To appear in Paul Cohen (Ed.) *50 years to AI*.

TECHNICAL REPORT
R-331
June 2006

Two journeys into human reasoning

Judea Pearl
Cognitive Systems Laboratory
Computer Science Department
University of California, Los Angeles, CA 90024
judea@cs.ucla.edu

Abstract

This essay is a personal account of two research journeys motivated by a bold, yet common AI paradigm: whatever people do well machine should do better, if only we could listen carefully to the way people do it.

The first journey takes us to reasoning with uncertainty and the development of Bayesian Networks, the second to causal reasoning and the formalization of causal and counterfactual relationships.

Admittedly, none of these journeys unveiled how people actually do it, and none led to algorithms that consistently outperform humans, yet the results achieved through these efforts far exceed those obtained from fields outside AI, guided by less ambitious paradigms. I hope that the lesson would inspire more such journeys in the futures.

1 Reasoning with Uncertainty

My journey into uncertainly land was motivated by a busy mixture of observations and speculations.

1. The consistent agreement between plausible reasoning and probability calculus could not be coincidental, but strongly suggests that human intuition invokes some crude form of probabilistic computation.

2. In light of the speed and effectiveness of human reasoning, the computational difficulties that plagued earlier probabilistic systems could not be very fundamental and should be overcome by making the right choice of simplifying assumptions which humans store in their head.

3. The most crucial type of assumptions needed for probabilistic computation is conditional independence and graphical forms are the only plausible way in which such assumptions could be represented.

4. If a graphical knowledge representation could be found, then it should be possible to use the lines as message-passing channels, and we could then

update beliefs by parallel distributed computations, reminiscent of neural architectures.

5. If belief updating could be achieved by such distributed mechanisms, then the update would be easier to explain, since the flow of information would transverse conceptually meaningful paths.

6. If distributed updating were feasible, then probabilistic inference would be as easy to program and execute (even on a serial machine) as rule-based systems, and only simple control mechanisms would be required.

In hindsight, some of these speculations were rather naive. For example, fully distributed updating turned out to be correct only in singly connected networks, and some conditional independence relationships were shown to defy graphical representation altogether. Nevertheless, many of these speculations have survived the test of time. In recent years, belief networks have become a tool of great versatility and power and are now considered the most common representation scheme for probabilistic knowledge.

My curiosity to study distributed probabilistic computations on graphical models began brewing in the late 1970s, after I read Rumelhart's paper on reading comprehension [11]. In this paper, Rumelhart presented compelling evidence that text comprehension must be a distributed process that combines both top-down and bottom-up inferences. Strangely, this dual mode of inference, so characteristic of Bayesian analysis, did not match the capabilities of the ruling paradigms for uncertainty management in the 1970s. I thus began to explore the possibility of achieving distributed computation in a "pure" Bayesian framework, so as not to compromise its basic capacity to combine bi-directional inferences (i.e., predictive and abductive).

Not caring much about generality at that point, I pieced the simplest structure I could thin of (i.e., a tree) and tried to see if anything useful can be computed by assigning each variable a simple processor, forced to communicate only with its neighbors. This gave rise to the tree-propagation algorithm reported in [5] and, a year later, to belief propagation on poly-trees [3], which supported not only bi-directional inferences but also intercausal interactions, such as "explaining-away."

In the course of developing these algorithms, it became clear that *conditional independence* is the most fundamental relation behind the organization of probabilistic knowledge and the most crucial factor facilitating distributed computations. I therefore decided to investigate systematically how directed and undirected graphs could be used as a language for encoding, decoding, and reasoning with such independencies.

Probabilities and graphs are rather dissimilar mathematical objects, so, I was forced to as some fundamental questions about the relationships between the two. I began by asking how a directed acyclic graph (dag) can be extracted from a given probability distribution, whether the extracted dag is unique, what kind of distributions can be specified by a given dag, how we can read off the

independencies that are captured by the dag, and whether they match those associated with causal organizations.

This line of inquiry led to the formulation of Bayesian belief networks (a name I coined in 1986), their interpretation, and their use in probabilistic inference. In parallel, this inquiry also gave rise to the axiomatic theory of *graphoids* [10][8][1], in which directed and undirected graphs are treated as abstract mathematical objects, called *dependency models*, and in which direct from indirect dependencies are distinguished by "path separation."[1]

2 Reasoning with Cause and Effect

One regrettable step in this line of research was my failure to recognize causation as totally distinct ingredient, different from all the probabilistic notions that support evidential reasoning and statistical inference. Although I acknowledged the ubiquitous role of causation in conceptualizing the world, so intoxicated was I with the power of probabilities that I mistook causation to be subservient to probability. A statement such as "Causation is a language with which one can talk efficiently about certain structures of relevance relationships" [8] would embarrass me today, as it should embarrass thousands of readers of my latest boo (Causality, 2000) in which I made the following confession:

"Ten years ago, when I began writing Probabilistic Reasoning in Intelligent Systems (1988), I was working within the empiricist tradition. In this tradition, probabilistic relationships constitute the foundations of human knowledge, whereas causality simply provides useful ways of abbreviating and organizing intricate patterns of probabilistic relationships. Today, my view is quite different. I now take causal relationships to be the fundamental building blocs both of physical reality and of human understanding of that reality, and I regard probabilistic relationships as but the surface phenomena of the causal machinery that underlies and propels our understanding of the world."

What I discovered in these years is that causality does not mix with probability. If probabilities encode beliefs and how beliefs change with observations, causality encodes how probabilities themselves change or, more accurately, what aspects of a probability function remain invariant when others undergo change.

And this brings us to the story of my first encounter with causality.

I got my first hint of the dark world of causality during my junior year of high school. My science teacher, Dr. Feuchtwanger, introduced us to the study of logic by discussing the 19th century finding that more people died from smallpox inoculations than from smallpox itself. Some people used this information to argue that inoculation was harmful when, in fact, the data proved the opposite, that inoculation was saving lives by eradicating smallpox. "And here is where logic comes in," concluded Dr. Feuchtwanger, "To protect us from cause-effect

[1] Bayesian belief networks have been criticized for "substituting mathematics for clarity" (e.g., R. E. Barlow, in [4], page 117). In my judgment, it was precisely this conversion of networks and diagrams to mathematically defined objects that led to their current acceptance in practical reasoning systems.

fallacies of this sort." We were all enchanted by the marvels of logic, even though Dr. Feuchtwanger never actually showed us how logic protects us from such fallacies.

It doesn't, I realized in the early 1990's, as I began to seriously examine the relations between causation, logic and probability. Neither logic, nor any branch of mathematics had developed adequate tools for managing problems, such as the smallpox inoculations, involving cause-effect relationships. Even an innocent sentence such as "the rooster crow does not cause the sun to rise" could not be written in any mathematical notation, let alone processed by mathematical methods. Most of my colleagues considered causal vocabulary to be dangerous, avoidable, ill-defined, and nonscientific. "Causality is endless controversy," one of them warned. The accepted style in scientific papers was to write "A implies B" even if one really meant "A causes B," or to state "A is related to B" if one was thinking "A affects B."

Clearly, such denial of causal thought could not last forever. The influence of artificial intelligence gave my generation the expectation that intuition should be expressed, not suppressed. And causality, it turns out, is not nearly as nasty as her reputation suggests. Once I got past a few mental blocs, and began formalizing the obvious, I found causality to be smiling with clarity, bursting with new ideas and new possibilities. As the epilogue of my boo summarizes:

> "Causality is not mystical or metaphysical. It can be understood in terms of simple processes, and it can be expressed in a friendly mathematical language, ready for computer analysis."

This sweeping statement evoked obvious criticism in some traditional statistically-minded circles, but the findings that have ensued gave comfort to many new audiences: students of statistics who wondered (many still do) why instructors are reluctant to discuss causality in class; students of epidemiology who wondered (some still do) why simple concepts such as "confounding" are so terribly complex when expressed mathematically; students of economics and social science who demand to now the meaning of the parameters they are asked to estimate; and, naturally, students of artificial intelligence and cognitive science, who write programs and theories for knowledge discovery, causal explanations and causal speech.

I thin students of AI should draw inspiration from the that the mathematization of causal reasoning, which has affected and enlightened so many scientific disciplines, emanated from a naive, simple minded AI paradigm: whatever people do well machine can do as well, if only we listen carefully and formulate what we hear.

References

[1] D. Geiger, Graphoids: A qualitative framework for probabilistic inference, Ph.D. dissertation, University of California, Los Angeles, CA (1990).

[2] D. Geiger, T.S. Verma and J. Pearl, Identifying independence in Bayesian networks, *Networks* 20 (5) (1990) 507-534.

[3] J.H. Kim and J. Pearl, A computational model for combined causal and diagnostic reasoning in inference systems, in: *Proceedings IJCAI-83*, Karlsruhe, Germany (1983) 190-193.

[4] R.M. Oliver and J.Q. Smith, eds., *Influence Diagrams, Belief Nets and Decision Analysis* (John Wiley, Rexdale, Ontario, Canada, 1990).

[5] J. Pearl, Reverend Bayes on inference engines: A distributed hierarchical approach, in: *Proceedings AAAI National Conference on AI*, Pittsburgh, PA (1982) 133-136.

[6] J. Pearl, How to do with probabilities what people say you can't, in: *Proceedings 2nd IEEE Conference on AI Applications*, Miami, FL (1985) 6-12.

[7] J. Pearl, Evidential reasoning using stochastic simulation of causal models, *Artificial Intelligence* 32 (2) (1987) 245-258.

[8] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*, Morgan Kaufmann, Palo Alto, CA, 1988). Revised second printing (1991).

[9] J. Pearl, *Causality: Models, Reasoning and Inference*. Cambridge University Press, Cambridge, (2000).

[10] J. Pearl and A. Paz, 1989. On the logic of representing dependencies by graphs, in: *Proceedings 1986 Canadian AI Conference*, Montreal, Ontario, Canada (1986) 94-98.

[11] D.E. Rumelhart, Toward an interactive model of reading, Tech. Report #CHIP-56, University of California, La Jolla, CA (1976).