider not only the rules that converge onto a given hypothesis but also those that diverge from the hypothesis. The answer in this case is that the dependencies in the domain must conform to a polytree structure (i.e., directed, singly connected networks) (Pearl 1986, 1988). Again, this structure dictates the precise nature of the updating functions which, unlike those of Mycin, provide a coherent account of bi-directional inferences (i.e., from evidence to hypothesis and from hypothesis to expectations). Thus, the propagation rules that emerge reflect richer patterns of qualitative reasoning and can be used to better approximate complex situations, where neither the assumption of tree dependence nor that of polytree dependence are valid.

The process of successive approximation can be continued in a similar manner, where at each step we widen the scope of the updating function. For example, we can add into consideration all rules that emanate from the antecedents of a given rule, all those that emanate from their consequents, and so on... In each level of locality, we obtain precise prescription of what the updating functions ought to be, and these functions, we conjecture, embody richer and richer structure of qualitative arguments that can be used to support reasoning and explanations.

This process of successively incrementing the scope of local operations resembles the practice of enforcing wider and wider levels of local consistency in Constraint Satisfaction Problems (so called K-consistency (Freuder 1982, Dechter 1987)). In truth maintenance systems it is embodied in the practice of applying hyperresolution with higher and higher arity (De Kleer 1989). However, whereas in these latter two applications, it is fairly obvious what the local operations should be at any given level of locality, the same is not true for reasoning under uncertainty -- additional analysis is needed to determine the precise nature of these operations.

References

Dechter, R., and Pearl, J. 1987c. Network-based hueristics for constraint- satisfaction problems. *Artificial Intelligence* 34 (no. 1): 1-38.

De Kleer, J. 1989. Personal Communication.

Frueder, E. C. 1982. A sufficient condition of backtrack-free search. *JACM* 29 (no. 1):24-32.

Hájek, P. 1985. Combining functions for certainty degrees in consulting systems. *Intl. Journal of Man-Machine Studies* 22:59-65.

Pearl, J. 1982. Reverend Bayes on inference engines: A distributed hierarchical approach. *Proc., Natl. Conf. on AI*, Pittsburgh, 133-36.

Pearl, J. 1986. Fusion, propagation and structuring in belief networks. *Artificial Intelligence* 29 (no. 3):241-88.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, San Mateo: Morgan Kaufmann Publishers.

Polya, G. 1954. *Patterns of plausible inference.* Princeton: Princeton University Press.

Shortliffe, E. H. 1976. *Computer-based medical consultation: MYCIN.* New York: Elsevier.