5 Causal Inference

Temporal inference, as we have seen, is relatively conceptually simple from a probabilistic perspective. It leads to a number of new link types and a fair amount of bookkeeping complication (the node-and-link constructs shown in the context of the fetch example won't win any prizes for elegance), but is not fundamentally conceptually problematic. The tricky issues that arise, such as the frame problem, are really more basic AI issues rather than temporal inference issues in particular.

Next, what about causality? This turns out to be a much subtler matter. There is much evidence that human causal inference is pragmatic and heterogeneous rather than purely mathematical (see discussion and references in Goertzel, 2006).

One illustration of this is the huge variance in the concept of causality that exists among various humans and human groups. Given this, it's not to be expected that PLN or any other logical framework could, in itself, give a thorough foundation for understanding causality. But even so, there are interest connections to be drawn between PLN structures and aspects of causal inference.

Predictive implication, as discussed above, allows us to discuss temporal correlation in a pragmatic way. But this brings us to what is perhaps the single most key conceptual point regarding causation: that *correlation and causation are distinct*. To take the classic example, if a rooster regularly crows before dawn, we do not want to infer that he causes the sun to rise.

In general, if X appears to cause Y, it may be actually due to Z causing both X and Y, with Y appearing later than X. We can only be sure that this is not the case if we have a way to identify alternate causes and test them in comparison to the causes we think are real. Or, as in the rooster/dawn case, we may have background knowledge that makes the "X causes Y" scenario intrinsically implausible in terms of the existence of potential causal mechanisms.

Let's consider this example in a little more detail. In the case of roosters and dawn, clearly, we have both implication and temporal precedence. Hence there will be a PredictiveImplication from "rooster crows" to "sun rises." But will the reasoning system conclude from this PredictiveImplication that, if a rooster happens to crow at 1 AM, the sun is going to rise really early that morning – say, at 2 AM? How is this elementary error avoided?

There are a couple answers here. The first has to do with the intension/extension distinction. It says: *The strength of this particular PredictiveImplication may be set high by direct observation, but it will be drastically lowered by inference from more general background knowledge.* Specifically, much of this inference will be *intensional* in nature, as opposed to the purely extensional information (direct evidence-counting) that is used to conclude that roosters crowing imply sun rising. We thus conclude that one signifier of bogus causal relationships is when

```
ExtensionalPredictiveImplication A B
```

has a high strength but

```
IntensionalPredictiveImplication A B
```

has a low strength. In the case of

```
A = rooster crows

B = sun rises
```

the count of the intensional relationship is much higher than that of the extensional relationship, so that the overall PredictiveImplication relationship comes out with a fairly low strength.

To put it more concretely, if the reasoning system had never seen roosters crow except an hour before sunrise, and had never seen the sun rise except after rooster crowing, the posited causal relation might indeed be created. What would keep it from surviving for long would be some knowledge about the mechanisms underlying sunrise. If the system knows that the sun is very large and rooster crows are physically insignificant forces, then, this tells it that there are many possible contexts in which rooster crows would not precede the sun rising. Conjectural reasoning about these possible contexts leads to negative evidence in favor of the implication

```
PredictiveImplication rooster_crows sun_rises
```

which counterbalances – probably overwhelmingly – the positive evidence in favor of this relationship derived from empirical observation.

Yet more concretely, one has the following pieces of evidence:

```
PredictiveImplication <.00, .99>

      small_physical_force

      movement_of_large_object

PredictiveImplication <.99,.99>

      rooster_crows

      small_physical_force

PredictiveImplication <.99, .99>

      sun_rises

      movement_of_large_object

PredictiveImplication <.00,.99>

      rooster_crows

      sun_rises
```

which must be merged with

```
      PredictiveImplication rooster_crows sun_rises  <1,c>
```

derived from direct observation. So it all comes down to: how much more confident is the system that a small force can't move a large object, than that rooster crows always precede the sunrise? How big is the parameter we've denoted c compared to the confidence we've arbitrarily set at .99?

Of course, for this illustrative example we've chosen only one of many general world-facts that contradicts the hypothesis that rooster crows cause the sunrise… in reality many, many such facts combine to effect this contradiction. This simple example just illustrates the general point that reasoning can invoke background knowledge to contradict the simplistic "correlation implies causation" conclusions that sometimes arise from direct empirical observation.

5.1 Aspects of Causality Missed By a Purely Logical Analysis

In this section we will briefly discuss a couple aspects of causal inference that seem to go beyond pure probabilistic logic – and yet are fairly easily integrable into a PLN-based framework. This sort of discussion highlights what we feel will ultimately be the greatest value of the PLN formalism: it formulates logical inference in a way that fits in naturally with a coherent overall picture of cognitive function. Here we will content ourselves with a very brief sketch of these ideas, as to pursue it further would lead us too far afield.

3.1.1 Simplicity of Causal Mechanisms

The first idea we propose has to do with the notion of causal mechanism. The basic idea is, given a potential cause-effect pair, to seek a concrete function mapping the cause into to the effect, and to consider the causality as more substantial if this function is simpler. In PLN terms, this means that one is not only looking at the IntensionalPredictiveImplication relationship underlying a posited causal relationship, but one is weighting the count of this relationship more highly if the Predicates involved in deriving the relationship are simpler. This heuristic for count-biasing means that one is valuing simple causal mechanisms as opposed to complex ones. The subtlety lies in the definition of the "simplicity" of a predicate, which relies on pattern theory as introduced above in the context of intensional inference.

3.1.2 Distal Causes, Enabling Conditions

As another indication of the aspects of the human judgment of causality that are omitted by a purely logical analysis, consider the distinction between *local* and *distal* causes. For example, does an announcement by Greenspan cause the market to change, or is he just responding to changed economic conditions on interest rates, and they are the ultimate cause? Or, to take another example, suppose a man named Bill drops a stone, breaking a car windshield. Do we want to blame (assign causal status to) Bill for dropping the stone that broke the car windshield, or his act of releasing the stone, or perhaps to the anger behind his action, or his childhood mistreatment by the owner of the car, or even the law of gravity pulling the rock down? Most commonly we would cite Bill as the cause because he was a free agent. But different causal ascriptions will be optimal in different contexts: typically, childhood mistreatment would be a mitigating factor in legal proceedings in such a case.

Related to this is the distinction between causes and so-called *enabling conditions*. Enabling conditions predictively imply their "effect," but they display no significant variation within the context considered pertinent. For, example oxygen is necessary to use a match to start a fire, but since it is normally always present, we usually ignore it as a cause, and it would be called an enabling condition. If it really is always present, we can ignore it in practice; the problem occurs when it is very often present but sometimes is not, as for example when new unforeseen conditions occur.

We believe it is fairly straightforward to explain phenomena like distal causes and enabling conditions, but only at the cost of introducing some notions that exist in Novamente but not in PLN proper. In Novamente, Atoms are associated with quantitative "importance" values as well as truth values. The importance value of an Atom has to do with how likely it is estimated to be that this Atom will be useful to the system in the future. There are short and long term importance values associated with different future time horizons. Importance may be assessed via PLN inference, but this is PLN inference based regarding propositions about how useful a given Atom has been over a given time interval.

It seems that the difference between a cause and an enabling condition often has to do with nonlogical factors. For instance, in Novamente PLN Atoms are associated not only with truth values but also with other numbers called attention values, including for instance "importances" value indicating the expected utility of the system to thinking about the Atom. For instance, the relationship

```
PredictiveImplication oxygen fire
```

may have a high strength and count, but it is going to have a very low importance unless the AI system in question is dealing with some cases where there is insufficient oxygen available to light fires. A similar explanation may help with the distinction between distal and local causes. Local causes are the ones associated with more important predictive implications – where importance needs to be assigned, by a reasoning system, based on inferences regarding which relationships are more likely to be useful in future inferences.