VII
PERCEIVING AND REASONING ABOUT A CHANGING WORLD

Previous chapters have detailed the construction of the OSCAR architecture. This chapter begins the application of that architecture to concrete problems in agent building. These are problems that must be faced in order to build a rational agent that is able to get around in a realistically complex world, but can be solved largely by using the OSCAR architecture as a tool rather than by fiddling with the architecture itself. By and large, such problems will be addressed by constructing reason-schemas that will enable OSCAR to reason about various aspects of the world.

A traditional problem of philosophical epistemology is that of explaining how it is possible for human beings to acquire knowledge of the external world. Essentially the same problem arises for artificial rational agents. The designers of such agents must provide them with procedures for accessing the world and forming reliable beliefs about it. Some knowledge may be built in, but in a complex changing environment, an agent cannot be equipped from its inception with all the information it needs. It must be capable of gathering new information by sensing its surroundings. This is perception, in a generic sense. All of an agent's knowledge of the world must be inferred from perception and background knowledge. The problems that an agent designer faces are essentially similar to those faced by the philosopher addressing the problem of our knowledge of the external world. These problems are at least threefold. First, perception need not be veridical—the world can be other than it appears. Second, perception is really a form of sampling. An agent cannot perceptually monitor the entire state of the world at all time. The best perception can do is provide the agent with images of small parts of the world at discrete times or over short time intervals. Perception provides momentary snapshots of scattered nooks and crannies at disparate times, and it is up to the agent's cognitive faculties to make inferences from these to a coherent picture of the world. Third, the world changes. The agent must be able to make inferences that enable it to keep track of an evolving world. For this it must be able to reason about both persistence and change, using knowledge of causal processes. Building an artificial agent that is able to perform these cognitive feats is no less difficult than solving the philosophical problem of our knowledge of the external world. In fact, the best way to solve the engineering problem is most likely to figure out how humans perform these tasks and then build AI systems that work similarly. This paper makes a start at providing this kind of analysis. The analysis is based upon decades of work in philosophical epistemology. The procedures that will be proposed are reason-schemas for defeasible reasoning. They have been implemented using the system of defeasible reasoning that is incorporated into the OSCAR architecture for rational agents.[1] Along the way, solutions will be proposed for the Frame Problem, the Qualification Problem, the Ramification Problem, and the Yale Shooting Problem.

1. Reasoning from Percepts

An agent must be capable of gathering new information by sensing its surroundings. This is perception, in a generic sense. Perception is a process that begins with the stimulation of sensors, and ends with beliefs about the agent's immediate surroundings. In artificial

[1] This architecture is detailed in Pollock [1995] and [1995a].

agents, this should be understood sufficiently broadly to include the input of information by a human operator. It is useful to draw a line between the last non-doxastic (non-belief) states in this process and the first beliefs. The production, in human beings, of the non-doxastic states is the subject of psychology and neuroscience. In AI it is the subject of research in machine vision. The reasoning from the beliefs is studied partly by epistemology and partly by psychology.[2] What is at issue here is the theory of the interface between the non-doxastic states and the beliefs.

I will refer to the final non-doxastic states from which beliefs are obtained in perception as *percepts*. Two mechanisms are possible for moving from percepts to beliefs. On the one hand, an agent could implement a purely automatic process whereby percepts give rise to beliefs automatically, and reasoning begins with the beliefs thus generated. This has the consequence that the beliefs thus produced are not *inferred* from anything, and hence are not rationally correctable. This has been a favorite view of philosophers regarding the nature of human cognition. But perception need not be veridical, and humans *can* discover that particular percepts are not accurate representations of the world, so the beliefs that are the automatic progeny of percepts cannot be beliefs about the world as such—they must be beliefs about the perceiver's sensory input. On this view, beliefs about physical objects (tables, chairs, people, plants, buildings, etc.) are inferred from beliefs about sensory input. I have attacked this view of human perception elsewhere.[3] The basic observation to be made about human beings is that when we perceive our surroundings, the resulting beliefs are usually beliefs about physical objects (tables, chairs, people, plants, buildings, etc.), and not beliefs about our own inner perceptual experiences. We *can* focus our attention on our perceptual experiences, forming beliefs about them by introspection, but that requires an explicit change of attention. On the other hand, because we can perform such a change of attention, we can evaluate the quality of the inference from those experiences to the beliefs about physical objects to which they give rise. This suggests that we should regard our reasoning as beginning from the percepts themselves, and not from beliefs about our percepts. Some philosophers have objected to this view on the grounds that reasoning is, by definition, a process of making transitions from beliefs to beliefs, and hence percepts cannot enter into reasoning. But this is just a verbal quibble. I have urged that the structure and evaluation of the transitions from percepts to beliefs about physical objects is sufficiently inference-like to warrant calling them inferences.[4] I propose to further substantiate that claim here by giving a precise description of the inferences and implementing them within OSCAR for incorporation into an artificial agent.

The preceding observations are just about human beings, but there are lessons to be drawn from them about rational agents in general. I see no reason why we *couldn't* build rational agents by having percepts automatically give rise to beliefs about percepts, and having all reasoning begin from those percepts. But that seems like needless duplication. If, as I will argue below, we can construct a system of perceptual reasoning that begins directly from the percepts, then beliefs about the percepts will in most cases be otiose and their production will be a needless burden on cognitive resources.

Accordingly, I will take the basic inference in perceptual reasoning to be from a percept to a conclusion about the world. This enables us to assign propositional contents to percepts. The content of a percept will be taken to be the same as the content of the belief for which the percept provides a reason. But of course, the percept is not the same thing as the belief. Having the percept consists of having a perceptual experience that is *as if* the belief were true. Given this understanding of the content of percepts, we can, as a first approximation, formulate the reasoning from percepts to beliefs as follows:

---

[2] Just as in linguistics, there is a competence/performance distinction to be made in the study of human reasoning. Epistemology constructs normative theories of competence, and psychology studies human performance. For more on this, see Pollock [1995].

[3] Pollock [1987] and [1995].

[4] See Pollock [1987].

(1) Having a percept with the content P is a defeasible reason for the agent to believe P.[5]

In this principle, the variable 'P' ranges over the possible contents of percepts. That range will depend upon the perceptual apparatus of the agent in question.

This formulation of the reasoning captures the obvious but important point that perceptual reasoning must be defeasible—appearances can be deceptive. However, that this formulation is not entirely adequate becomes apparent when we consider perceptual updating. The world changes, and accordingly percepts produced at different times can support inferences to conflicting conclusions. We can diagram this roughly as in figure 1, where t_1 is a later time than t_0, '⇢' symbolizes defeasible inference, and '⇠' symbolizes defeat relations. The reasoning seems to produce a case of collective defeat—the inferences to P and ~P defeat each other. But this should not be a case of collective defeat. The initial percept supports the belief that P holds *at* t_0, and the second percept supports the belief that P does not hold *at* t_1. These conclusions do not conflict. We can hold both beliefs simply by acknowledging that the world has changed.

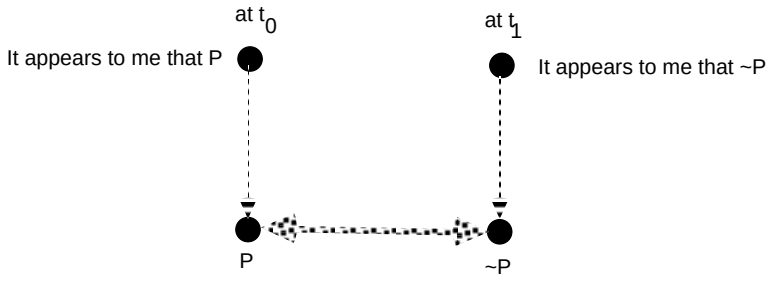


Figure 1. Perceptual updating.

We can accommodate this by building temporal reference into the belief produced by perception, and giving the percept a date. It will be convenient to build the time of the percept into the formula representing the content of the percept. Accordingly, I will adopt the convention of saying that a percept at time t with content P is a percept of P-at-t. This allows us to reformulate the above defeasible reason as follows:

PERCEPTION
Having a percept at time t with the content P is a defeasible reason for the agent to believe P-at-t.

We can then redraw the diagram of the reasoning as in figure 2, with the result that the apparent conflict has gone away.

There is a large literature on how, exactly, to build temporal reference into beliefs. This includes the literature on the situation calculus, temporal logic, possible worlds, etc. However, for present purposes, nothing very complicated is required. I will simply attach a term designating the time to the formula. No assumptions will be made about time other than that it is linearly ordered. More complex kinds of temporal reasoning may require a more sophisticated treatment, but I presume that nothing in this paper will be incompatible with such a treatment.

[5] This is based upon proposals in my [1967], [1971], [1974], [1987], and [1995].

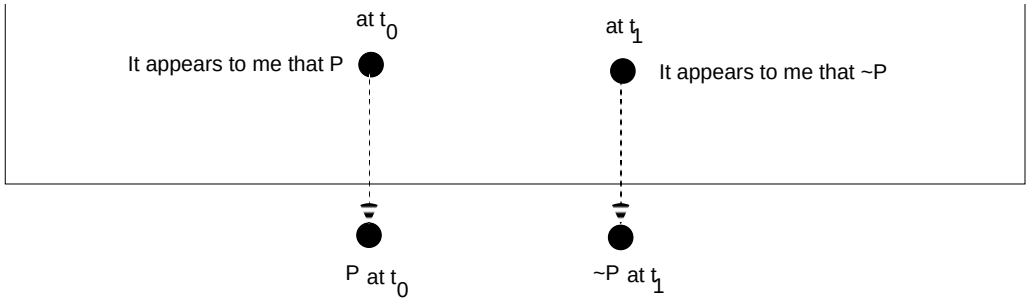


Figure 2. Perceptual updating revised.

2. Perceptual Reliability

When giving an account of a species of defeasible reasoning, it is as important to characterize the defeaters for the defeasible reasons as it is to state the reasons themselves. This paper assumes the theory of defeasible reasons and reasoning implemented in OSCAR and described in detail in Pollock [1995] and [1995a]. One of the central doctrines of that theory is that there are just two kinds of defeaters—*rebutting defeaters* and *undercutting defeaters*. Any reason for denying P-at-t is a rebutting defeater for PERCEPTION. An undercutting defeater for an inference from a belief in P to a belief in Q attacks the connection between P and Q rather than merely denying the conclusion. An undercutting defeater is a reason for the formula (P ⊗ Q) (read "It is false that P would not be true unless Q were true", or abbreviated as "P does not guarantee Q"). The only obvious undercutting defeater is a reliability defeater, which is of a general sort applicable to all defeasible reasons. Reliability defeaters result from observing that the inference from P to Q is not, under the present circumstances, reliable. To make this precise it is necessary to understand how reason-strengths work in OSCAR. Some reasons are better than others. In OSCAR, reason-strengths range from 0 to 1. Reason-strengths are calibrated by comparing them with the statistical syllogism. According to the statistical syllogism, when $r > 0.5$, "Bc & prob(A/B) = r" is a defeasible reason for "Ac", the strength of the reason being a function of r.[6] A reason of strength r is taken to have the same strength as an instance of the statistical syllogism from a probability of $2\cdot(r - 0.5)$. The inference rule PERCEPTION will have some strength r, although this may vary from agent to agent. The value of r should correspond roughly to the reliability of an agent's system of perceptual input in the circumstances in which it normally functions. PERCEPTUAL-RELIABILITY constitutes a defeater by informing us that under the present circumstances, perception is not as reliable as it is normally assumed to be:

PERCEPTUAL-RELIABILITY

Where R is projectible, r is the strength of PERCEPTION , and $s < 0.5\cdot(r + 1)$, "R-at-t, and the probability is less than or equal to r of P's being true given R and that I have a percept with content P" is an undercutting defeater for PERCEPTION as a reason of strength $\geq r$.

The projectibility constraint in this principle is a perplexing one. To illustrate its need, suppose I have a percept of a red object, and am in improbable but irrelevant circumstances of some type C_1. For instance, C_1 might consist of my having been born in the first second of the first minute of the first hour of the first year of the twentieth century. Let C_2 be

[6] This is a slight oversimplification. See my [1990] for a detailed discussion of the statistical syllogism.

circumstances consisting of wearing rose-colored glasses. When I am wearing rose-colored glasses, the probability is not particularly high that an object is red just because it looks red, so if I were in circumstances of type C_2, that would quite properly be a reliability defeater for a judgment that there is a red object before me. However, if I am in circumstances of type C_1 but not of C_2, there should be no reliability defeater. The difficulty is that if I am in circumstances of type C_1, then I am also in the disjunctive circumstances $(C_1 \vee C_2)$. Furthermore, the probability of being in circumstances of type C_2 given that one is in circumstances of type $(C_1 \vee C_2)$ is very high, so the probability is not high that an object is red given that it looks red to me but I am in circumstances $(C_1 \vee C_2)$. Consequently, if $(C_1 \vee C_2)$ were allowed as an instantiation of R in PERCEPTUAL-RELIABILITY, being in circumstances of type C_1 would suffice to indirectly defeat the perceptual judgment.

The preceding examples show that the set of circumstance-types appropriate for use in PERCEPTUAL-RELIABILITY is not closed under disjunction. This is a general characteristic of projectibility constraints. The need for a projectibility constraint in induction is familiar to most philosophers (although unrecognized in many other fields).[7] I showed in Pollock [1990] that the same constraint occurs throughout probabilistic reasoning, and the constraint on induction can be regarded as derivative from a constraint on the statistical syllogism.[8] However, similar constraints occur in other contexts and do not appear to be derivative from the constraints on the statistical syllogism. The constraint on reliability defeaters is one example of this, and another example will be given below. Unfortunately, at this time there is no generally acceptable theory of projectibility. The term "projectible" serves more as the label for a problem than as an indication of the solution to the problem.

PERCEPTUAL-RELIABILITY constitutes a defeater by informing us that under the present circumstances, perception is not as reliable as it is normally assumed to be. Notice, however, that this should not prevent our drawing conclusions with a weaker level of justification. The probability recorded in PERCEPTUAL-RELIABILITY should function merely to weaken the strength of the perceptual inference rather than completely blocking it. This can be accomplished by supplementing PERCEPTION with the following rule:

DISCOUNTED-PERCEPTION

Where R is projectible, r is the strength of PERCEPTION , and $0.5 < s < 0.5 \cdot (r + 1)$, having a percept at time t with the content P and the belief "R-at-t, and the probability is less than s of P's being true given R and that I have a percept with content P" is a defeasible reason of strength $2 \cdot (s - 0.5)$ for the agent to believe P-at-t.

DISCOUNTED-PERCEPTION must be defeasible in the same way PERCEPTION is:

PERCEPTUAL-UNRELIABILITY

Where A is projectible and $s^* < s$, "A-at-t, and the probability is less than or equal to s^* of P's being true given A and that I have a percept with content P" is a defeater for DISCOUNTED-PERCEPTION.

In a particular situation, the agent may know that a number of facts hold each of which is sufficient to lower the reliability of perception. The preceding principles have the consequence that the only undefeated inference from the percept will be that made in accordance with the weakest instance of DISCOUNTED-PERCEPTION.[9]

---

[7] The need for the projectibility constraint on induction was first noted by Goodman [1955].

[8] The material on projectibility in my [1990] has been collected into a paper and reprinted in my [1994b].

[9] In such a case, we might also know that the reliability of perception on the combination of facts is higher than it is on the individual facts (interfering considerations might cancel out). In that case, we should be able to make an inference from the percept to its content in accordance with that higher probability. However, that inference can be made straightforwardly using the statistical syllogism, and does not require any further principles specifically about perception.

3. Implementation

Epistemic reasoning begins from contingent information input into the system in the form of percepts. Percepts are encoded as structures with the following fields:

- percept-content—a formula, without temporal reference built in.
- percept-clarity—a number between 0 and 1, indicating how strong a reason the percept provides for the conclusion of a perceptual inference.
- percept-date—a number.

When a new percept is presented to OSCAR, an inference-node of kind :percept is constructed, having a node-formula that is the percept-content of the percept (this includes the percept-date). This inference-node is then inserted into the inference-queue for processing.

We can implement PERCEPTION as a simple forwards-reason:

```
(def-forwards-reason PERCEPTION
   :forwards-premises "(p at time)" (:kind :percept)
   :conclusions "(p at time)"
   :variables p time
   :defeasible? t
   :strength .98
   :description "When information is input, it is defeasibly reasonable to believe it.")
```

The strength of .98 has been chosen arbitrarily.

PERCEPTUAL-RELIABILITY was formulated as follows:

> *PERCEPTUAL-RELIABILITY*
>
> Where R is projectible, r is the strength of PERCEPTION , and $s < 0.5 \cdot (r + 1)$, "R-at-t, and the probability is less than or equal to s of P's being true given R and that I have a percept with content P" is a defeater for PERCEPTION.

It seems clear that this should be treated as a backwards-reason. That is, given an interest in the undercutting defeater for PERCEPTION, this reason schema should be activated, but if the reasoner is not interested in the undercutting defeater, this reason schema should have no effect on the reasoner. However, treating this as a simple backwards-reason is impossible, because there are no constraints (other than projectibility) on R. We do not want interest in the undercutting defeater to lead to interest in every projectible R. Nor do we want the reasoner to spend its time trying to determine the reliability of perception given everything it happens to know about the situation. This can be avoided by making this a degenerate backwards-reason, taking R-at-t (where t is the time of the percept) and the probability premise to be forwards-premises. This suggests to the following definition:

```
(def-backwards-undercutter PERCEPTUAL-RELIABILITY
   :defeatee PERCEPTION
   :forwards-premises
   "((the probability of p given ((I have a percept with content p) & R)) <= s)"
     (:condition (s < 0.99))
   "(R at time0)"
      (:condition (and (projectible R) (time0 < time)))
   :backwards-premises "(R at time)"
   :variables p time R time0 s
   :defeasible? t
   :description "When perception is unreliable, it is not reasonable to accept its representations.")
```

(DEF-BACKWARDS-UNDERCUTTER is a variant of DEF-BACKWARDS-REASON that computes the reason-conclusions for us.) For now, I will take the projectible formulas to

be any conjunctions of literals, although it must be recognized that this is simplistic and must ultimately be refined.

A problem remains for this implementation. PERCEPTUAL-RELIABILITY requires us to know that R is true at the time of the percept. We will typically know this only by inferring it from the fact that R was true earlier. The nature of this inference is the topic of the next section. Without this inference, it is not possible to give interesting illustrations of the implementation just described, so that will be postponed until section six.

4. Temporal Projection

The reason-schema PERCEPTION enables an agent to draw conclusions about its current surroundings on the basis of its current percepts. However, that is of little use unless the agent can also draw conclusions about its current surroundings on the basis of earlier (at least fairly recent) percepts. For instance, imagine a robot whose task is to visually check the readings of two meters and then press one of two buttons depending upon which reading is higher. This should not be a hard task, but if we assume that the robot can only look at one meter at a time, it will not be able to acquire the requisite information about the meters using only the reason-schema PERCEPTION. The robot can look at one meter and draw a conclusion about its value, but when the robot turns to read the other meter, it no longer has a percept of the first and so is no longer in a position to hold a justified belief about what that meter reads *now*. This is a reflection of the observation made at the beginning of the paper that perception samples bits and pieces of the world at disparate times, and an agent must be supplied with cognitive faculties enabling it to build a coherent picture of the world out of those bits and pieces. In the case of our robot, what is needed is some basis for believing that the first meter still reads what it read a moment ago. In other words, the robot must have some basis for regarding the meter reading as a *stable property*—one that tends not to change quickly over time.

It is natural to suppose that a rational agent, endowed with the ability to learn by induction, can discover inductively that some properties (like the meter readings) are stable to varying degrees, and can bring that knowledge to bear on tasks like the meter-reading task. However, I argued long ago (Pollock [1974]) that such inductive learning is not epistemically possible—it presupposes the very stability that is the object of the learning. The argument for this somewhat surprising conclusion is as follows. To say that a property is stable is to say that objects possessing it tend to retain it. To confirm this inductively, an agent would have to re-examine the same object at different times and determine whether the property has changed. The difficulty is that in order to do this, the agent must be able to reidentify the object as the same object at different times. Although this is a complex matter, it seems clear that the agent makes essential use of the perceptible properties of objects in reidentifying them. If all perceptible properties fluctuated wildly, we would be unable to reidentify anything. If objects tended to exchange their perceptible properties abruptly and unpredictably, we would be unable to tell which object was which.[10] The upshot of this is that it is epistemically impossible to investigate the stability of perceptible properties inductively without presupposing that most of them tend to be stable. If we make that general supposition, then we can use induction to refine it by discovering that some perceptible properties are more stable than others, that particular properties tend to be unstable under specifiable circumstances, etc. But our conceptual framework must include a general presumption of stability for perceptible properties before any of this refinement can take place. In other words, the built-in epistemic arsenal of a rational agent must include reason-schemas of the following sort for at least some choices of P:

(2) If $t_0 < t_1$, believing P-at-t_0 is a defeasible reason for the agent to believe P-at-t_1.

---

[10] A more detailed presentation of this argument can be found in chapter six of Pollock [1974].

Principle (2) amounts to a presumption that P's being true is a stable property of a time (i.e., a stable fluent, to use the jargon of the situation calculus). A stable property is one such that if it holds at one time, the probability is high that it will continue to hold at a later time. Let ρ be the probability that P will hold at time t+1 given that it holds at time t. Assuming independence, it follows that the probability that P will hold at time (t+Δt) given that it holds at time t is $\rho^{\Delta t}$. In other words, the strength of the presumption that a stable property will continue to hold over time decays as the time interval increases. This is important for understanding the logic of reasoning about stable properties. To illustrate, consider what I will call *the perceptual updating problem*. Suppose an agent has a percept of P at time t_0, and a percept of ~P at a later time t_1. What an agent *should* conclude (defeasibly) under these circumstances is that the world has changed between t_0 and t_1, and although P was true at t_0, it is no longer true at t_1 and hence no longer true at a later time t_2. If we attempt to reconstruct this reasoning using principle (2), we do not seem to get the right answer. Principle (2) produces the inference-graph of figure 3, and it is a straightforward case of collective defeat. This is intuitively incorrect.

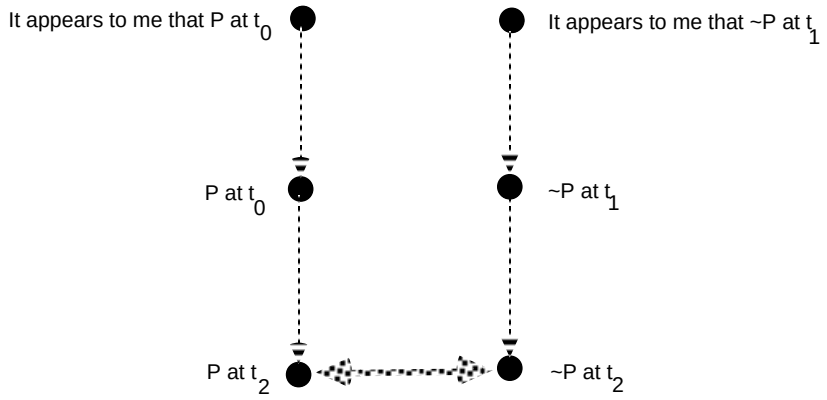


Figure 3. The perceptual updating problem.

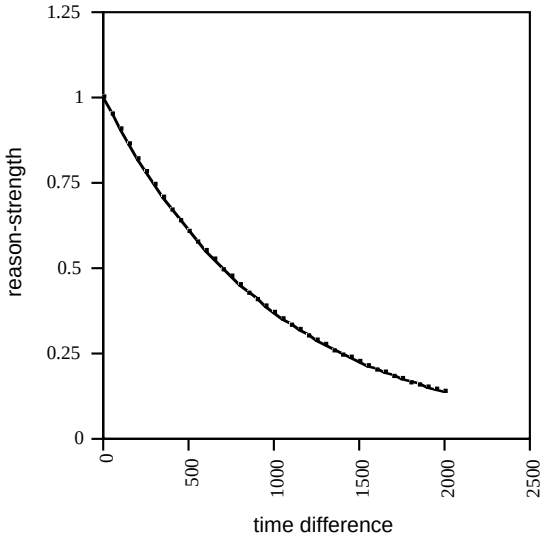


Figure 4. Decaying reason-strengths.

The solution to getting the reasoning to come out right is to embrace the observation that the strength of the presumption that a stable property will continue to hold over time decays as the time interval increases, and build this into principle (2). The strength of the reason provided by principle (2) must decrease as $(t_1 - t_0)$ increases. This will have the result that the support for ~P-at-t_2 is greater than the support for P-at-t_2, and hence the latter is defeated but the former is not. I propose then that we take the reason-strength of temporal projection to have the form $\rho^{\Delta t}$ where ρ is a constant I will call *the temporal-decay factor*. I have fairly arbitrarily set ρ to .999 (OSCAR uses an arbitrary time-scale anyway—the only constraint is that reasoning should be completed in a reasonable amount of time). This produces the decay curve of figure 4.

A probability of $\rho^{\Delta t}$ corresponds to a reason-strength of $2\cdot(\rho^{\Delta t} - .5)$. $\rho^{\Delta t} > .5$ iff $\Delta t < \log(.5)/\log(\rho)$, so the proposal is that we reformulate (1) as follows:

(2) When $\Delta t < \log(.5)/\log(\rho)$, believing *P-at-t* is a defeasible reason of strength $2\cdot(\rho^{\Delta t} - .5)$ for the agent to believe *P-at-*$(t+\Delta t)$.

5. Temporal Projectibility

Let P and Q be unrelated propositions, and consider the inference-graph of figure 5. In this inference-graph, '⟶' symbolizes deductive inferences, and bars connecting arrows indicate that the inference is from multiple premises. In this inference-graph, the conclusion Q-at-t_2 is undefeated. But this is unreasonable. Q-at-t_2 is inferred from $(P \vee Q)$-at-t_2. $(P \vee Q)$ is expected to be true at t_2 only because it was true at t_0, and it was only true at t_0 because P was true at t_0. This makes it reasonable to believe $(P \vee Q)$-at-t_2 only insofar as it is reasonable to believe P-at-t_2, but the latter is defeated.

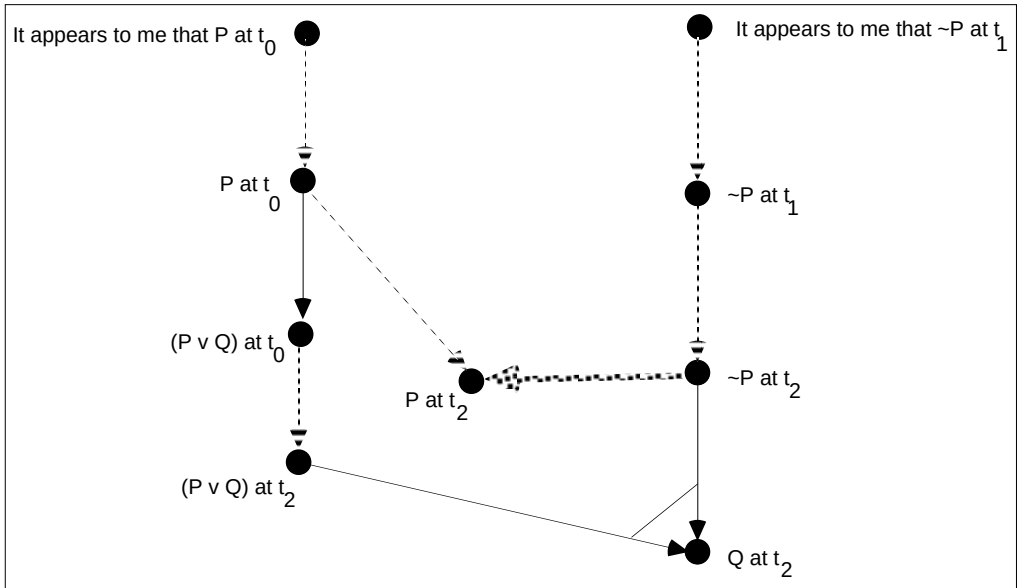


Figure 5. The need for a temporal projectibility constraint.

This appears to be a projectibility problem, analogous to that discussed above in connection with reliability defeaters. In temporal projection, the use of arbitrary disjunctions, and other non-projectible constructions, must be precluded. It is unclear precisely what the connection is between the projectibility constraint involved in temporal

projection and that involved in induction, so I will refer to it neutrally as "temporal-projectibility". Notice that in temporal-unprojectibility, disjunctions are not the only culprits. The ascriptions of properties to objects will generally be projectible, but the negations of such ascriptions need not be. For instance, "x is red" would seem to be temporally projectible. But "x is not red" is equivalent to a disjunction "x is blue or green or yellow or orange or ...", and as such it would seem to be temporally unprojectible. On the other hand, there are "bivalent" properties, like "dead" and "alive" for which the negation of an ascription is projectible because it is equivalent to ascribing the other (temporally-projectible) property.

Using the concept of temporal-projectibility, temporal projection should be reformulated as follows:

TEMPORAL-PROJECTION

If P is temporally-projectible and $\Delta t < \log(.5)/\log(\rho)$ then believing P-at-t is a defeasible reason of strength $2\cdot(\rho^{\Delta t} - .5)$ for the agent to believe P-at-$(t+\Delta t)$.

TEMPORAL-PROJECTION is based on an *a-priori* presumption of stability for temporally-projectible properties. However, it must be possible to override or modify the presumption by discovering that the probability of P's being true at time t+1 given that P is true at time t is something other than the constant ρ. This requires the following defeater:

DEFEAT-FOR-TEMPORAL-PROJECTION

"The probability of P-at-(t+1) given P-at-t $\neq \rho$" is a conclusive undercutting defeater for TEMPORAL-PROJECTION.

If we know that the probability of P-at-(t+1) given P-at-t is σ where $\sigma \neq \rho$, we may still be able to project P forwards in time, but now the inference will be based upon the statistical syllogism and the known high probability rather than an *a-priori* principle of temporal projection.

6. Implementing Temporal Projection

In order to implement TEMPORAL-PROJECTION, we must have a test for the temporal-projectibility of formulas. This is a problem, because as indicated above, I do not have a theory of temporal-projectibility to propose. For present purposes, I will finesse this by assuming that atomic formulas, negations of atomic formulas whose predicates are on a list *bivalent-predicates*, and conjunctions of the above, are temporally projectible. This will almost certainly be inadequate in the long run, but it will suffice for testing the proposed reason-schemas.

It seems clear that TEMPORAL-PROJECTION must be treated as a backwards-reason. That is, given some fact P-at-t, we do not want the reasoner to automatically infer P-at-t* for every one of the infinitely many times t* > t. An agent should only make such an inference when the conclusion is of interest. For the same reason, the premise P-at-t_0 should be a forwards-premise rather than a backwards-premise—we do not want the reasoner adopting interest in P-at-t for every t < t*. I propose to handle this by implementing it as a backwards reason. This will have the effect that when the reasoner adopts interest in P-at-t, it will check to see whether it already has a conclusion of the form P-at-t* for t* < t, and if so it will infer P-at-t. This can be done in either of two ways. We could use a mixed-backwards-reason:

```
(def-backwards-reason TEMPORAL-PROJECTION
   :conclusions  "(p at time)"
   :condition  (and (temporally-projectible p) (numberp time))
   :forwards-premises
      "(p at time0)"
```

```
   :backwards-premises
      "(time0 < time)"
      "((time* - time0) < 693)"
   :variables p time0 time
   :defeasible? T
   :strength  (- (* 2 (expt *temporal-decay* (- time time0))) 1)
   :description
   "It is defeasibly reasonable to expect temporally projectible truths to remain unchanged.")
```

This requires the reasoner to engage in explicit arithmetical reasoning about whether (time0 < time). It is more efficient to make this a condition on the forwards-premise rather than an independent premise. This produces a degenerate backwards-reason:

```
(def-backwards-reason TEMPORAL-PROJECTION
   :conclusions  "(p at time)"
   :condition  (and (temporally-projectible p) (numberp time))
   :forwards-premises
      "(p at time0)"
      (:condition (and (time0 < time*) ((time* - time0) < 693)))
   :variables p time0 time
   :defeasible? T
   :strength  (- (* 2 (expt *temporal-decay* (- time time0))) 1)
   :description
   "It is defeasibly reasonable to expect temporally projectible truths to remain unchanged.")
```

PROBABILISTIC-DEFEAT-FOR-TEMPORAL-PROJECTION is implemented as a conclusive degenerate backwards-reason:

```
(def-backwards-undercutter PROBABILISTIC-DEFEAT-FOR-TEMPORAL-PROJECTION
  :defeatee  TEMPORAL-PROJECTION
   :forwards-premises
     "((the probability of (p at (t + 1)) given (p at t)) = s)"
       (:condition (not (s = *temporal-decay*)))
   :variables  p s time0 time)
```

To illustrate, consider the perceptual updating problem:

```
=========================================================================
Problem number 6:  This is the perceptual updating problem.  First, Fred looks red to me.  Later, Fred looks
blue to me.  What should I conclude  about the color of Fred?

Forwards-substantive-reasons:
       PERCEPTION

Backwards-substantive-reasons:
       TEMPORAL-PROJECTION
       INCOMPATIBLE-COLORS

Inputs:
       (the color of Fred is red) : at cycle 1 with justification 1.0
       (the color of Fred is blue) : at cycle 30 with justification 1.0

Ultimate epistemic interests:
    (? x)((the color of Fred is x) at 50)    degree of interest = 0.5
=========================================================================
THE FOLLOWING IS THE REASONING INVOLVED IN THE SOLUTION
Nodes marked DEFEATED have that status at the end of the reasoning.
```

```
                              # 1
                              interest: ((the color of Fred is y0) at 50)
                              This is of ultimate interest
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
It appears to me that ((the color of Fred is red) at 1)
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 # 1
 It appears to me that ((the color of Fred is red) at 1)
 # 2
 ((the color of Fred is red) at 1)
 Inferred by:
         support-link #1 from { 1 } by PERCEPTION
 undefeated-degree-of-support = 0.98
 # 3
 ((the color of Fred is red) at 50)          DEFEATED
 undefeated-degree-of-support = 0.904
 Inferred by:
         support-link #2 from { 2 } by TEMPORAL-PROJECTION  defeaters: { 7 }   DEFEATED
 This discharges interest 1
                              # 5
                              interest: ~((the color of Fred is red) at 50)
                              Of interest as a defeater for support-link 2 for node 3
         =============================================
         Justified belief in ((the color of Fred is red) at 50)
         with undefeated-degree-of-support 0.904
         answers #<Query 1: (? x)((the color of Fred is x) at 50)>
         =============================================
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
It appears to me that ((the color of Fred is blue) at 30)
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 # 4
 It appears to me that ((the color of Fred is blue) at 30)
 # 5
 ((the color of Fred is blue) at 30)
 Inferred by:
         support-link #3 from { 4 } by PERCEPTION
 undefeated-degree-of-support = 0.98
 # 6
 ((the color of Fred is blue) at 50)
 Inferred by:
         support-link #4 from { 5 } by TEMPORAL-PROJECTION  defeaters: { 8 }
 undefeated-degree-of-support = 0.960
 This discharges interest 1
                              # 9
                              interest: ~((the color of Fred is blue) at 50)
                              Of interest as a defeater for support-link 4 for node 6
         =============================================
         Justified belief in ((the color of Fred is blue) at 50)
         with undefeated-degree-of-support 0.960
         answers #<Query 1: (? x)((the color of Fred is x) at 50)>
         =============================================
 # 7
 ~((the color of Fred is red) at 50)
 Inferred by:
         support-link #5 from { 6 } by INCOMPATIBLE-COLORS
 undefeated-degree-of-support = 0.960
```

```
 defeatees: { link 2 for node 3 }
          vvvvvvvvvvvvvvvvvvvvvvvvvvv
          #<Node 3> has become defeated.
          vvvvvvvvvvvvvvvvvvvvvvvvvvv
          ========================================
          Lowering the undefeated-degree-of-support of ((the color of Fred is red) at 50)
          retracts the previous answer to #<Query 1: (? x)((the color of Fred is x) at 50)>
          ========================================
================== ULTIMATE EPISTEMIC INTERESTS ==================
 Interest in (? x)((the color of Fred is x) at 50)
 is answered by node 6:  ((the color of Fred is blue) at 50)
------------------------------------------------
```

The previous example was produced by the macro MAKE-SIMULATION-PROBLEM and run by executing (SIMULATE-OSCAR). The following is an example of the use of MAKE-SIMULATION-PROBLEM:

```
(make-simulation-problem
 :number 8
 :message
 "First, Fred looks red to me.  Later, I am informed by Merrill that I am then
wearing blue-tinted glasses.  Later still, Fred looks blue to me.  All along, I know that the
probability is not high of Fred being blue given that Fred looks blue to me, but I am
wearing blue tinted glasses.  What should I conclude about the color of Fred?"
 :reasons
 *INDEXICAL-PERCEPTION*
 *PERCEPTION*
 *RELIABLE-INFORMANT*
 *INDEXICAL-PERCEPTUAL-RELIABILITY*
 *PERCEPTUAL-RELIABILITY*
 *TEMPORAL-PROJECTION*
 *INDEXICAL-INCOMPATIBLE-COLORS*
 :start-time 50
 :inputs
 (100  "(the color of Fred is red)" 0.8)
 (120  "(Merrill reports that I_am_wearing_blue_tinted_glasses)" 1.0)
 (130  "(the color of Fred is blue)" 0.8)
 :premises
 ("((the probability of (the color of Fred is blue) given
   ((I have a percept with content (the color of Fred is blue)) & I_am_wearing_blue_tinted_glasses)) <= .8)"
   1.0)
 ("(Merrill is a reliable informant)" 1.0)
 :interests
 ("(? x)(the color of Fred is x)" 0.65)
 )
```

All fields are optional, and can be omitted in defining the problem. The number is used for identifying the problem. The message is a string describing the problem. If the start-time is non-nil, *cycle* is set to that number when the reasoning begins. The start-time defaults to 0. Inputs are triples, consisting of a time (a cycle number) at which a percept is produced, a string (pretty-formula) giving the content of the percept, and a number determining the percept-clarity. Premises are pairs or triples. The first member of the premise is a pretty-formula, the second member is the degree of justification, and the optional third number is the cycle at which the premise is supplied to the reasoner. If the third number is missing, the premise is supplied at the beginning of the reasoning. All premises and percepts are treated as background knowledge. Simulation problems are run by first executing the definition of the problem (using MAKE-SIMULATION-PROBLEM)

and then executing (SIMULATE-OSCAR). The source code for MAKE-SIMULATION-PROBLEM and SIMULATE-OSCAR, together with the definitions of all the utility functions and reason-schemas discussed in this chapter, are contained in the file "Perception-causes". The problems are defined in the file "P/C-examples". These simulation problems are usually run with reductio-off, as reductio is not needed for any of the problems and slows the reasoning. However, they can also be run with reductio-on.

Now let us return to the problem noted above for PERCEPTUAL-RELIABILITY. This is that we will typically know R-at-t only by inferring it from R-at-t_0 for some $t_0 < t$ (by TEMPORAL-PROJECTION). TEMPORAL-PROJECTION is a backwards-reason. That is, given some fact P-at-t, the reasoner only infers P-at-t* (for $t^* > t$) when that conclusion is of interest. Unfortunately, in PERCEPTUAL-RELIABILITY, R-at-t is not an interest, and so it will not be inferred from R-at-t_0 by TEMPORAL-PROJECTION. This difficulty can be circumvented by formulating PERCEPTUAL-RELIABILITY with an extra forwards-premise R-at-t_0 which is marked as a *clue*, and a backwards-premise R-at-t:

```
(def-backwards-undercutter PERCEPTUAL-RELIABILITY
   :defeatee PERCEPTION
   :forwards-premises
   "((the probability of p given ((I have a percept with content p) & R)) <= s)"
     (:condition (s < 0.99))
   "(R at time0)"
      (:condition (and (projectible R) (time0 < time)))
      (:clue? t)
   :backwards-premises "(R at time)"
   :variables p time R time0 s
   :defeasible? t
   :description "When perception is unreliable, it is not reasonable to accept its representations.")
```

The difference between ordinary forwards-premises and clues is that when a clue is instantiated by a node, that node is not inserted into the basis for the inference. The function of clues is to guide the reasoning. Thus in an application of PERCEPTUAL-RELIABILITY, if R-at-t_0 is concluded, this suggests that R-at-t is true and leads to an interest in it, which can then be inferred from R-at-t_0 by TEMPORAL-PROJECTION. An example of such reasoning follows:

```
========================================================================
Problem number 1:  Fred looks red to me.  However, I also know that my surroundings are illuminated by
red light.  All along, I know that the probability is not high of Fred being red given that Fred looks red to me,
but my surroundings are illuminated by red light.  What should I conclude about the color of Fred?

Forwards-substantive-reasons:
      PERCEPTION

Backwards-substantive-reasons:
      PERCEPTUAL-RELIABILITY
      TEMPORAL-PROJECTION

Inputs:
      (The color of Fred is red) : at cycle 1 with justification 1.0

Given:
      (My surroundings are illuminated by red light (at 0))     justification = 1.0
      ((The probability of (The color of Fred is red) given
        ((I have a percept with content (The color of Fred is red)) & My surroundings are illuminated by red
light))
       <= 0.8)     justification = 1.0
```

```
Ultimate epistemic interests:
    (? x)((the color of Fred is x) at 1)    degree of interest = 0.75
===========================================================================
THE FOLLOWING IS THE REASONING INVOLVED IN THE SOLUTION
Nodes marked DEFEATED have that status at the end of the reasoning.

  # 1
  (my surroundings are illuminated by red light at 0)
  given
  # 2
  ((the probability of (the color of Fred is red) given ((I have a percept with content (the color of Fred is red))
& my surroundings are illuminated by red light)) <= 0.8)
  given
                                  # 1
                                  interest: ((the color of Fred is y0) at 1)
                                  This is of ultimate interest
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
It appears to me that ((the color of Fred is red) at 1)
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
  # 3
  It appears to me that ((the color of Fred is red) at 1)
  # 4
  ((the color of Fred is red) at 1)                DEFEATED
  Inferred by:
              support-link #3 from { 3 } by PERCEPTION  defeaters: { 6 }   DEFEATED
  This discharges interest 1
                                  # 2
                                  interest: (((it appears to me that (the color of Fred is red)) at 1) ⊗ ((the color of
                                    Fred is red) at 1))
                                  Of interest as a defeater for support-link 3 for node 4
              =========================================
              Justified belief in ((the color of Fred is red) at 1)
              answers #<Query 1: (? x)((the color of Fred is x) at 1)>
              =========================================
                                  # 4
                                  interest: (my surroundings are illuminated by red light at 1)
                                  For interest 2 by PERCEPTUAL-RELIABILITY
                                  This interest is discharged by node 5
  # 5
  (my surroundings are illuminated by red light at 1)
  Inferred by:
              support-link #4 from { 1 } by TEMPORAL-PROJECTION
  This discharges interest 4
  # 6
  (((it appears to me that (the color of Fred is red)) at 1) ⊗ ((the color of Fred is red) at 1))
  Inferred by:
              support-link #5 from { 2 , 5 } by PERCEPTUAL-RELIABILITY
  defeatees: { link 3 for node 4 }
  This node is inferred by discharging interest #2
              vvvvvvvvvvvvvvvvvvvvvvvvvvvvvv
              #<Node 4> has become defeated.
              vvvvvvvvvvvvvvvvvvvvvvvvvvvvvv
              =========================================
              Lowering the undefeated-degree-of-support of ((the color of Fred is red) at 1)
              retracts the previous answer to #<Query 1: (? x)((the color of Fred is x) at 1)>
              =========================================
================== ULTIMATE EPISTEMIC INTERESTS ===================
```

```
 Interest in (? x)((the color of Fred is x) at 1)
 is unsatisfied.
--------------------------------------------------
```

Note that node 1 is not listed as a premise of the inference to node 6.

DISCOUNTED-PERCEPTION and PERCEPTUAL-UNRELIABILITY can be implemented similarly:

```
(def-forwards-reason DISCOUNTED-PERCEPTION
   :forwards-premises
   "((the probability of p given ((I have a percept with content p) & R)) <= s)"
      (:condition (and (projectible R) (0.5 < s) (s < 0.99)))
   "(p at time)"
     (:kind :percept)
   "(R at time0)"
      (:condition (time0 < time))
      (:clue? t)
   :backwards-premises "(R at time)"
   :conclusions "(p at time)"
   :variables p time R time0 s
   :strength  (2 * (s - 0.5))
   :defeasible? t
   :description "When information is input, it is defeasibly reasonable to believe it.")

(def-backwards-undercutter PERCEPTUAL-UNRELIABILITY
   :defeatee DISCOUNTED-PERCEPTION
   :forwards-premises
   "((the probability of p given ((I have a percept with content p) & A)) <= s*)"
     (:condition (and (projectible A) (s* < s)))
   "(A at time1)"
     (:condition (time1 <= time))
     (:clue? t)
   :backwards-premises "(A at time)"
   :variables p time R A time0 time1 s s*
   :defeasible? t
   :description "When perception is unreliable, it is not reasonable to accept its representations.")
```

These rules are illustrated by the following example:

```
========================================================================
Problem number 9:  This illustrates the use of discounted-perception and perceptual-unreliability.

Forwards-substantive-reasons:
        PERCEPTION
        DISCOUNTED-PERCEPTION

Backwards-substantive-reasons:
        PERCEPTUAL-RELIABILITY
        PERCEPTUAL-UNRELIABILITY
        TEMPORAL-PROJECTION
        NEG-AT-INTRO

Inputs:
        (the color of Fred is red) : at cycle 10 with justification 1.0

Given:
        ((the probability of (the color of Fred is red) given ((I have a percept with content (the color of Fred is
```

```
red)) &
            my surroundings are illuminated by red light)) <= 0.7)  :  with justification = 1.0
        ((the probability of (the color of Fred is red) given ((I have a percept with content (the color of Fred is
red)) &
            I am wearing red tinted glasses)) <= 0.8)  :  with justification = 1.0
        (I am wearing red tinted glasses at 1)  :  at cycle 15 with justification = 1.0
        (my surroundings are illuminated by red light at 1)  :  at cycle 30 with justification = 1.0
        (~my surroundings are illuminated by red light at 8)  :  at cycle 50 with justification = 1.0

Ultimate epistemic interests:
    ((the color of Fred is red) at 10)    degree of interest = 0.5
===========================================================================
THE FOLLOWING IS THE REASONING INVOLVED IN THE SOLUTION
Nodes marked DEFEATED have that status at the end of the reasoning.

 # 1
 ((the probability of (the color of Fred is red) given ((I have a percept with content (the color of Fred is red))
& my surroundings are illuminated by red light)) <= 0.7)
 given
 # 2
 ((the probability of (the color of Fred is red) given ((I have a percept with content (the color of Fred is red))
& I am wearing red tinted glasses)) <= 0.8)
 given
                                    # 1
                                    interest: ((the color of Fred is red) at 10)
                                    This is of ultimate interest
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
It appears to me that ((the color of Fred is red) at 10)
|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 # 3
 It appears to me that ((the color of Fred is red) at 10)
 # 4
 ((the color of Fred is red) at 10)
 Inferred by:
            support-link #3 from { 3 } by PERCEPTION  defeaters: { 7 }   DEFEATED
 This node is inferred by discharging interests (1 1)
                                    # 2
                                    interest: (((it appears to me that (the color of Fred is red)) at 10) ⊗ ((the color of
                                      Fred is red) at 10))
                                    Of interest as a defeater for support-link 3 for node 4
            ==========================================
            Justified belief in ((the color of Fred is red) at 10)
            with undefeated-degree-of-support 0.98
            answers #<Query 1: ((the color of Fred is red) at 10)>
            ==========================================
 # 5
 (I am wearing red tinted glasses at 1)
 given
                                    # 5
                                    interest: (I am wearing red tinted glasses at 10)
                                    For interest 1 by DISCOUNTED-PERCEPTION
                                    For interest 2 by PERCEPTUAL-RELIABILITY
                                    This interest is discharged by node 6
 # 6
 (I am wearing red tinted glasses at 10)
 Inferred by:
            support-link #5 from { 5 } by TEMPORAL-PROJECTION
```

This discharges interest 5
4
((the color of Fred is red) at 10)
Inferred by:
support-link #6 from { 2 , 3 , 6 } by DISCOUNTED-PERCEPTION defeaters: { 10 }
support-link #3 from { 3 } by PERCEPTION defeaters: { 7 } DEFEATED
This node is inferred by discharging interests (1 1)
8
interest: ((((the probability of (the color of Fred is red) given ((I have a percept with content (the color of Fred is red)) & I am wearing red tinted glasses)) <= 0.8) & (((it appears to me that (the color of Fred is red)) at 10) & (I am wearing red tinted glasses at 10))) ⊗ ((the color of Fred is red) at 10))
Of interest as a defeater for support-link 6 for node 4
7
(((it appears to me that (the color of Fred is red)) at 10) ⊗ ((the color of Fred is red) at 10))
Inferred by:
support-link #7 from { 2 , 6 } by PERCEPTUAL-RELIABILITY
defeatees: { link 3 for node 4 }
This node is inferred by discharging interests (2 2)
vvvvvvvvvvvvvvvvvvvvvvvvvvv
The undefeated-degree-of-support of #<Node 4> has decreased to 0.6
vvvvvvvvvvvvvvvvvvvvvvvvvvv
8
(my surroundings are illuminated by red light at 1)
given
12
interest: (my surroundings are illuminated by red light at 10)
For interest 1 by DISCOUNTED-PERCEPTION
For interest 2 by PERCEPTUAL-RELIABILITY
For interest 8 by PERCEPTUAL-UNRELIABILITY
This interest is discharged by node 9
9
(my surroundings are illuminated by red light at 10) DEFEATED
Inferred by:
support-link #9 from { 8 } by TEMPORAL-PROJECTION defeaters: { 13 } DEFEATED
This discharges interest 12
14
interest: ~(my surroundings are illuminated by red light at 10)
Of interest as a defeater for support-link 9 for node 9
4
((the color of Fred is red) at 10)
Inferred by:
support-link #10 from { 1 , 3 , 9 } by DISCOUNTED-PERCEPTION DEFEATED
support-link #6 from { 2 , 3 , 6 } by DISCOUNTED-PERCEPTION defeaters: { 10 }
support-link #3 from { 3 } by PERCEPTION defeaters: { 7 } DEFEATED
This node is inferred by discharging interests (1 1)
7
(((it appears to me that (the color of Fred is red)) at 10) ⊗ ((the color of Fred is red) at 10))
Inferred by:
support-link #11 from { 1 , 9 } by PERCEPTUAL-RELIABILITY DEFEATED
support-link #7 from { 2 , 6 } by PERCEPTUAL-RELIABILITY
defeatees: { link 3 for node 4 }
This node is inferred by discharging interests (2 2)
10
((((the probability of (the color of Fred is red) given ((I have a percept with content (the color of Fred is red)) & I am wearing red tinted glasses)) <= 0.8) & (((it appears to me that (the color of Fred is red)) at 10) & (I am wearing red tinted glasses at 10))) ⊗ ((the color of Fred is red) at 10)) DEFEATED

```
 Inferred by:
          support-link #12 from { 1 , 9 } by PERCEPTUAL-UNRELIABILITY   DEFEATED
 defeatees: { link 6 for node 4 }
 This node is inferred by discharging interest #8
          vvvvvvvvvvvvvvvvvvvvvvvvvvv
          The undefeated-degree-of-support of #<Node 4> has decreased to 0.4
          vvvvvvvvvvvvvvvvvvvvvvvvvvv
          ========================================
          Lowering the undefeated-degree-of-support of ((the color of Fred is red) at 10)
          retracts the previous answer to #<Query 1: ((the color of Fred is red) at 10)>
          ========================================
                    # 19
                    interest: (~my surroundings are illuminated by red light at 10)
                    For interest 14 by NEG-AT-INTRO
                    This interest is discharged by node 12
 # 11
 (~my surroundings are illuminated by red light at 8)
 given
 # 12
 (~my surroundings are illuminated by red light at 10)
 Inferred by:
          support-link #14 from { 11 } by TEMPORAL-PROJECTION  defeaters: { 14 }
 This discharges interest 19
                    # 21
                    interest: ~(~my surroundings are illuminated by red light at 10)
                    Of interest as a defeater for support-link 14 for node 12
 # 13
 ~(my surroundings are illuminated by red light at 10)
 Inferred by:
          support-link #15 from { 12 } by NEG-AT-INTRO
 defeatees: { link 9 for node 9 }
 This node is inferred by discharging interest #14
          vvvvvvvvvvvvvvvvvvvvvvvvvvv
          The undefeated-degree-of-support of #<Node 4> has increased to 0.6
          vvvvvvvvvvvvvvvvvvvvvvvvvvv
          #<Node 9> has become defeated.
          vvvvvvvvvvvvvvvvvvvvvvvvvvv
          #<Node 10> has become defeated.
          vvvvvvvvvvvvvvvvvvvvvvvvvvv
          ========================================
          Justified belief in ((the color of Fred is red) at 10)
          with undefeated-degree-of-support 0.6
          answers #<Query 1: ((the color of Fred is red) at 10)>
          ========================================
================== ULTIMATE EPISTEMIC INTERESTS ===================
 Interest in ((the color of Fred is red) at 10)
 is answered affirmatively by node 4
---------------------------------------------------
```

7. Extending Temporal Projection

Sometimes we want to reason about something being true throughout an interval rather than at an instant. For example, given that Fred is red at 10, it is reasonable to conclude that for each time t between 20 and 30, Fred is red at t, and hence that Fred is red throughout the interval [20,30]. This conclusion can be expressed using quantifiers over time as:

$(\forall t)[(20 \leq t \leq 30) \supset ((\text{Fred is red}) \text{ at } t)]$.

Furthermore, if we ignore considerations of reason-strength, OSCAR can perform this reasoning using the existing principle of TEMPORAL-PROJECTION:

```
========================================================================
Forwards-substantive-reasons:

Backwards-substantive-reasons:
      TEMPORAL-PROJECTION+
      ARITHMETICAL-INEQUALITY
      INEQUALITY-TRANSITIVITY

Inputs:

Given:
      ((Fred is red) at 10)  :  with justification = 1.0

Ultimate epistemic interests:
   (∀time)(((20 ≤ time) & (time ≤ 30)) → ((Fred is red) at time))    degree of interest = 0.75
========================================================================
 # 1
 ((Fred is red) at 10)
 given
                              # 1
                              interest: (∀time)(((20 ≤ time) & (time ≤ 30)) → ((Fred is red) at time))
                              This is of ultimate interest
                              # 2
                              interest: (((20 ≤ x0) & (x0 ≤ 30)) → ((Fred is red) at x0))
                              For interest 1 by UG
                              This interest is discharged by node 8
 # 2
 ((20 ≤ x0) & (x0 ≤ 30))    supposition: { ((20 ≤ x0) & (x0 ≤ 30)) }
 supposition
 generated by interest 2
                              # 3
                              interest: ((Fred is red) at x0)    supposition: { ((20 ≤ x0) & (x0 ≤ 30)) }
                              For interest 2 by conditionalization
                              This interest is discharged by node 7
 # 3
 (20 ≤ x0)    supposition: { ((20 ≤ x0) & (x0 ≤ 30)) }
 Inferred by:
          support-link #2 from { 2 } by simp
                              # 4
                              interest: (10 ≤ x0)    supposition: { ((20 ≤ x0) & (x0 ≤ 30)) }
                              For interest 3 by TEMPORAL-PROJECTION+
                              This interest is discharged by node 6
                              # 5
                              interest: (10 ≤ 20)    supposition: { ((20 ≤ x0) & (x0 ≤ 30)) }
                              For interest 4 by INEQUALITY-TRANSITIVITY
                              This interest is discharged by node 5
 # 5
 (10 ≤ 20)
 Inferred by:
          support-link #4 from {  } by ARITHMETICAL-INEQUALITY
 This discharges interest 5
```

```
 # 6
 (10 ≤ x0)    supposition: { ((20 ≤ x0) & (x0 ≤ 30)) }
 Inferred by:
          support-link #5 from { 3 , 5 } by INEQUALITY-TRANSITIVITY
 This node is inferred by discharging interest #4
 # 7
 ((Fred is red) at x0)    supposition: { ((20 ≤ x0) & (x0 ≤ 30)) }
 Inferred by:
          support-link #6 from { 1 , 6 } by TEMPORAL-PROJECTION+
 This node is inferred by discharging interest #3
 # 8
 (((20 ≤ x0) & (x0 ≤ 30)) → ((Fred is red) at x0))
 Inferred by:
          support-link #7 from { 7 } by conditionalization
 This node is inferred by discharging interest #2
 # 9
 (∀time)(((20 ≤ time) & (time ≤ 30)) → ((Fred is red) at time))
 Inferred by:
          support-link #8 from { 8 } by UG
 This node is inferred by discharging interest #1
          ========================================
          Justified belief in (∀time)(((20 ≤ time) & (time ≤ 30)) → ((Fred is red) at time))
          answers #<Query 1: (∀time)(((20 ≤ time) & (time ≤ 30)) → ((Fred is red) at time))>
          ========================================
============================================================
```

However, in getting OSCAR to perform this reasoning, I have replaced TEMPORAL-PROJECTION by TEMPORAL-PROJECTION+, which is just like TEMPORAL-PROJECTION except that the reason-strength is left unspecified (and hence defaults to 1.0). OSCAR cannot perform this reasoning using TEMPORAL-PROJECTION unmodified, because the applicability of that principle requires that the times be numbers, whereas in this example OSCAR must reason about variable times. There is no way to modify TEMPORAL-PROJECTION to allow reasoning about variable times, because if the times are not specified as numbers then there is no way to compute an appropriate reason-strength.

In this example, it is clear what the strength of support should be for the conclusion. It should be the weakest support for any conclusion of the form ((Fred is red) at t) where $20 \leq t \leq 30$, and that in turn occurs when t = 30. Consequently, we can capture the appropriate form of this reasoning by adopting an analogue of TEMPORAL-PROJECTION for intervals:

```
(def-backwards-reason INTERVAL-PROJECTION
   :conclusions  "(p throughout (time* time))"
   :condition  (and (temporally-projectible p) (numberp time*) (numberp time) (<= time* time))
   :forwards-premises
   "(p at time0)"
      (:condition (and (time0 < time*) ((time* - time0) < 693)))
   :variables p time0 time* time
   :defeasible? T
   :strength  (- (* 2 (expt *temporal-decay* (- time time0))) 1)
   :description
   "It is defeasibly reasonable to expect temporally projectible truths to remain unchanged.")
```

OSCAR can then reason trivially as follows:

```
========================================================================
 # 1
 ((Fred is red) at 10)
```

```
given
                        # 1
                        interest: ((Fred is red) throughout (20 30))
                        This is of ultimate interest
# 2
((Fred is red) throughout (20 30))
Inferred by:
        support-link #2 from { 1 } by INTERVAL-PROJECTION
This discharges interest 1
==========================================================================
```

We can simplify things still further by noting that TEMPORAL-PROJECTION can be regarded as a special case of INTERVAL-PROJECTION. Using an idea of Shoham [1987], we can take (P at t) to mean (P throughtout [t , t]).[11] Then we can dispense with INTERVAL-PROJECTION and redefine TEMPORAL-PROJECTION as we defined INTERVAL-PROJECTION above.

A further generalization is desirable. INTERVAL-PROJECTION projects a conclusion throughout a closed interval. There will be cases in which we want to project a conclusion throughout an open interval (an interval of the form (x , y)) or a clopen interval (of the form (x , y]) rather than a closed interval. As TEMPORAL-PROJECTION is an backwards-reason, we can have it license inferences to any of these conclusions. To accomplish this, let us symbolize open intervals as (open x y), closed intervals as (closed x y), and clopen intervals as (clopen x y). These will be printed as "(x , y)", "[x , y]", and "(x , y]", respectively. When I want to refer to an interval without specifying whether it is open, clopen, or closed, I will write it in the form "⟨x , y⟩". Then we can redefine TEMPORAL-PROJECTION as follows:

> TEMPORAL-PROJECTION
> If P is temporally-projectible, $t < t^* \leq t^{**}$ and $(t^{**} - t) < \log(.5)/\log(\rho)$ then believing P-at-t is a defeasible reason of strength $2\cdot(\rho^{(t^{**} - t)} - .5)$ for the agent to believe P-throughout-⟨t* , t**⟩.

and impliment it as follows:

```
(def-backwards-reason *TEMPORAL-PROJECTION*
   :conclusions  "(p throughout (op time* time))"
   :condition  (and (temporally-projectible p) (numberp time*) (numberp time) (<= time* time)
                    (or (eq op 'open) (eq op 'closed) (eq op 'clopen)))
   :forwards-premises
      "(p at time0)"
      (:condition (and (time0 < time) ((time - time0) < 693)))
   :variables p time0 time* time op
   :defeasible? T
   :strength  (- (* 2 (expt *temporal-decay* (- time time0))) 1)
   :description
   "It is defeasibly reasonable to expect temporally projectible truths to remain unchanged.")
```

8. Temporal Indexicals

An agent that did all of its temporal reasoning using the reason-schemas described above would be led into crippling computational complexities. Every time the agent

---

[11] Implementation node: this is accomplished very simply in OSCAR by translating the pretty-formula "(P at t)" into the formula (throughout P t t).

wanted to reuse a belief about its surroundings, it would have to reinfer it for the present time. Inference takes time, so by the time it had reinferred the belief, other beliefs with which the agent might want to combine this belief in further inference would themselves no longer be current. To get around this difficulty, the agent would have to make inferences about some time in the near future rather than the present, inferring a number of properties of that time, and then combine those properties to make a further inference about that time, and finally project that new property into the future for use in further inference. This would not be computationally impossible, but it would make life difficult for an agent that had to reason in this way.

Human beings achieve the same result in a more efficient way by employing the temporal indexical "now". Rather than repeatedly reinferring a property as time advances, they infer it once as holding *now*, and that single belief is retained until it becomes defeated. The mental representation (i.e., formula) believed remains unchanged as time advances, but the content of the belief changes continuously in the sense that at each instant it is a belief about *that instant*. This has the effect of continuously updating the agent's beliefs in accordance with TEMPORAL-PROJECTION, but no actual reasoning need occur.

The use of "now" can be implicit or explicit. That is, we can either write "x is red" or "x is red now". The latter is used primarily for emphasis. The representation is simpler if we drop the temporal reference rather than putting in the "now", so that is the course that will be followed below.

Percepts are always percepts of the agent's present situation, so we can regard them as providing defeasible reasons for inferences about the present:

INDEXICAL-PERCEPTION
Having a percept at time t with the content P is a defeasible reason for the agent to believe P.

INDEXICAL-PERCEPTION is defeated by considerations of reliability just as PERCEPTION is:

INDEXICAL-PERCEPTUAL-RELIABILITY
Where R is projectible, r is the strength of INDEXICAL-PERCEPTION, "R-at-t, and the probability is less than $0.5\cdot(r + 1)$ of P's being true given R and that I have a percept with content P at t" is a defeasible undercutting defeater for INDEXICAL-PERCEPTION.

The conclusion P is automatically projected into the future just by retaining it, so INDEXICAL-PERCEPTION can be viewed as combining PERCEPTION and TEMPORAL-PROJECTION into a single reason-scheme. The projectibility constraint on temporal projection has not been included here, on the assumption that only temporally projectible propositions can be the contents of percepts.

Because INDEXICAL-PERCEPTION builds in an application of TEMPORAL-PROJECTION, it must be defeated by the same considerations that defeat TEMPORAL-PROJECTION:

PROBABILISTIC-DEFEAT-FOR-INDEXICAL-PERCEPTION
"The probability of P-at-(t+1) given P-at-t $\neq \rho$" is a conclusive undercutting defeater for INDEXICAL-PERCEPTION.

A complication arises for the reasoner in dealing with conclusions containing an implicit or explicit "now" representing an implicit use of TEMPORAL-PROJECTION. The degree of support for a conclusion inferred by TEMPORAL-PROJECTION decreases as the time interval increases. Retaining a conclusion containing "now" is equivalent to making an inference by TEMPORAL-PROJECTION. Accordingly, the degree of support for that conclusion must decay over time, just as if it were being continually re-inferred by TEMPORAL-PROJECTION. To handle this, we must make a distinction between *temporal* and *atemporal* conclusions, where the former are those containing an explicit or implicit "now". Atemporal conclusions have fixed degrees-of-support, but the degree of support for a temporal

conclusion must decay proportionally to $\rho^{\Delta t}$ where Δt is the length of time since it was initially inferred. To handle this, inference-nodes are marked "temporal" or "atemporal", and their construction times are stored with them.[12]

When a new argument is constructed for a temporal conclusion, it begins to decay from the time it is constructed. This means that the new construction time must be stored as well. However, we need not retain a record of the construction time for the original argument. One of the advantages of using $\rho^{\Delta t}$ as the decay factor is that $\rho^{\Delta t + \Delta t^*} = \rho^{\Delta t} \cdot \rho^{\Delta t^*}$. Consequently, we can replace the old construction time by the new construction time, store the decayed strengths of any earlier arguments in place of their original strengths, and then let everything decay further as if it had all begun at the new construction time.

To implement this, reasons are given a new field *temporal?* that determines whether the application of the reason produces a temporal conclusion. This allows us to implement INDEXICAL-PERCEPTION as follows:

```
(def-forwards-reason INDEXICAL-PERCEPTION
   :forwards-premises "(p at time)"
       (:kind :percept)
   :conclusions "p"
   :variables p time
   :strength  (- (* 1.98 (expt *temporal-decay* (- *cycle* time))) 1)
   :defeasible? t
   :temporal? t
   :description  "When information is input, it is defeasibly reasonable to believe it.")
```

In effect, INDEXICAL-PERCEPTION combines an application of PERCEPTION and an application of TEMPORAL-PROJECTION. INDEXICAL-PERCEPTUAL-RELIABILITY defeats INDEXICAL-PERCEPTION by defeating the imbedded application of PERCEPTION. However, unlike PERCEPTION, the strength of INDEXICAL-PERCEPTION decays as the time interval increases. The ability of INDEXICAL-PERCEPTUAL-RELIABILITY to defeat an application of INDEXICAL-PERCEPTION should not increase as the time interval increases, so the strength of INDEXICAL-PERCEPTUAL-RELIABILITY must also decay:

```
(def-backwards-undercutter INDEXICAL-PERCEPTUAL-RELIABILITY
   :defeatee *indexical-perception*
   :forwards-premises
   "((the probability of p given ((I have a percept with content p) & R)) <= s)"
     (:condition (and (projectible R) (s < 0.99)))
   "(R at time0)"
      (:condition (and (time0 < time) ((time - time0) < 693)))
      (:clue? t)
   :backwards-premises "(R at time)"
   :variables p time R time0 s
   :defeasible? t
   :strength  (- (* 2 (expt *temporal-decay* (- now time))) 1)
   :temporal? t
   :description "When perception is unreliable, it is not reasonable to accept its representations.")
```

Here is an example that combines PERCEPTION, INDEXICAL-PERCEPTION, INDEXICAL-

---

[12] I assume that temporal nodes can only be supported by arguments with decaying strengths. This seems obvious when we think about examples. How could you have a non-decaying argument for "Fred is now red"? Without that assumption, we could have a strong decaying argument and a weaker argument of fixed strength for the same conclusion. The undefeated-degree-of-support of the conclusion would then decay until the stronger argument decayed to the strength of the weaker argument, and then it would cease decaying. But with this assumption, it follows that the undefeated-degree-of-support of a temporal conclusion decays uniformly proportionally to $\rho^{\Delta t}$.

PERCEPTUAL-RELIABILITY (and also an undiscussed principle about reliable testimony):

==

Problem number 8: First, Fred looks red to me. Later, I am informed by Merrill that I am then wearing blue-tinted glasses. Later still, Fred looks blue to me. All along, I know that the probability is not high of Fred being blue given that Fred looks blue to me but I am wearing blue-tinted glasses. What should I conclude about the color of Fred?

Forwards-substantive-reasons:
INDEXICAL-PERCEPTION
PERCEPTION
RELIABLE-INFORMANT

Backwards-substantive-reasons:
INDEXICAL-PERCEPTUAL-RELIABILITY
PERCEPTUAL-RELIABILITY
TEMPORAL-PROJECTION
INDEXICAL-INCOMPATIBLE-COLORS

Inputs:
(the color of Fred is red) : at cycle 1 with justification 0.8
(Merrill reports that I am wearing blue tinted glasses) : at cycle 20 with justification 1.0
(the color of Fred is blue) : at cycle 30 with justification 0.8

Given:
((the probability of (the color of Fred is blue) given ((I have a percept with content (the color of Fred is blue)) & I am wearing blue tinted glasses)) <= 0.8) : with justification = 1.0
(Merrill is a reliable informant) : with justification = 1.0

Ultimate epistemic interests:
(? x)(the color of Fred is x) degree of interest = 0.65

===

THE FOLLOWING IS THE REASONING INVOLVED IN THE SOLUTION
Nodes marked DEFEATED have that status at the end of the reasoning.

1
((the probability of (the color of Fred is blue) given ((I have a percept with content (the color of Fred is blue)) & I am wearing blue tinted glasses)) <= 0.8)
given
undefeated-degree-of-support = 1.0 at cycle 1.
2
(Merrill is a reliable informant)
given
undefeated-degree-of-support = 1.0 at cycle 1.
1
interest: (the color of Fred is y0)
This is of ultimate interest
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
It appears to me that ((the color of Fred is red) at 1)
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
3
It appears to me that ((the color of Fred is red) at 1)
5
(the color of Fred is red)
Inferred by:
support-link #4 from { 3 } by INDEXICAL-PERCEPTION defeaters: { 13 }
undefeated-degree-of-support = 0.8 at cycle 2.

```
This discharges interest 1
                        # 5
                        interest: ~(the color of Fred is red)
                        Of interest as a defeater for support-link 4 for node 5
            ========================================
            Justified belief in (the color of Fred is red)
            with undefeated-degree-of-support 0.8
            answers #<Query 1: (? x)(the color of Fred is x)>
            ========================================
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
It appears to me that ((Merrill reports that I am wearing blue tinted glasses) at 20)
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 # 6
 It appears to me that ((Merrill reports that I am wearing blue tinted glasses) at 20)
 # 7
 ((Merrill reports that I am wearing blue tinted glasses) at 20)
 Inferred by:
          support-link #5 from { 6 } by PERCEPTION
 undefeated-degree-of-support = 0.99 at cycle 20.
 # 9
 (I am wearing blue tinted glasses at 20)
 Inferred by:
          support-link #7 from { 2 , 7 } by RELIABLE-INFORMANT
 undefeated-degree-of-support = 0.98 at cycle 22.
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
It appears to me that ((the color of Fred is blue) at 30)
||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 # 10
 It appears to me that ((the color of Fred is blue) at 30)
                        # 13
                        interest: (I am wearing blue tinted glasses at 30)
                        For interest 15 by INDEXICAL-PERCEPTUAL-RELIABILITY
                        This interest is discharged by node 15
 # 12
 (the color of Fred is blue)            DEFEATED
 Inferred by:
          support-link #9 from { 10 } by INDEXICAL-PERCEPTION  defeaters: { 16 , 14 }   DEFEATED
 undefeated-degree-of-support = 0.8 at cycle 30.
 This discharges interest 1
                        # 15
                        interest: (((it appears to me that (the color of Fred is blue)) at 30) ⊗ (the color of
                          Fred is blue))
                        Of interest as a defeater for support-link 9 for node 12
                        # 16
                        interest: ~(the color of Fred is blue)
                        Of interest as a defeater for support-link 9 for node 12
            ========================================
            Justified belief in (the color of Fred is blue)
            with undefeated-degree-of-support 0.8
            answers #<Query 1: (? x)(the color of Fred is x)>
            ========================================
 # 13
 ~(the color of Fred is red)            DEFEATED
 Inferred by:
          support-link #10 from { 12 } by INDEXICAL-INCOMPATIBLE-COLORS   DEFEATED
 undefeated-degree-of-support = 0.7982 at cycle 31.
 defeatees: { link 4 for node 5 }
```

```
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
        #<Node 5> has become defeated.
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
        ========================================
        Lowering the undefeated-degree-of-support of (the color of Fred is red)
        retracts the previous answer to #<Query 1: (? x)(the color of Fred is x)>
        ========================================
# 14
~(the color of Fred is blue)
Inferred by:
         support-link #11 from { 5 } by inversion from contradictory nodes 13 and 5
defeatees: { link 9 for node 12 }
undefeated-degree-of-support = 0.0 at cycle 31.
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
        #<Node 14> has become defeated.
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
# 15
(I am wearing blue tinted glasses at 30)
Inferred by:
         support-link #12 from { 9 } by TEMPORAL-PROJECTION
undefeated-degree-of-support = 0.98 at cycle 34.
This discharges interest 13
# 16
(((it appears to me that (the color of Fred is blue)) at 30) ⊗ (the color of Fred is blue))
Inferred by:
         support-link #13 from { 1 , 15 } by INDEXICAL-PERCEPTUAL-RELIABILITY
defeatees: { link 9 for node 12 }
undefeated-degree-of-support = 0.98 at cycle 34.
This node is inferred by discharging interest #15
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
        The undefeated-degree-of-support of #<Node 14> has increased to 0.773
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
        The undefeated-degree-of-support of #<Node 5> has increased to 0.773
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
        #<Node 12> has become defeated.
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
        #<Node 13> has become defeated.
        vvvvvvvvvvvvvvvvvvvvvvvvvvv
        ========================================
        Justified belief in (the color of Fred is red)
        with undefeated-degree-of-support 0.773
        answers #<Query 1: (? x)(the color of Fred is x)>
        ========================================
        ========================================
        Lowering the undefeated-degree-of-support of (the color of Fred is blue)
        retracts the previous answer to #<Query 1: (? x)(the color of Fred is x)>
        ========================================
================== ULTIMATE EPISTEMIC INTERESTS ==================
 Interest in (? x)(the color of Fred is x)
 is answered by node 5:  (the color of Fred is red)
--------------------------------------------------
```

It may be illuminating to draw the inference-graph for this latter problem, as in figure 6.

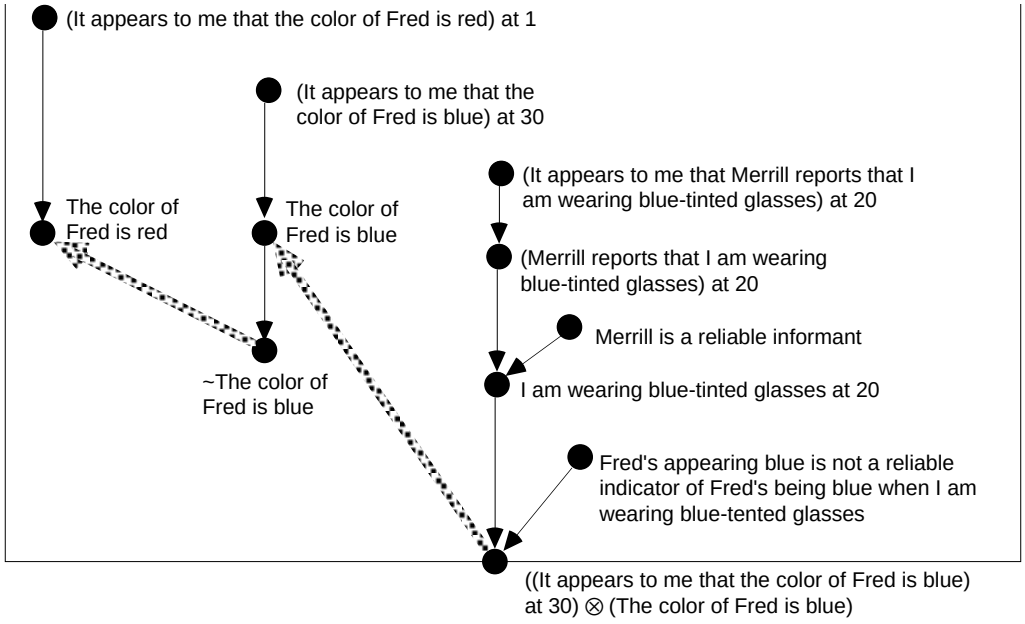


Figure 6. Inference graph

The use of temporal indexicals calls for indexical versions of TEMPORAL-PROJECTION, DISCOUNTED-PERCEPTION, and PERCEPTUAL-UNRELIABILITY:

INDEXICAL-TEMPORAL-PROJECTION

If P is temporally-projectible and t < now and (now–t) < log(.5)/log(ρ) then believing P-at-t is a defeasible reason of strength $2\cdot(\rho^{(now - t)} - .5)$ for the agent to believe P.

DISCOUNTED-INDEXICAL-PERCEPTION

Where R is projectible, r is the strength of INDEXICAL-PERCEPTION , $0.5 < s < 0.5\cdot(r + 1)$ and (now–t) < log(.5)/log(ρ), having a percept at time t with the content P and the belief "R-at-t, and the probability is less than s of P's being true given R and that I have a percept with content P" is a defeasible reason of strength $2\cdot(s - 0.5)\cdot\rho^{(now - t)}$ for the agent to believe P-at-t.

INDEXICAL-PERCEPTUAL-UNRELIABILITY

Where A is projectible, (now–t) < log(.5)/log(ρ), and s* < s, "A-at-t, and the probability is less than or equal to s* of P's being true given A and that I have a percept with content P" is a defeater of strength $2\cdot(\rho^{(now - t)} - .5)$ for DISCOUNTED-INDEXICAL-PERCEPTION.

INDEXICAL-TEMPORAL-PROJECTION must be defeated by the same considerations that defeat TEMPORAL-PROJECTION:

PROBABILISTIC-DEFEAT-FOR-INDEXICAL-TEMPORAL-PROJECTION

"The probability of P-at-(t+1) given P-at-t ≠ ρ" is a conclusive undercutting defeater for INDEXICAL-TEMPORAL-PROJECTION.

These are implimented as follows:

These are implimented as follows:

```
(def-backwards-reason INDEXICAL-TEMPORAL-PROJECTION
   :conclusions  "p"
   :forwards-premises
   "(p at time0)"
   (:condition (and (time0 < now) ((now - time0) < 693)))
   :condition  (and (temporally-projectible p) (not (occur 'at p)))
   :variables p time0
   :defeasible? T
   :temporal? T
   :strength  (- (* 2 (expt *temporal-decay* (- now time0))) 1)
   :description
   "It is defeasibly reasonable to expect temporally projectible truths to remain unchanged.")

(def-backwards-undercutter
      PROBABILISTIC-DEFEAT-FOR-INDEXICAL-TEMPORAL-PROJECTION
  :defeatee  INDEXICAL-TEMPORAL-PROJECTION
   :forwards-premises
      "((the probability of (p at (t + 1)) given (p at t)) = s)"
       (:condition (not (s = *temporal-decay*)))
   :variables  p s time0 time)

(def-forwards-reason DISCOUNTED-INDEXICAL-PERCEPTION
   :forwards-premises
   "((the probability of p given ((I have a percept with content p) & R)) <= s)"
      (:condition (and (projectible R) (0.5 < s) (s < 0.995)))
   "(p at time)"
     (:kind :percept)
   "(R at time0)"
      (:condition (and (time0 < now) ((now - time0) < 693)))
      (:clue? t)
   :backwards-premises "(R at time)"
   :conclusions "p"
   :variables p R time0 time s
   :strength (* 2 (s - 0.5) (expt *temporal-decay* (- *cycle* time)))
   :defeasible? t
   :temporal? t
   :description "When information is input, it is defeasibly reasonable to believe it.")

(def-backwards-undercutter INDEXICAL-PERCEPTUAL-UNRELIABILITY
   :defeatee DISCOUNTED-INDEXICAL-PERCEPTION
   :forwards-premises
   "((the probability of p given ((I have a percept with content p) & A)) <= s*)"
      (:condition (and (projectible A) (s* < s)))
   "(A at time1)"
      (:condition (and (time1 <= now) ((now - time1) < 693)))
      (:clue? t)
   :backwards-premises "(A at time)"
   :variables p time R A time0 time1 s s*
   :defeasible? t
   :temporal? t
   :strength  (- (* 2 (expt *temporal-decay* (- now time))) 1)
   :description "When perception is unreliable, it is not reasonable to accept its representations.")
```

We can move back and forth between indexical and non-indexical representations of time. "now" is short for "at the present time". "now" is an adverb, and "the present time" functions as a pronoun. Thus we have the following reason-schemas:

(7) If P does not contain temporal reference then "the present time = t and P" is a conclusive reason for P-at-t.

(8) If P does not contain temporal reference then "the present time = t and P-at-t" is a conclusive reason for P.

To make use of these reason-schemas, the agent must have some way of knowing what the present time is. In human beings, this is a complex matter, but in artificial agents implemented on a computer, we can use the computer's clock for the time reference and give the agent the ability to form a true belief of the form "the present time = t" whenever it adopts interest in what the present time is. Even more simply, the present implementation uses the reasoning-cycle as the measure of time.

9. Implementing Temporal Conclusions

To implement temporal conclusions with decaying strengths, we must store not only the strength of an inference-node but also the point from which that strength begins to decay, and then whenever we reuse the strength we must update it. I have talked generically about the "strength" of an inference-node, but in OSCAR there are three different strengths that must decay for temporal conclusions. These are the maximal-degree-of-support, the undefeated-degree-of-support, and the discounted-node-strength. The maximal-degree-of-support is the maximal strength of all arguments supporting the node, and is used in computing defeat-statuses. The undefeated-degree-of-support is the maximal strength of the undefeated arguments supporting the node. The discounted-node-strength is a constant times the undefeated-degree-of-support, and is only used for prioritizing the reasoning. All three decay uniformly from the time the node is constructed.

To accommodate this, we add a slot to inference-nodes for *temporal-node* and a slot to support-links for *temporal-link*. For those that are temporal, this stores the cycle at which they begin decaying. If BUILD-SUPPORT-LINK adds a new temporal support-link whose link-strength is greater than the decayed maximal-degree-of-support for the link-target, it must recompute *maximum-degree-of-support*, *discounted-node-strength*, *old-undefeated-degree-of-support*, and *temporal-node* if the target-node is temporal. This is because a node has just one start-time (recorded in *temporal-node*), so we cannot update just one support-link at a time. These changes must then be recursed through the node-consequences by ADJUST-SUPPORT-FOR-CONSEQUENCES.

When UPDATE-BELIEFS is called, *maximal-degree-of-support*, *discounted-node-strength*, *old-undefeated-degree-of-support*, and *temporal-node* must be reset for all affected-nodes that are temporal-nodes, and COMPUTE-LINK-STRENGTH must be run on all their consequent-links. *Undefeated-degree-of-support* will then be recomputed normally.

The main complication for all this is that link-strengths are used in computing affected-nodes and links. We have to be sure that updated values are used. This can be resolved by updating-strengths for all inference/defeat-descendants of the new link.

It will be assumed that descendants of temporal-nodes and links are temporal. Somewhat surprisingly, this assumption can conflict with the weakest link principle. Suppose we have a reason allowing us to infer R from the two premises P and Q, where P is temporal and Q is not. Then R will be temporal, and so will begin decaying from the time it is inferred. However, suppose that at the time of the inference, the undefeated-degree-of-support of Q is less than that of P. By the weakest link principle, the undefeated-degree-of-support of R should be that of Q, and should remain constant at that value until the undefeated-degree-of-support of P decays to the level of the undefeated-degree-of-support of Q, at which time the undefeated-degree-of-support of R will begin decaying. It follows that the undefeated-degree-of-support of R should be constant for a while and then begin decaying. Combining this observation with the fact that there may be multiple arguments for a single conclusion, we can construct nodes whose decay curves have multiple flat sections followed by decay. I have been unable to implement the computation of undefeated-

degrees-of-support for such nodes in an efficient way. Introspection suggests that human beings do not compute undefeated-degrees-of-support in this way either. If a conclusion contains an implicit "now", we take its undefeated-degree-of-support to decay uniformly, without taking account of its ancestry. Accordingly, I have followed this same course in dealing with temporal nodes, but it must be recognized that this is, in a sense, a divergence from ideal rationality.

10. Reasoning about Change

Reasoning about what will change if an action is performed or some other change occurs often presupposes knowing what will not change. Early attempts to model such reasoning deductively proceeded by adopting a large number of "frame axioms", which were axioms to the effect that if something occurs then something else will not change. For instance, in a blocks world one of the frame axioms might be "If a block is moved, its color will not change". It soon became apparent that complicated situations required more frame axioms than axioms about change, and most of the system resources were being occupied by proofs that various properties did not change. In a realistically complicated situation, this became unmanageable. What became known as the *Frame Problem* is the problem of reorganizing reasoning about change so that reasoning about non-change can be done efficiently (McCarthy and Hayes [1969]; Janlert, [1987]).

AI hackers, as Hayes [1987] calls them, avoided this problem by adopting the "sleeping dog strategy" (Haugeland [1987]). Starting with STRIPS, actual planning systems maintained databases of what was true in a situation, and with each possible action they stored lists of what changes those actions would produce. For planning systems intended to operate only in narrowly circumscribed situations, this approach is effective, although for general-purpose planning it quickly becomes unwieldy. In the attempt to provide a more general theory that justifies this approach as a special case, several authors (Sandewall [1970], McDermott [1982], McCarthy [1986]) proposed reasoning about change defeasibly and adopting some sort of defeasible inference scheme to the effect that it is reasonable to believe that something doesn't change unless you are forced to conclude otherwise. But to make the idea work, one needs both a precise framework for defeasible reasoning and a precise formulation of the requisite defeasible inference schemes. That proved to be a difficult problem.

The temporal projection principles defended in sections five and seven can be regarded as a precise formulation of the defeasible inference schemes sought. Unfortunately, these principles do not solve the Frame Problem. Steve Hanks and Drew McDermott [1986] were the first to observe that even with defeasible principles of non-change, a reasoner will often be unable to determine what changes and what does not. They illustrated this with what has become known as "the Yale shooting problem". The general form of the problem is this. Suppose we have a causal law to the effect that if P is true at a time t and action A is performed at that time, then Q will be true shortly thereafter. (More generally, A could be anything that becomes true at a certain time. What is significant about actions is that they are changes.) Suppose we know that P is true now, and Q false. What should we conclude about the results of performing action A in the immediate future? Hanks and McDermott illustrate this by taking P to be "The gun is loaded, in working condition, and pointed at Jones", Q to be "Jones is dead", and A to be the action of pulling the trigger. We suppose (simplistically) that there is a causal law dictating that if the trigger is pulled on a loaded gun that is in working condition and pointed at someone, that person will shortly be dead. Under these circumstances, it seems clear that we should conclude that Jones will be dead shortly after the trigger is pulled.

The difficulty is that all we can infer from what we are given is that when A is performed either P will no longer be true or Q will be true shortly thereafter. Intuitively, we want to conclude (at least defeasibly) that P will remain true at the time A is performed and Q will therefore become true shortly thereafter. But none of our current machinery enables us to distinguish between P and Q. Because P is now true and Q is now false, we have a defeasible reason for believing that P will still be true when A is performed, and

we have a defeasible reason for believing that Q will still be false shortly thereafter. We know that one of these defeasible conclusions will be false, but we have no basis for choosing between them, so this becomes a case of collective defeat. That, however, is the intuitively wrong answer.

When we reason about causal mechanisms, we think of the world as "unfolding" temporally, and changes only occur when they are forced to occur by what has already happened. In our example, when A is performed, nothing has yet happened to force a change in P, so we conclude defeasibly that P remains true. But given the truth of P, we can then deduce that at a slightly later time, Q will become true. Thus when causal mechanisms force there to be a change, we conclude defeasibly that the change occurs in the later states rather than the earlier states. This seems to be part of what we mean by describing something as a causal mechanism. Causal mechanisms are systems that force changes, where "force" is to be understood in terms of temporal unfolding.[13]

When reasoning about such a causal system, part of the force of describing it as causal must be that the defeasible presumption against the effect occurring is somehow removed. Thus, although we normally expect Jones to remain alive, we do not expect this any longer when he is shot. To remove a defeasible presumption is to defeat it. This suggests that there is some kind of general "causal" defeater for the temporal projection principles adumbrated above. The problem is to state this defeater precisely. As a first approximation we might try:

(9) For every $\varepsilon \geq 0$ and $\delta > 0$, "A&P-at-$(t+\varepsilon)$ & (A&P causes Q)" is an undercutting defeater for the defeasible inference from ~Q-at-t to ~Q-at-$(t+\varepsilon+\delta)$ by TEMPORAL-PROJECTION.

The temporal-unfolding view of causal reasoning requires causation to be temporally asymmetric. That is, "A&P causes Q" means, in part, that if A&P becomes true then Q will *shortly* become true. This precludes *simultaneous* causation, in which Q is caused to be true at t by A&P being true at t. This may seem problematic, on the grounds that simultaneous causation occurs throughout the real world. For instance, colliding billiard balls in classical physics might seem to illustrate simultaneous causation. However, this is a mistake. If two billiard balls collide at time t with velocity vectors pointing towards each other, they do not also have velocity vectors pointing away from each other at the very same time. Instead, this illustrates what I have elsewhere [1984] called *instantaneous* causation. Instantaneous causation requires that if A&P becomes true at t, then for some $\delta > 0$, Q will be true throughout the clopen interval $(t , t+\delta]$.[14] I believe that instantaneous causation is all that is required for describing the real world.

I have followed AI-convention here in talking about causal change in terms of causation. However, that introduces unnecessary complexities. For example, it is generally assumed in the philosophical literature on causation that if P causes Q then Q would not have been true if P were not true.[15] This has the consequence that when there are two independent factors each of which would be sufficient by itself to cause the same effect, if both occur then neither causes it. These are cases of causal overdetermination. A familiar example of causal overdetermination occurs when two assailants shoot a common victim at the same time. Either shot would be fatal. The result is that neither shot is such that if it had not occurred then the victim would not have died, and hence, it is generally maintained,

---

[13] This intuition is reminiscent of Shoham's [1987] "logic of chronological ignorance", although unlike Shoham, I propose to capture the intuition without modifying the structure of the system of defeasible reasoning. This is also related to the proposal of Gelfond and Lifschitz [1993]. This same idea underlies my analysis of counterfactual conditionals in Pollock [1979] and [1984].

[14] I assume that time has the structure of the reals, although that assumption is not required for the implementation.

[15] See Lewis [1973]. See my [1984] for more details about the relationship between causes and counterfactual conditionals.

neither shot caused the death of the victim. However, this kind of failure of causation ought to be irrelevant to the kind of causal reasoning under discussion in connection with change. Principle (9) ought to apply to cases of causal overdetermination as well as to genuine cases of causation. This indicates that the intricacies of the analysis of "cause" are irrelevant in the present context.

I take it (and have argued in my [1984]) that all varieties of causation (including causal overdetermination) arise from the instantiation of "causal laws". These are what I have dubbed *nomic generalizations*, and have discussed at length in my [1990]. Nomic generalizations are symbolized as "P ≡> Q", where P and Q are formulas and '≡>' is a variable-binding operator, binding all free occurrences of variables in P and Q. An informal gloss on "P ≡> Q" is "Any physically-possible P would be a Q". For example, the law that electrons are negatively charged could be written "(x is an electron) ≡> (x is negatively charged)". The free occurrences of 'x' are bound by '≡>'.

A rule of universal instantiation applies to nomic generalizations, allowing us to derive less general nomic generalizations:

> If 'x' is free in P and Q, and P(x/a) and Q(x/a) result from substituting the constant term 'a' for 'x', then (P ≡> Q) entails (P(x/a) ≡> Q(x/a)).

I propose that we replace "(A&P causes Q)" in (9) by "(A&P ≡> Q will shortly be true)", where the latter typically results from instantiating more general laws. More precisely, let us define "A when P is causally sufficient for Q after an interval ε" to mean

(∀t){(A-at-t & P-at-t) ≡> (∃δ)Q-throughout-(t+ε , t+ε+δ]}.

Instantaneous causation is causal sufficiency with an interval 0.

My proposal is to replace "causes" by "causal sufficiency" in (9). Modifying it to take account of the interval over which the causation occurs:

> *CAUSAL-UNDERCUTTER*
> Where $t_0 \leq t_1$ and $(t_1+\varepsilon) < t$, "A-at-t_1 & Q-at-t_1 & (A when Q is causally sufficient for ~P after an interval ε)" is a defeasible undercutting defeater for the inference from P-at-t_0 to P-throughout-⟨t* t⟩ by TEMPORAL-PROJECTION.

This can be implemented as follows:

```
(def-backwards-undercutter CAUSAL-UNDERCUTTER
   :defeatee *temporal-projection*
   :forwards-premises
      "(A when Q is causally sufficient for ~p after an interval interval)"
      "(A at time1)"
        (:condition (and (time0 <= time1) ((time1 + interval) < time)))
   :backwards-premises
      "(Q at time1)"
   :variables  A Q p time0 time time* time1 interval op
   :defeasible?  T)
```

We can also construct an indexical version of this principle is as follows:

```
(def-backwards-undercutter INDEXICAL-CAUSAL-UNDERCUTTER
   :defeatee INDEXICAL-TEMPORAL-PROJECTION
   :forwards-premises
      "(A when Q is causally sufficient for ~p after an interval interval)"
      "(A at time1)"
        (:condition (and (time0 <= time1) ((time1 + interval) < now) ((now - (time1 + interval) < 693))))
   :backwards-premises
```

```
    "(Q at time1)"
  :variables  A Q p time0 time1 interval
  :defeasible?  T
  :temporal? T)
```

For causal reasoning, we want to use the causal connection to support inferences about what will happen. This is more complicated than it might initially seem. The difficulty is that, for example,

> the gun is fired when the gun is loaded is causally sufficient for ~(Jones is alive) after an interval 20

does not imply that if the gun is fired at t and the gun is loaded at t then Jones is dead at t+20. Recall the discussion of instantaneous causation. All that is implied is that Jones is dead over some interval open on the left and with t+20 as the lower bound. We can conclude that *there is* at time > t+20 at which Jones is dead, but it does not follow as a matter of logic that Jones is dead at any particular time because, at least as far as this causal law is concerned, Jones could become alive again after becoming dead. To infer that Jones is dead at a particular time after t+20, we must combine the causal sufficiency with temporal projection. This yields the following principle:

CAUSAL-IMPLICATION

- If Q is temporally projectible, $(t^{**}-(t+\varepsilon)) < \log(.5)/\log(\rho)$, and $((t+\varepsilon) \leq t^* < t^{**})$, then "(A when P is causally sufficient for Q after an interval ε) & A-at-t & P-at-t" is a defeasible reason for "Q-throughout-(t* , t**]" and for "Q-throughout-(t* , t**)".
- If Q is temporally projectible, $(t^{**}-(t+\varepsilon)) < \log(.5)/\log(\rho)$, and $((t+\varepsilon) < t^* \leq t^{**})$, then "(A when P is causally sufficient for Q after an interval ε) & A-at-t & P-at-t" is a defeasible reason for "Q-throughout-[t* , t**]".

This is implemented as follows:

```
(def-backwards-reason CAUSAL-IMPLICATION
  :conclusions  "(Q throughout (op time* time**))"
  :condition (and (<= time* time**) ((time** - time*) < 693))
  :forwards-premises
  "(A when P is causally sufficient for Q after an interval interval)"
    (:condition (every #'temporally-projectible (conjuncts Q)))
  "(A at time)"
  (:condition
   (or (and (eq op 'clopen) ((time + interval) <= time*) (time* < time**) ((time** - (time + interval)) < 693))
       (and (eq op 'closed) ((time + interval) < time*) (time* <= time**) ((time** - (time + interval)) < 693))
       (and (eq op 'open) ((time + interval) <= time*) (time* < time**) ((time** - (time + interval)) < 693))))
  :backwards-premises
  "(P at time)"
  :variables  A P Q interval time time* time** op
  :strength  (- (* 2 (expt *temporal-decay* (- time** time))) 1)
  :defeasible?  T)
```

We also need an indexical version of CAUSAL-IMPLICATION:

```
(def-backwards-reason INDEXICAL-CAUSAL-IMPLICATION
  :conclusions  "Q"
  :forwards-premises
  "(A when P is causally sufficient for Q after an interval interval)"
    (:condition (every #'temporally-projectible (conjuncts Q)))
  "(A at time)"
```

```
(:condition (and ((time + interval) < now) ((now - (time + interval)) < 693)))
:backwards-premises
"(P at time)"
:variables  A P Q interval time
:defeasible?  T
:strength  (- (* 2 (expt *temporal-decay* (- now time))) 1)
:temporal? T)
```

These principles can be illustrated by applying them to the Yale Shooting Problem. The problem arises from the fact that if we have INDEXICAL-CAUSAL-IMPLICATION but do not have INDEXICAL-CAUSAL-UNDERCUTTER, then it cannot be inferred that Jones is dead after the shooting. Instead we get collective defeat, as indicated in figure 7.

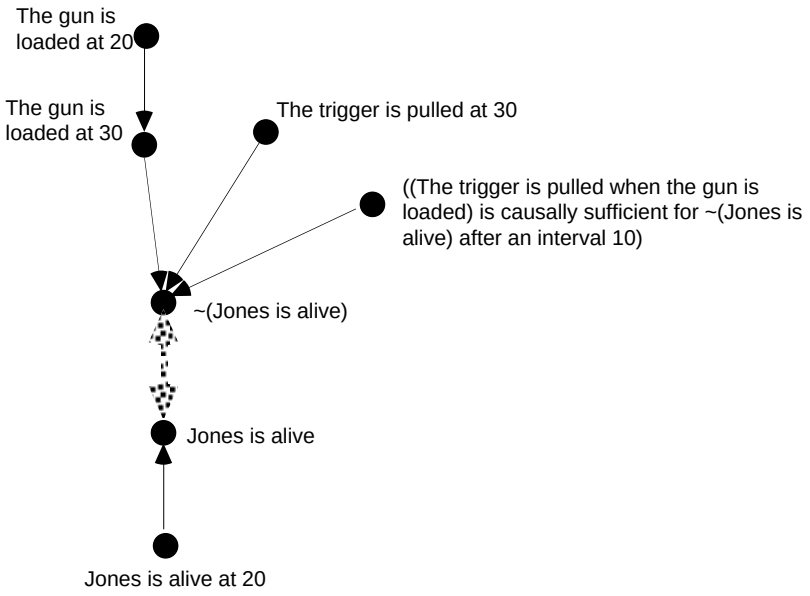


Figure 7. The Unsolved Yale Shooting Problem

On the other hand, if we allow the reasoner to use INDEXICAL-CAUSAL-UNDERCUTTER, it is able to conclude that Jones is dead after the shooting:

```
======================================================================
Problem number 13:  This is the solved Yale Shooting Problem.  I know that the gun being fired while loaded will cause Jones to become dead.  I know that the gun is initially loaded, and Jones is initially alive.  Later, the gun is fired.  Should I conclude that Jones becomes dead?  (It is assumed that ~(Jones is alive) is temporally-projectible.)

Forwards-substantive-reasons:

Backwards-substantive-reasons:
     INDEXICAL-TEMPORAL-PROJECTION
     TEMPORAL-PROJECTION
     INDEXICAL-CAUSAL-UNDERCUTTER
     INDEXICAL-CAUSAL-IMPLICATION

Start reasoning at cycle 50
```

Inputs:

Given:
((Jones is alive) at 20) : with justification = 1.0
(the gun is loaded at 20) : with justification = 1.0
(the gun is fired at 30) : with justification = 1.0
(the gun is fired when the gun is loaded is causally sufficient for ~(Jones is alive) after an interval 10) : with
justification = 1.0

Ultimate epistemic interests:
(? (Jones is alive)) degree of interest = 0.75
==========
THE FOLLOWING IS THE REASONING INVOLVED IN THE SOLUTION
Nodes marked DEFEATED have that status at the end of the reasoning.

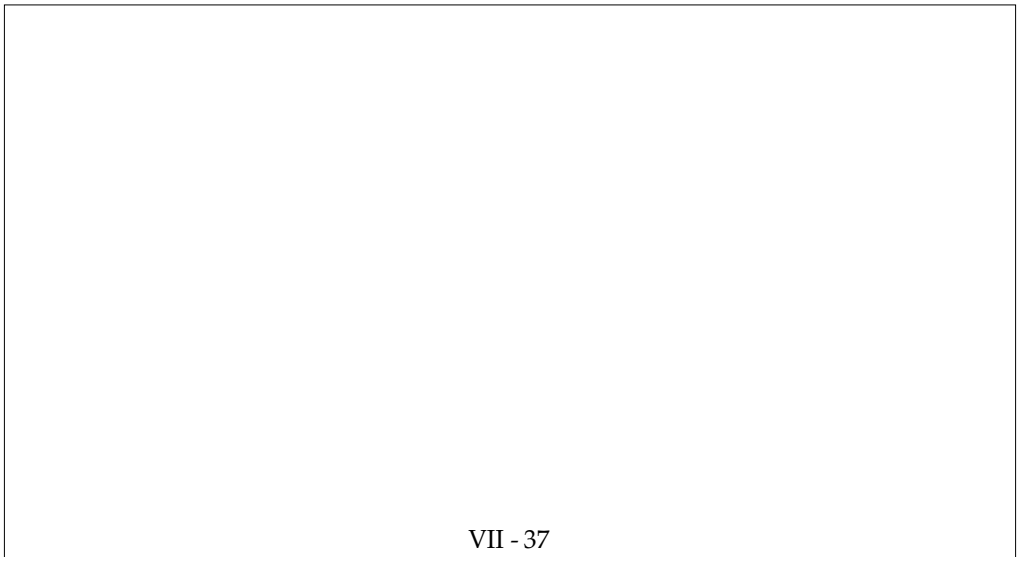
1
((Jones is alive) at 20)
given
2
(the gun is loaded at 20)
given
3
(the gun is fired at 30)
given
4
(the gun is fired when the gun is loaded is causally sufficient for ~(Jones is alive) after an interval 10)
given
1
interest: (Jones is alive)
This is of ultimate interest
Of interest as a defeater for support-link 7 for node 7
2
interest: ~(Jones is alive)
This is of ultimate interest
Of interest as a defeater for support-link 5 for node 5
5
(Jones is alive) DEFEATED
Inferred by:
support-link #5 from { 1 } by INDEXICAL-TEMPORAL-PROJECTION defeaters: { 8 , 7 }
DEFEATED
defeatees: { link 7 for node 7 }
This discharges interest 1
3
interest: (((Jones is alive) at 20) ⊗ (Jones is alive))
Of interest as a defeater for support-link 5 for node 5
==========
Justified belief in (Jones is alive)
with undefeated-degree-of-support 0.935
answers #<Query 1: (? (Jones is alive))>
==========
4
interest: (the gun is loaded at 30)
For interest 2 by INDEXICAL-CAUSAL-IMPLICATION
For interest 3 by INDEXICAL-CAUSAL-UNDERCUTTER
This interest is discharged by node 6
6

```
 (the gun is loaded at 30)
 Inferred by:
          support-link #6 from { 2 } by TEMPORAL-PROJECTION
 This discharges interest 4
 # 7
 ~(Jones is alive)
 Inferred by:
          support-link #7 from { 4 , 3 , 6 } by INDEXICAL-CAUSAL-IMPLICATION  defeaters: { 5 }
 defeatees: { link 5 for node 5 }
 This node is inferred by discharging interest #2
 # 8
 (((Jones is alive) at 20) ⊗ (Jones is alive))
 Inferred by:
          support-link #8 from { 4 , 3 , 6 } by INDEXICAL-CAUSAL-UNDERCUTTER
 defeatees: { link 5 for node 5 }
 This node is inferred by discharging interest #3
          vvvvvvvvvvvvvvvvvvvvvvvvvvvv
          #<Node 5> has become defeated.
          vvvvvvvvvvvvvvvvvvvvvvvvvvvv
          ========================================
          Justified belief in ~(Jones is alive)
          with undefeated-degree-of-support 0.949
          answers #<Query 1: (? (Jones is alive))>
          ========================================
          ========================================
          Lowering the undefeated-degree-of-support of (Jones is alive)
          retracts the previous answer to #<Query 1: (? (Jones is alive))>
          ========================================
============================================================


================== ULTIMATE EPISTEMIC INTERESTS ====================
 Interest in (? (Jones is alive))
 is answered by node 7:  ~(Jones is alive)
------------------------------------------------
```

This reasoning produces the inference-graph of figure 8.

This, as far as I know, is the first *implemented* fully general solution to the Frame Problem. It is noteworthy how simple it is to implement such principles in OSCAR, making OSCAR a potent tool for epistemological analysis.



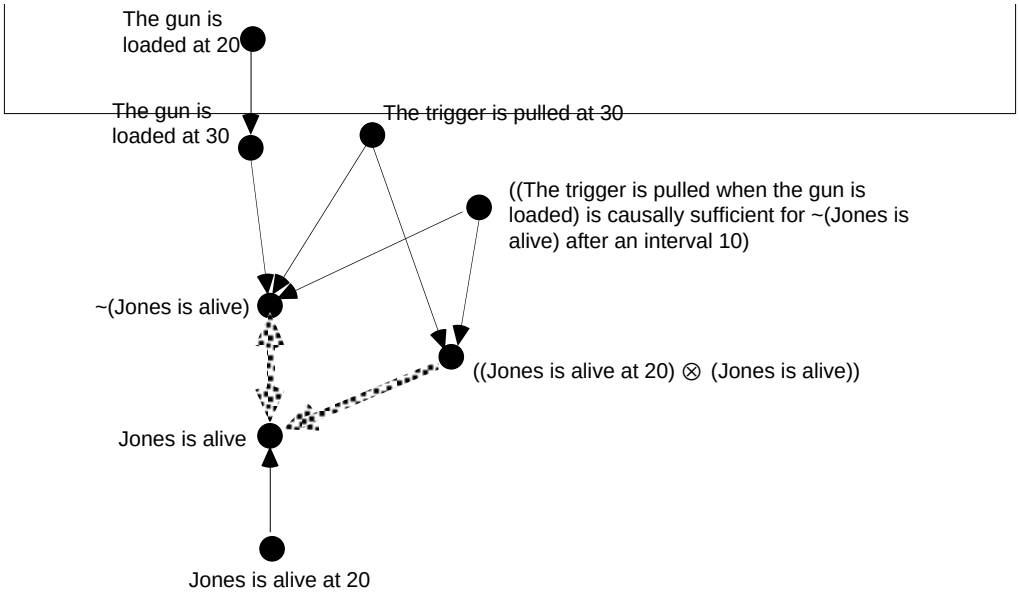


Figure 8. The Solved Yale Shooting Problem

11. The Qualification and Ramification Problems

Two problems that are often associated with the Frame Problem are the Qualification and Ramification Problems. Like the Frame Problem, these arose initially within the framework of attempts to reason about change deductively. The Frame Problem concerned the proliferation of frame axioms—axioms concerning what does not change. The Qualification and Ramification Problems concerned the difficulty in correctly formulating axioms about what does change. The Qualification Problem is the problem of getting the antecedent right in axioms like "A match's being struck when it is dry, in the presence of oxygen, ... , is causally sufficient for it to light". The difficulty is that we are typically unable to fill in the ellipsis and give a complete list of the conditions required to cause a particular effect. McCarthy [1977] illustrated this with his famous "banana in the tailpipe" example. Most people are unaware that a car will not start if the tailpipe is blocked, e.g., by a banana. Thus if asked to state the conditions under which turning the ignition key will start the car, they will be unable to do so. An allied difficulty is that even if we could completely enumerate the conditions required, deductive reasoning about change would require us to then deductively verify that all of those conditions are satisfied—something that human beings clearly do not generally do.

Within the present framework, the solution to the Qualification Problem seems to be fairly simple. I defined "A when P is causally sufficient for Q after an interval ε" to mean

$$(\forall t)\{(\text{A-at-t} \;\&\; \text{P-at-t}) \equiv\!> (\exists \delta)\text{Q-throughout-}(t+\varepsilon\,,\, t+\varepsilon+\delta]\}.$$

So defined, the causal knowledge that we use in reasoning about change is not generally of this form. This is for two reasons. First, we rarely have more than a rough estimate of the value of ε. Second, we are rarely in a position to formulate P precisely. That latter is just the Qualification Problem. Our knowledge actually takes the form:

$(\exists P^*)(\exists \varepsilon)[P^*$ is true $\& \varepsilon \leq \varepsilon^* \;\&$ (A when (P & P^*) is causally sufficient for Q after an interval ε)].

P formulates the known preconditions for the causal sufficiency, P* the unknown precondtions, and ε∗ is the known upper bound on ε. Let us abbreviate this as "A when P is weakly causally sufficient for Q after an interval ε∗". We acquire knowledge of weak causal sufficiency inductively. For example, we learn inductively that striking a dry match is usually weakly causally sufficient for it to light after a negligible interval. If we then examine CAUSAL-UNDERCUTTER and CAUSAL-IMPLICATION:

CAUSAL-UNDERCUTTER

Where $t_0 \leq t_1$ and $(t_1+\varepsilon) < t$, "A-at-t_1 & Q-at-t_1 & (A when Q is causally sufficient for ~P after an interval ε)" is a defeasible undercutting defeater for the inference from P-at-t_0 to P-throughout-⟨t* t⟩ by TEMPORAL-PROJECTION.

CAUSAL-IMPLICATION

- If Q is temporally projectible, $(t^{**}-(t+\varepsilon)) < \log(.5)/\log(\rho)$, and $((t+\varepsilon) \leq t^* < t^{**})$, then "(A when P is causally sufficient for Q after an interval ε) & A-at-t & P-at-t" is a defeasible reason for "Q-throughout-(t* , t**]" and for "Q-throughout-(t* , t**)".
- If Q is temporally projectible, $(t^{**}-(t+\varepsilon)) < \log(.5)/\log(\rho)$, and $((t+\varepsilon) < t^* \leq t^{**})$, then "(A when P is causally sufficient for Q after an interval ε) & A-at-t & P-at-t" is a defeasible reason for "Q-throughout-[t* , t**]".

we find that both continue to hold if we reconstrue "causally suffficient" to mean "weakly causally sufficient". Thus we can reason about change in the same way even with incomplete causal knowledge. This resolves the Qualification Problem.

The Ramification Problem arises from the observation that in realistically complex environments, we cannot formulate axioms that completely specify the effects of actions or events. People sometimes refer to these as "actions with ramifications", as if these were peculiar actions. But in the real world, all actions have infinitely many ramifications stretching into the indefinite future. This is a problem for reasoning about change deductively, but does not seem to be a problem for reasoning about change defeasibly in the present framework. Consider how human beings deal with this difficulty. Our inability to enumerate all the effects of an action means that we cannot formulate true successor-state axioms (axioms that roll frame axioms and effect axioms into a single axiom in the situation calculus to completely describe the next situation). But we do not have to. CAUSAL-UNDERCUTTER and CAUSAL-IMPLICATION allow us to reason defeasibly about change on the basis of our incomplete knowledge.

Another aspect of the Ramification Problem is that even if it were sufficient to formulate successor-state axioms using just the effects that we actually know to be produced by an action, listing all of the known effects would make the axiom so complex that a theorem prover would find it too unwieldy to use. For example, if we think about it for a while, we must enumerate among the effects of striking a match such things as displacing air around the match, marginally depleting the ozone layer, raising the temperature of the earth's atmosphere, marginally illuminating Alpha Centauri, making that match unavailable for future use, etc. These are effects that we typically do not care about, and so we do not reason about them. But this does not mean that we can omit them from a successor-state axiom with impunity, because occasionally we might care about one of them.

Within the present framework, this is not a problem. Reasoning in OSCAR is interest-driven, and CAUSAL-IMPLICATION is a backwards-reason. This means that we only reason about potential effects of actions and events when they are of interest to us. Whether they are of interest can vary from circumstance to circumstance, allowing our reasoning to vary similarly, without our having to revise our knowledge base or rules of inference to accommodate the change. Deductive reasoning in terms of successor-state axioms is too crude an implement to reflect this feature of human reasoning, but the current framework handles it automatically. The conclusion is that the Ramification Problem simply does

not arise in this framework.

11. The Extended Prediction Problem

The literature on the Frame Problem has generally focused on toy problems like the Yale Shooting Problem. Experience elsewhere in AI should make us wary of such an approach. Solutions to toy problems may not scale up to problems of realistic complexity. To see how the present proposals fare in this connection, consider "the extended prediction problem" introduced by Shoham [1987]. He suggests that even if reasoners are able to reason about the immediate future, as in the Yale Shooting Problem, they will have difficulty using causal information to make predictions about the relatively distant future. He takes this to be illustrated by the traditional classical physics problem of predicting the future positions of colliding billiard balls. Consider two (dimensionless) billiard balls whose positions and velocities are known at initial times, and suppose they are rolling on a flat frictionless surface. Suppose further that they are on a collision course. The problem is to predict their positions at some particular time after the impending collision.

Although this seems like a simple problem, it is considerably more difficult than the toy problems addressed elsewhere in the literature. It is worth noting that even human physicists have one kind of difficulty solving it. Once it is recognized that a collision is going to occur, there is no difficulty solving the problem, but *noticing* the impending collision is not a trivial task. If this is not obvious, consider trying to solve the same problem for 100 billiard balls rolling about on a billiard table. In trying to solve such a problem, a human being uses various heuristics to detect what may be collisions (e.g., seeing whether the directions of motion of two billiard balls cross), and then is able to determine by explicit reasoning whether these possible collisions are real collisions. What I will show is that if OSCAR's attention is drawn to the appropriate possible collisions, then OSCAR, equipped with the reasons formulated in this paper, can solve this problem as readily as human beings.

Solving this problem requires assuming that the velocity of a billiard ball (the vector quantity of the speed in a certain direction) will not change unless it collides with another billiard ball. This is intimately connected with Newton's laws of motion. It is of interest to think about Newtonian kinematics in connection with the preceding account of causal change. Newton's laws of motion tell us that the velocity of an object remains unchanged in the absense of external forces. If velocity is taken to be temporally projectible, this is very much like applying TEMPORAL-PROJECTION to it. If it seems dubious that velocity should be temporally projectible, notice that velocity and position must both be relativized to inertial frames of reference. This has the consequence that velocity remains unchanged relative to one inertial frame iff there is an inertial frame relative to which the velocity remains constantly zero and hence position remains unchanged. Thus velocity is a stable property iff position is. And it seems eminently plausible that position is temporally projectible. Thus we can regard this part of Newtonian kinematics as a mathematicization of this aspect of commonsense reasoning. So construed, Newtonian force laws have the same status as the putative law about loaded guns that is employed in the Yale Shooting Problem. Correct reasoning about the consequences of billiard balls colliding encounters precisely the same difficulties as those encountered in predicting the consequences of pulling the triggers on loaded guns. Newtonian physics avoids this difficulty at a theoretical level by adopting a mathematical model of the entire world, and solving the "world equation". But of course, this is not something that rational agent can really do in a complex environment (even if he is a Newtonian physicist). Using Newtonian physics to get around in the real world requires epistemic principles like TEMPORAL-PROJECTION and CAUSAL-UNDERCUTTER just as much as the use of more naive physical principles does.

Suppose, then, that we begin with two billiard balls whose positions and velocities are known at an initial time t_0. Suppose further that if those velocities remain unchanged until time t_1, the billiard balls will collide at that time. Assuming that it is a perfectly

elastic collision, and the balls have the same mass, their new velocities are easily computed—the balls simply exchange trajectories. If we assume that the resulting velocities remain unchanged until time t_2, we can then compute the positions of the balls at t_2. This reasoning uses TEMPORAL-PROJECTION to conclude that the velocities are unchanged prior to the collision, and CAUSAL-UNDERCUTTER and CAUSAL-IMPLICATION (the latter builds in TEMPORAL-PROJECTION) to infer the velocities after the collision.

To make the problem concrete, suppose b_1 is at (0 3) at time 0 with velocity (.5 0), and b_2 is at (1 0) at time 0 with velocity (.4 .3). Then the balls should move as diagrammed in figure 9.

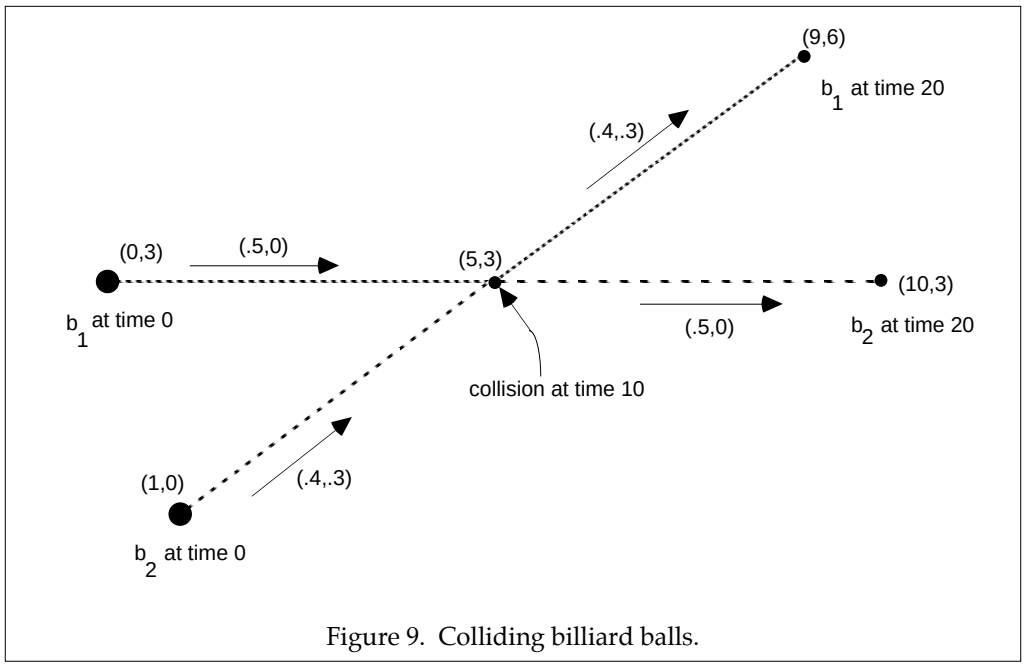


Figure 9. Colliding billiard balls.

Positions can be represented as pairs (x y) of x– and y–coordinates, and velocities as pairs (v_x v_y) of speeds along the x– and y–axes. To compute positions after an object has moved from an initial position with a constant velocity for a specified amount of time, we need the following kinematic principle:

NEW-POSITION
"(the position of s is (x_0 y_0))-at-t_0 & (the velocity of s is (v_x v_y) throughout (t_0 , t_1]" is a conclusive reason for "(the position of s is ((x_0 + (v_x· (t_1 – t_0))) (·y_0 + (v_y· (t_1 – t_0)))))-at-t_1".

```
(def-backwards-reason NEW-POSITION
   :conclusions  "((the position of b is (x y)) at time1)"
   :forwards-premises
      "((the position of b is (x0 y0)) at time0)"
       (:condition (time0 < time1))
   :backwards-premises
      "(∃vx)(∃vy)
       (& ((the velocity of b is (vx vy)) throughout (clopen time0 time1))
          (x = (x0 + (vx * (time1 - time0))))
          (y = (y0 + (vy * (time1 - time0)))))"
   :variables  b time1 x y x0 y0 time0)
```

We also need the following causal principle governing elastic collisions between billiard

balls of the same mass. To keep the mathematics simple, I will just consider the case of dimensionless colliding billiard balls having the same mass. In that case, the colliding balls simply exchange velocities:

(10) $(\forall b_1)(\forall b_2)(\forall {}_1v_x)(\forall {}_1v_y)(\forall {}_2v_x)(\forall {}_2v_y)$
{[b_1 is a dimensionless billiard ball & b_2 is a dimensionless billiard ball & b_1 and b_2 have the same mass & $({}_1v_x^2 + {}_1v_y^2) = ({}_2v_x^2 + {}_2v_y^2)$] $\supset$ [(b_1 and b_2 collide) when (the velocity of b_1 is $({}_1v_x\ {}_1v_y)$) is causally sufficient for (the velocity of b_2 is $({}_1v_x\ {}_1v_y)$) after an interval 0]}.

Dimensionless billiard balls collide iff they have the same position:

COLLISION

"$(\exists x)$[(the position of b_1 is x)-at-t & (the position of b_2 is x)-at-t]" is a conclusive reason for "(b_1 and b_2 collide)-at-t".

```
(def-backwards-reason COLLISION
  :conclusions  "((b1 and b2 collide) at time)"
  :backwards-premises
    "(∃x)(∃y)(((the position of b1 is (x y)) at time) & ((the position of b2 is (x y)) at time))"
  :variables   b1 b2 time)
```

Now suppose b_1 is at (0 3) at time 0 with velocity (.5 0), and b_2 is at (1 0) at time 0 with velocity (.4 .3). If the velocities remain unchanged, b_1 and b_2 should collide at time 10. If we pose this as a problem for OSCAR, by COLLISION, an interest in whether b_1 and b_2 collide at 10 generates an interest in their positions at 10. Because we know their positions at 0, NEW-POSITION generates an interest in their velocities between 0 and 10. We know the velocities at 0, and TEMPORAL-PROJECTION leads to an inference that those velocities remain unchanged between 0 and 10. From that we can compute the positions at 10, and infer that b1 and b2 collide at 10.

However, TEMPORAL-PROJECTION also leads to an inference that the positions at 10 are the same as those at 0. That inference must be defeated somehow, but the principles described so far will not accomplish that. This can be accomplished by adding another principle of causal sufficiency. It is a logically contingent but physically necessary fact (at least, according to Newtonian physics) that billiard balls, and other objects of nonzero mass, do not undergo infinite accelerations. As such, if a billiard ball has nonzero velocity at a time, that is causally sufficient for there being a subsequent time at which its position has changed. Thus we have:

(11) $(\forall b)(\forall x)(\forall y)(\forall v_x)(\forall v_y)$
{(the position of b is (x y)) when ((the velocity of b is $(v_x\ v_y)$) & $\sim((v_x\ v_y) = (0.0\ 0.0))$) is causally sufficient for ~(the position of b is (x , y)) after an interval 0]

Because the velocities at 0 are nonzero, CAUSAL-UNDERCUTTER defeats the problematic inference that the billiard balls do not move from 0 to 10, leaving the single undefeated conclusion regarding the positions at 10 that both balls are at (5.0 3.0).

So far, we are given the positions and velocities of b_1 and b_2 at 0, and we have inferred their velocities between 0 and 10, their positions at 10, and have concluded that they collide at 10. Now suppose we want to know the position of b_1 at 20. Given a knowledge of the position of b_1 at 10, NEW-POSITION generates an interest in the velocity of b_1 between 10 and 20. By CAUSAL-IMPLICATION, we can infer that the velocity of b_1 between 10 and 20 is (.4 .3). From this NEW-POSITION enables us to infer (correctly) that the position of b_1 at 20 is (9.0 6.0).

However, there are conflicting inferences that can also be made, and they must be defeated. By TEMPORAL-PROJECTION, we can infer that the velocity of b_1 between 10 and 20 is the same as at 10. This is defeated as above by CAUSAL-UNDERCUTTER, using

(10), because we know the velocities of b_1 and b_2 at 10 and know they collide. Similarly, we can use TEMPORAL-PROJECTION to infer that the velocity of b_1 between 0 and 20 is the same as at 0, which we know to be (.5 0). This is also defeated by CAUSAL-UNDERCUTTER, using (10).

By TEMPORAL-PROJECTION, we can infer that the position of b_1 at 20 is the same as at 0. But this is defeated by CAUSAL-UNDERCUTTER, using (11), because we know that the velocity of b_1 at 0 is nonzero. Finally, by TEMPORAL-PROJECTION, we can infer that the position of b_1 at 20 is the same as at 10. But this is also defeated by CAUSAL-UNDERCUTTER, using (11), because we know that the velocity of b_1 at 10 is nonzero. Thus all the undesirable inferences are defeated, leaving OSCAR with the desired conclusion that the position of b_1 at 20 is (9.0 6.0).

==========

Problem number 17: This is the Extended Prediction Problem.

1. We are given the velocities of b1 and b2 at 0, and are told they collide at (5 3) at 10. We are interested in the position of b1 at 20. Given knowledge of the position of b1 at 10, this generates an interest in the velocity of b1 between 10 and 20.

2. By causal-implication, we can infer that the velocity of b1 between 10 and 20 is (.4 .3). From this we can compute that the position of b1 at 20 is (9.0 6.0).

3. By temporal projection, we can also infer that the velocity of b1 at 20 is (.5 .0). But this is defeated by causal-undercutter, because we also know that if the velocity is (.4 .3) then it is not (.5 .0).

4. By temporal projection, we can infer that the position of b1 at 20 is the same as at 0. But this is defeated by causal-undercutter, because we know that the velocity of b1 at 0 is nonzero.

5. By temporal projection, we can infer that the position of b1 at 20 is the same as at 10. This is defeated in the same fashion as (4), because we know the velocity of b1 between 0 and 10, and we are given that 10 is between 0 and 10.

Forwards-substantive-reasons:
- POSITION-INCOMPATIBILITY

Backwards-substantive-reasons:
- CAUSAL-IMPLICATION
- TEMPORAL-PROJECTION
- CAUSAL-UNDERCUTTER
- COLLISION
- NEW-POSITION
- PAIR-NONIDENTITY
- PAIR-NONIDENTITY-AT-TIME
- &-AT-INTRO

Given:
- ((the position of b1 is (0.0 3.0)) at 0) : with justification = 1.0
- ((the position of b2 is (1.0 0.0)) at 0) : with justification = 1.0
- ((the velocity of b1 is (0.5 0.0)) at 0) : with justification = 1.0
- ((the velocity of b2 is (0.4 0.3)) at 0) : with justification = 1.0
- (b1 is a dimensionless billiard ball) : with justification = 1.0
- (b2 is a dimensionless billiard ball) : with justification = 1.0
- (∀b)(∀x)(∀y)(∀vx)(∀vy)((the position of b is (x y)) when ((the velocity of b is (vx vy)) & ~((vx vy) = (0.0 0.0))) is causally sufficient for ~(the position of b is (x y)) after an interval 0) : with justification = 1.0
- (∀b1)(∀b2)(∀v1x)(∀v1y)(∀v2x)(∀v2y)((((b1 is a dimensionless billiard ball) & (b2 is a dimensionless billiard ball)) & ((same-mass b1 b2) & (((v1x expt 2) + (v1y expt 2)) = ((v2x expt 2) + (v2y expt 2))))) → ((b1 and b2 collide) when (the velocity of b2 is (v2x v2y)) is causally sufficient for (the velocity of b1 is (v2x v2y)) after an interval 0)) : with justification = 1.0

(same-mass b1 b2) : with justification = 1.0
(5.0 = (0.0 + (0.5 * (10 - 0)))) : with justification = 1.0
(3.0 = (3.0 + (0.0 * (10 - 0)))) : with justification = 1.0
(5.0 = (1.0 + (0.4 * (10 - 0)))) : with justification = 1.0
(3.0 = (0.0 + (0.3 * (10 - 0)))) : with justification = 1.0
(9.0 = (5.0 + (0.4 * (20 - 10)))) : with justification = 1.0
(6.0 = (3.0 + (0.3 * (20 - 10)))) : with justification = 1.0
(((0.5 expt 2) + (0.0 expt 2)) = ((0.4 expt 2) + (0.3 expt 2))) : with justification = 1.0

Ultimate epistemic interests:
(? ((b1 and b2 collide) at 10)) degree of interest = 0.75
(? x)(? y)((the position of b1 is (x y)) at 20) degree of interest = 0.75

==========

THE FOLLOWING IS THE REASONING INVOLVED IN THE SOLUTION
Nodes marked DEFEATED have that status at the end of the reasoning.

1
((the position of b1 is (0.0 3.0)) at 0)
given
2
((the position of b2 is (1.0 0.0)) at 0)
given
3
((the velocity of b1 is (0.5 0.0)) at 0)
given
This discharges interest 75
4
((the velocity of b2 is (0.4 0.3)) at 0)
given
5
(b1 is a dimensionless billiard ball)
given
6
(b2 is a dimensionless billiard ball)
given
7
(∀b) (∀x) (∀y) (∀vx) (∀vy)((the position of b is (x y)) when ((the velocity of b is (vx vy)) & ~((vx vy) = (0.0 0.0))) is causally sufficient for ~(the position of b is (x y)) after an interval 0)
given
8
(∀b1) (∀b2) (∀v1x) (∀v1y) (∀v2x) (∀v2y)((((b1 is a dimensionless billiard ball) & (b2 is a dimensionless billiard ball)) & ((same-mass b1 b2) & (((v1x expt 2) + (v1y expt 2)) = ((v2x expt 2) + (v2y expt 2))))) -> ((b1 and b2 collide) when (the velocity of b2 is (v2x v2y)) is causally sufficient for (the velocity of b1 is (v2x v2y)) after an interval 0))
given
9
(same-mass b1 b2)
given
10
(5.0 = (0.0 + (0.5 * (10 - 0))))
given
This discharges interest 37
11
(3.0 = (3.0 + (0.0 * (10 - 0))))
given
This discharges interest 38
12
(5.0 = (1.0 + (0.4 * (10 - 0))))

given
This discharges interest 49
13
(3.0 = (0.0 + (0.3 * (10 - 0))))
given
This discharges interest 50
14
(9.0 = (5.0 + (0.4 * (20 - 10))))
given
This discharges interest 61
15
(6.0 = (3.0 + (0.3 * (20 - 10))))
given
This discharges interest 62
16
(((0.5 expt 2) + (0.0 expt 2)) = ((0.4 expt 2) + (0.3 expt 2)))
given
1
interest: ((the position of b1 is (y1 y0)) at 20)
This is of ultimate interest
2
interest: ((b1 and b2 collide) at 10)
This is of ultimate interest
3
interest: (∃x)(∃y)(((the position of b1 is (x y)) at 10) & ((the position of b2 is (x y)) at 10))
For interest 2 by *collision*
This interest is discharged by node 64
17
((the position of b1 is (0.0 3.0)) at 20) DEFEATED
Inferred by:
support-link #17 from { 1 } by *temporal-projection* defeaters: { 82 , 56 } DEFEATED
This discharges interest 1
6
interest: (((the position of b1 is (0.0 3.0)) at 0) ⊗ ((the position of b1 is (0.0 3.0)) at 20))
Of interest as a defeater for support-link 17 for node 17
7
interest: ~((the position of b1 is (0.0 3.0)) at 20)
Of interest as a defeater for support-links: { link 28 for node 17, link 17 for node 17}
==========
Justified belief in ((the position of b1 is (0.0 3.0)) at 20)
with undefeated-degree-of-support 0.960
answers #<Query 2: (? x)(? y)((the position of b1 is (x y)) at 20)>
==========
18
(∀x) (∀y) (∀vx) (∀vy)((the position of x2 is (x y)) when ((the velocity of x2 is (vx vy)) & ~((vx vy) = (0.0 0.0))) is causally sufficient for ~(the position of x2 is (x y)) after an interval 0)
Inferred by:
support-link #18 from { 7 } by UI
19
(∀y) (∀vx) (∀vy)((the position of x2 is (x3 y)) when ((the velocity of x2 is (vx vy)) & ~((vx vy) = (0.0 0.0))) is causally sufficient for ~(the position of x2 is (x3 y)) after an interval 0)
Inferred by:
support-link #19 from { 18 } by UI
20

(∀vx) (∀vy)((the position of x2 is (x3 x4)) when ((the velocity of x2 is (vx vy)) & ~((vx vy) = (0.0 0.0))) is causally sufficient for ~(the position of x2 is (x3 x4)) after an interval 0)
Inferred by:
support-link #20 from { 19 } by UI
21
(∀vy)((the position of x2 is (x3 x4)) when ((the velocity of x2 is (x5 vy)) & ~((x5 vy) = (0.0 0.0))) is causally sufficient for ~(the position of x2 is (x3 x4)) after an interval 0)
Inferred by:
support-link #21 from { 20 } by UI
22
((the position of x2 is (x3 x4)) when ((the velocity of x2 is (x5 x6)) & ~((x5 x6) = (0.0 0.0))) is causally sufficient for ~(the position of x2 is (x3 x4)) after an interval 0)
Inferred by:
support-link #22 from { 21 } by UI
8
interest: (((the velocity of b1 is (x5 x6)) & ~((x5 x6) = (0.0 0.0))) at 0)
For interest 13 by *causal-undercutter*
For interest 6 by *causal-undercutter*
This interest is discharged by node 80
23
(∀b2) (∀v1x) (∀v1y) (∀v2x) (∀v2y)((((x7 is a dimensionless billiard ball) & (b2 is a dimensionless billiard ball)) & ((same-mass x7 b2) & (((v1x expt 2) + (v1y expt 2)) = ((v2x expt 2) + (v2y expt 2))))) -> ((x7 and b2 collide) when (the velocity of b2 is (v2x v2y)) is causally sufficient for (the velocity of x7 is (v2x v2y)) after an interval 0))
Inferred by:
support-link #23 from { 8 } by UI
24
(∀v1x) (∀v1y) (∀v2x) (∀v2y)((((x7 is a dimensionless billiard ball) & (x8 is a dimensionless billiard ball)) & ((same-mass x7 x8) & (((v1x expt 2) + (v1y expt 2)) = ((v2x expt 2) + (v2y expt 2))))) -> ((x7 and x8 collide) when (the velocity of x8 is (v2x v2y)) is causally sufficient for (the velocity of x7 is (v2x v2y)) after an interval 0))
Inferred by:
support-link #24 from { 23 } by UI
9
interest: (∃y)(((the position of b1 is (y9 y)) at 10) & ((the position of b2 is (y9 y)) at 10))
For interest 3 by EG
This interest is discharged by node 63
25
(∀v1y) (∀v2x) (∀v2y)((((x7 is a dimensionless billiard ball) & (x8 is a dimensionless billiard ball)) & ((same-mass x7 x8) & (((x10 expt 2) + (v1y expt 2)) = ((v2x expt 2) + (v2y expt 2))))) -> ((x7 and x8 collide) when (the velocity of x8 is (v2x v2y)) is causally sufficient for (the velocity of x7 is (v2x v2y)) after an interval 0))
Inferred by:
support-link #25 from { 24 } by UI
10
interest: (((the position of b1 is (y9 y11)) at 10) & ((the position of b2 is (y9 y11)) at 10))
For interest 9 by EG
This interest is discharged by node 62
26
(∀v2x) (∀v2y)((((x7 is a dimensionless billiard ball) & (x8 is a dimensionless billiard ball)) & ((same-mass x7 x8) & (((x10 expt 2) + (x12 expt 2)) = ((v2x expt 2) + (v2y expt 2))))) -> ((x7 and x8 collide) when (the velocity of x8 is (v2x v2y)) is causally sufficient for (the velocity of x7 is (v2x v2y)) after an interval 0))
Inferred by:
support-link #26 from { 25 } by UI

11
interest: ((the position of b1 is (y9 y11)) at 10)
For interest 10 by adjunction
This interest is discharged by node 50
12
interest: (∃vx)(∃vy)(((the velocity of b1 is (vx vy)) throughout (0 , 10]) & ((y9 = (0.0 + (vx * (10 - 0)))) & (y11 = (3.0 + (vy * (10 - 0))))))
For interest 11 by *new-position*
This interest is discharged by node 49

27
((the position of b1 is (0.0 3.0)) at 10) DEFEATED
Inferred by:
support-link #27 from { 1 } by *temporal-projection* defeaters: { 81 , 53 } DEFEATED
13
interest: (((the position of b1 is (0.0 3.0)) at 0) ⊗ ((the position of b1 is (0.0 3.0)) at 10))
Of interest as a defeater for support-link 27 for node 27
14
interest: ~((the position of b1 is (0.0 3.0)) at 10)
Of interest as a defeater for support-link 27 for node 27
17
interest: (((the velocity of b1 is (x5 x6)) & ~((x5 x6) = (0.0 0.0))) at 10)
For interest 41 by *causal-undercutter*
For interest 18 by *causal-undercutter*
For interest 6 by *causal-undercutter*
This interest is discharged by node 86

17
((the position of b1 is (0.0 3.0)) at 20) DEFEATED
Inferred by:
support-link #28 from { 27 } by *temporal-projection* defeaters: { 88 , 56 } DEFEATED
support-link #17 from { 1 } by *temporal-projection* defeaters: { 82 , 56 } DEFEATED
This discharges interest 1
18
interest: (((the position of b1 is (0.0 3.0)) at 10) ⊗ ((the position of b1 is (0.0 3.0)) at 20))
Of interest as a defeater for support-link 28 for node 17
vvvvvvvvvvvvvvvvvvvvvvvvvvvv
The undefeated-degree-of-support of #<Node 17> has increased to 0.980
vvvvvvvvvvvvvvvvvvvvvvvvvvvv

28
(∀v2y)((((x7 is a dimensionless billiard ball) & (x8 is a dimensionless billiard ball)) & ((same-mass x7 x8) & (((x10 expt 2) + (x12 expt 2)) = ((x13 expt 2) + (v2y expt 2))))) -> ((x7 and x8 collide) when (the velocity of x8 is (x13 v2y)) is causally sufficient for (the velocity of x7 is (x13 v2y)) after an interval 0))
Inferred by:
support-link #29 from { 26 } by UI

29
((((x7 is a dimensionless billiard ball) & (x8 is a dimensionless billiard ball)) & ((same-mass x7 x8) & (((x10 expt 2) + (x12 expt 2)) = ((x13 expt 2) + (x14 expt 2))))) -> ((x7 and x8 collide) when (the velocity of x8 is (x13 x14)) is causally sufficient for (the velocity of x7 is (x13 x14)) after an interval 0))
Inferred by:
support-link #30 from { 28 } by UI

30
(((x7 is a dimensionless billiard ball) & (x8 is a dimensionless billiard ball)) -> (((same-mass x7 x8) & (((x10 expt 2) + (x12 expt 2)) = ((x13 expt 2) + (x14 expt 2)))) -> ((x7 and x8 collide) when (the velocity of x8 is (x13 x14)) is causally sufficient for (the velocity of x7 is (x13 x14)) after an interval 0)))
Inferred by:
support-link #31 from { 29 } by exportation

31
((x7 is a dimensionless billiard ball) -> ((x8 is a dimensionless billiard ball) -> (((same-mass x7 x8) & (((x10 expt 2) + (x12 expt 2)) = ((x13 expt 2) + (x14 expt 2)))) -> ((x7 and x8 collide) when (the velocity of x8 is (x13 x14)) is causally sufficient for (the velocity of x7 is (x13 x14)) after an interval 0))))
Inferred by:
support-link #32 from { 30 } by exportation
33
((x8 is a dimensionless billiard ball) -> (((same-mass b1 x8) & (((x10 expt 2) + (x12 expt 2)) = ((x13 expt 2) + (x14 expt 2)))) -> ((b1 and x8 collide) when (the velocity of x8 is (x13 x14)) is causally sufficient for (the velocity of b1 is (x13 x14)) after an interval 0)))
Inferred by:
support-link #34 from { 31 , 5 } by modus-ponens1
38
(((same-mass b1 b2) & (((x10 expt 2) + (x12 expt 2)) = ((x13 expt 2) + (x14 expt 2)))) -> ((b1 and b2 collide) when (the velocity of b2 is (x13 x14)) is causally sufficient for (the velocity of b1 is (x13 x14)) after an interval 0))
Inferred by:
support-link #39 from { 33 , 6 } by modus-ponens1
40
((same-mass b1 b2) -> ((((x10 expt 2) + (x12 expt 2)) = ((x13 expt 2) + (x14 expt 2))) -> ((b1 and b2 collide) when (the velocity of b2 is (x13 x14)) is causally sufficient for (the velocity of b1 is (x13 x14)) after an interval 0)))
Inferred by:
support-link #41 from { 38 } by exportation
41
((((x10 expt 2) + (x12 expt 2)) = ((x13 expt 2) + (x14 expt 2))) -> ((b1 and b2 collide) when (the velocity of b2 is (x13 x14)) is causally sufficient for (the velocity of b1 is (x13 x14)) after an interval 0))
Inferred by:
support-link #42 from { 40 , 9 } by modus-ponens1
42
((b1 and b2 collide) when (the velocity of b2 is (0.4 0.3)) is causally sufficient for (the velocity of b1 is (0.4 0.3)) after an interval 0)
Inferred by:
support-link #43 from { 41 , 16 } by modus-ponens1
23
interest: (∃vy)(((the velocity of b1 is (y18 vy)) throughout (0 , 10]) & ((y9 = (0.0 + (y18 * (10 - 0)))) & (y11 = (3.0 + (vy * (10 - 0))))))
For interest 12 by EG
This interest is discharged by node 48
26
interest: (((the velocity of b1 is (y18 y21)) throughout (0 , 10]) & ((y9 = (0.0 + (y18 * (10 - 0)))) & (y11 = (3.0 + (y21 * (10 - 0))))))
For interest 23 by EG
This interest is discharged by node 47
28
interest: ((the velocity of b2 is (y17 y22)) throughout (0 , 10])
For interest 47 by adjunction
This interest is discharged by node 44
44
((the velocity of b2 is (0.4 0.3)) throughout (0 , 10])
Inferred by:
support-link #45 from { 4 } by *temporal-projection*
This discharges interest 28
33
interest: ((the velocity of b1 is (y18 y21)) throughout (0 , 10])
For interest 26 by adjunction
This interest is discharged by node 45

45
((the velocity of b1 is (0.5 0.0)) throughout (0 , 10])
Inferred by:
support-link #46 from { 3 } by *temporal-projection* defeaters: { 54 }
This discharges interest 33
35
interest: ~((the velocity of b1 is (0.5 0.0)) throughout (0 , 10])
Of interest as a defeater for support-link 46 for node 45
36
interest: ((y9 = (0.0 + (0.5 * (10 - 0)))) & (y11 = (3.0 + (0.0 * (10 - 0)))))
For interest 26 by adjunction
This interest is discharged by node 46
37
interest: (y9 = (0.0 + (0.5 * (10 - 0))))
For interest 36 by adjunction
This interest is discharged by node 10
38
interest: (y11 = (3.0 + (0.0 * (10 - 0))))
For interest 36 by adjunction
This interest is discharged by node 11
46
((5.0 = (0.0 + (0.5 * (10 - 0)))) & (3.0 = (3.0 + (0.0 * (10 - 0)))))
Inferred by:
support-link #47 from { 10 , 11 } by adjunction
This node is inferred by discharging interest #36
47
(((the velocity of b1 is (0.5 0.0)) throughout (0 , 10]) & ((5.0 = (0.0 + (0.5 * (10 - 0)))) & (3.0 = (3.0 + (0.0 * (10 - 0))))))
Inferred by:
support-link #48 from { 45 , 46 } by adjunction
This node is inferred by discharging interest #26
48
(∃vy)(((the velocity of b1 is (0.5 vy)) throughout (0 , 10]) & ((5.0 = (0.0 + (0.5 * (10 - 0)))) & (3.0 = (3.0 + (vy * (10 - 0))))))
Inferred by:
support-link #49 from { 47 } by EG
This node is inferred by discharging interest #23
49
(∃vx)(∃vy)(((the velocity of b1 is (vx vy)) throughout (0 , 10]) & ((5.0 = (0.0 + (vx * (10 - 0)))) & (3.0 = (3.0 + (vy * (10 - 0))))))
Inferred by:
support-link #50 from { 48 } by EG
This node is inferred by discharging interest #12
50
((the position of b1 is (5.0 3.0)) at 10)
Inferred by:
support-link #51 from { 1 , 49 } by *new-position*
This node is inferred by discharging interest #11
39
interest: ((the position of b2 is (5.0 3.0)) at 10)
For interest 10 by adjunction
This interest is discharged by node 61
40
interest: (∃vx)(∃vy)(((the velocity of b1 is (vx vy)) throughout (10 , 20]) & ((y1 = (5.0 + (vx * (20 - 10)))) & (y0 = (3.0 + (vy * (20 - 10))))))
For interest 1 by *new-position*
This interest is discharged by node 72

51
((the position of b1 is (5.0 3.0)) at 20) DEFEATED
Inferred by:
support-link #52 from { 50 } by *temporal-projection* defeaters: { 87 , 55 } DEFEATED
This discharges interest 1
41
interest: (((the position of b1 is (5.0 3.0)) at 10) ⊗ ((the position of b1 is (5.0 3.0)) at 20))
Of interest as a defeater for support-link 52 for node 51
42
interest: ~((the position of b1 is (5.0 3.0)) at 20)
Of interest as a defeater for support-link 52 for node 51
52
~((the position of b1 is (5.0 3.0)) at 10) DEFEATED
Inferred by:
support-link #53 from { 27 } by *position-incompatibility* DEFEATED
53
~((the position of b1 is (0.0 3.0)) at 10)
Inferred by:
support-link #54 from { 50 } by inversion_from_contradictory_nodes_52_and_50
defeatees: { link 27 for node 27 }
54
~((the velocity of b1 is (0.5 0.0)) throughout (0 , 10]) DEFEATED
Inferred by:
support-link #55 from { 27 } by inversion_from_contradictory_nodes_53_and_27 DEFEATED
defeatees: { link 46 for node 45 }
vvvvvvvvvvvvvvvvvvvvvvvvvvv
#<Node 52> has become defeated.
vvvvvvvvvvvvvvvvvvvvvvvvvvv
#<Node 54> has become defeated.
vvvvvvvvvvvvvvvvvvvvvvvvvvv
#<Node 27> has become defeated.
vvvvvvvvvvvvvvvvvvvvvvvvvvv
The undefeated-degree-of-support of #<Node 17> has decreased to 0.960
vvvvvvvvvvvvvvvvvvvvvvvvvvv
==========
Justified belief in ((the position of b1 is (5.0 3.0)) at 20)
with undefeated-degree-of-support 0.980
answers #<Query 2: (? x)(? y)((the position of b1 is (x y)) at 20)>
==========
55
~((the position of b1 is (5.0 3.0)) at 20)
Inferred by:
support-link #56 from { 17 } by *position-incompatibility* DEFEATED
defeatees: { link 52 for node 51 }
56
~((the position of b1 is (0.0 3.0)) at 20)
Inferred by:
support-link #57 from { 51 } by inversion_from_contradictory_nodes_55_and_51 DEFEATED
defeatees: { link 17 for node 17 , link 28 for node 17 }
vvvvvvvvvvvvvvvvvvvvvvvvvvv
#<Node 55> has become defeated.
vvvvvvvvvvvvvvvvvvvvvvvvvvv
#<Node 17> has become defeated.
vvvvvvvvvvvvvvvvvvvvvvvvvvv
==========
Lowering the undefeated-degree-of-support of ((the position of b1 is (0.0 3.0)) at 20)

retracts the previous answer to #<Query 2: (? x)(? y)((the position of b1 is (x y)) at 20)>
==========

43
interest: (∃vx)(∃vy)(((the velocity of b2 is (vx vy)) throughout (0 , 10]) & ((5.0 = (1.0 + (vx * (10 - 0)))) & (3.0 = (0.0 + (vy * (10 - 0))))))
For interest 39 by *new-position*
This interest is discharged by node 60
44
interest: (∃vy)(((the velocity of b1 is (y25 vy)) throughout (10 , 20]) & ((y1 = (5.0 + (y25 * (20 - 10)))) & (y0 = (3.0 + (vy * (20 - 10))))))
For interest 40 by EG
This interest is discharged by node 71
45
interest: (∃vy)(((the velocity of b2 is (y26 vy)) throughout (0 , 10]) & ((5.0 = (1.0 + (y26 * (10 - 0)))) & (3.0 = (0.0 + (vy * (10 - 0))))))
For interest 43 by EG
This interest is discharged by node 59
46
interest: (((the velocity of b1 is (y25 y27)) throughout (10 , 20]) & ((y1 = (5.0 + (y25 * (20 - 10)))) & (y0 = (3.0 + (y27 * (20 - 10))))))
For interest 44 by EG
This interest is discharged by node 70
47
interest: (((the velocity of b2 is (y26 y28)) throughout (0 , 10]) & ((5.0 = (1.0 + (y26 * (10 - 0)))) & (3.0 = (0.0 + (y28 * (10 - 0))))))
For interest 45 by EG
This interest is discharged by node 58
48
interest: ((5.0 = (1.0 + (0.4 * (10 - 0)))) & (3.0 = (0.0 + (0.3 * (10 - 0)))))
For interest 47 by adjunction
This interest is discharged by node 57
49
interest: (5.0 = (1.0 + (0.4 * (10 - 0))))
For interest 48 by adjunction
This interest is discharged by node 12
50
interest: (3.0 = (0.0 + (0.3 * (10 - 0))))
For interest 48 by adjunction
This interest is discharged by node 13

57
((5.0 = (1.0 + (0.4 * (10 - 0)))) & (3.0 = (0.0 + (0.3 * (10 - 0)))))
Inferred by:
support-link #58 from { 12 , 13 } by adjunction
This node is inferred by discharging interest #48
58
(((the velocity of b2 is (0.4 0.3)) throughout (0 , 10]) & ((5.0 = (1.0 + (0.4 * (10 - 0)))) & (3.0 = (0.0 + (0.3 * (10 - 0))))))
Inferred by:
support-link #59 from { 44 , 57 } by adjunction
This node is inferred by discharging interest #47
59
(∃vy)(((the velocity of b2 is (0.4 vy)) throughout (0 , 10]) & ((5.0 = (1.0 + (0.4 * (10 - 0)))) & (3.0 = (0.0 + (vy * (10 - 0))))))
Inferred by:
support-link #60 from { 58 } by EG
This node is inferred by discharging interest #45
60

(∃vx)(∃vy)(((the velocity of b2 is (vx vy)) throughout (0 , 10]) & ((5.0 = (1.0 + (vx * (10 - 0)))) & (3.0 = (0.0 + (vy * (10 - 0))))))
Inferred by:
support-link #61 from { 59 } by EG
This node is inferred by discharging interest #43
61
((the position of b2 is (5.0 3.0)) at 10)
Inferred by:
support-link #62 from { 2 , 60 } by *new-position*
This node is inferred by discharging interest #39
62
(((the position of b1 is (5.0 3.0)) at 10) & ((the position of b2 is (5.0 3.0)) at 10))
Inferred by:
support-link #63 from { 50 , 61 } by adjunction
This node is inferred by discharging interest #10
63
(∃y)(((the position of b1 is (5.0 y)) at 10) & ((the position of b2 is (5.0 y)) at 10))
Inferred by:
support-link #64 from { 62 } by EG
This node is inferred by discharging interest #9
64
(∃x)(∃y)(((the position of b1 is (x y)) at 10) & ((the position of b2 is (x y)) at 10))
Inferred by:
support-link #65 from { 63 } by EG
This node is inferred by discharging interest #3
65
((b1 and b2 collide) at 10)
Inferred by:
support-link #66 from { 64 } by *collision*
This node is inferred by discharging interest #2
==========
Justified belief in ((b1 and b2 collide) at 10)
with undefeated-degree-of-support 0.980
answers #<Query 1: (? ((b1 and b2 collide) at 10))>
==========
51
interest: ((the velocity of b1 is (y25 y27)) throughout (10 , 20])
For interest 46 by adjunction
This interest is discharged by node 68
52
interest: ((the velocity of b2 is (0.4 0.3)) at 10)
For interest 51 by *causal-implication*
This interest is discharged by node 67
67
((the velocity of b2 is (0.4 0.3)) at 10)
Inferred by:
support-link #68 from { 4 } by *temporal-projection*
This discharges interest 52
68
((the velocity of b1 is (0.4 0.3)) throughout (10 , 20])
Inferred by:
support-link #69 from { 42 , 65 , 67 } by *causal-implication* defeaters: { 75 }
This node is inferred by discharging interest #51
59
interest: ~((the velocity of b1 is (0.4 0.3)) throughout (10 , 20])
Of interest as a defeater for support-link 69 for node 68
60

interest: ((y1 = (5.0 + (0.4 * (20 - 10)))) & (y0 = (3.0 + (0.3 * (20 - 10)))))
For interest 46 by adjunction
This interest is discharged by node 69
61
interest: (y1 = (5.0 + (0.4 * (20 - 10))))
For interest 60 by adjunction
This interest is discharged by node 14
62
interest: (y0 = (3.0 + (0.3 * (20 - 10))))
For interest 60 by adjunction
This interest is discharged by node 15

69
((9.0 = (5.0 + (0.4 * (20 - 10)))) & (6.0 = (3.0 + (0.3 * (20 - 10)))))
Inferred by:
support-link #70 from { 14 , 15 } by adjunction
This node is inferred by discharging interest #60
70
(((the velocity of b1 is (0.4 0.3)) throughout (10 , 20]) & ((9.0 = (5.0 + (0.4 * (20 - 10)))) & (6.0 = (3.0 + (0.3 * (20 - 10))))))
Inferred by:
support-link #71 from { 68 , 69 } by adjunction
This node is inferred by discharging interest #46
71
(∃vy)(((the velocity of b1 is (0.4 vy)) throughout (10 , 20]) & ((9.0 = (5.0 + (0.4 * (20 - 10)))) & (6.0 = (3.0 + (vy * (20 - 10))))))
Inferred by:
support-link #72 from { 70 } by EG
This node is inferred by discharging interest #44
72
(∃vx)(∃vy)(((the velocity of b1 is (vx vy)) throughout (10 , 20]) & ((9.0 = (5.0 + (vx * (20 - 10)))) & (6.0 = (3.0 + (vy * (20 - 10))))))
Inferred by:
support-link #73 from { 71 } by EG
This node is inferred by discharging interest #40
73
((the position of b1 is (9.0 6.0)) at 20)
Inferred by:
support-link #74 from { 50 , 72 } by *new-position*
This node is inferred by discharging interest #1

==========

Justified belief in ((the position of b1 is (9.0 6.0)) at 20)
with undefeated-degree-of-support 0.980
answers #<Query 2: (? x)(? y)((the position of b1 is (x y)) at 20)>

==========

74
~((the position of b1 is (9.0 6.0)) at 20) DEFEATED
Inferred by:
support-link #79 from { 17 } by *position-incompatibility* DEFEATED
support-link #75 from { 51 } by *position-incompatibility* DEFEATED
55
~((the position of b1 is (5.0 3.0)) at 20)
Inferred by:
support-link #76 from { 73 } by inversion_from_contradictory_nodes_74_and_73
support-link #56 from { 17 } by *position-incompatibility* DEFEATED
defeatees: { link 52 for node 51 }
54
~((the velocity of b1 is (0.5 0.0)) throughout (0 , 10]) DEFEATED

```
Inferred by:
        support-link #81 from { 68 , 17 } by inversion_from_contradictory_nodes_56_and_17
 DEFEATED
        support-link #77 from { 68 , 51 } by inversion_from_contradictory_nodes_55_and_51
 DEFEATED
        support-link #55 from { 27 } by inversion_from_contradictory_nodes_53_and_27   DEFEATED
defeatees: { link 46 for node 45 }
# 75
~((the velocity of b1 is (0.4 0.3)) throughout (10 , 20])              DEFEATED
Inferred by:
        support-link #82 from { 45 , 17 } by inversion_from_contradictory_nodes_56_and_17
 DEFEATED
        support-link #78 from { 51 , 45 } by inversion_from_contradictory_nodes_54_and_45
 DEFEATED
defeatees: { link 69 for node 68 }
# 74
~((the position of b1 is (9.0 6.0)) at 20)              DEFEATED
Inferred by:
        support-link #79 from { 17 } by *position-incompatibility*   DEFEATED
        support-link #75 from { 51 } by *position-incompatibility*   DEFEATED
# 56
~((the position of b1 is (0.0 3.0)) at 20)
Inferred by:
        support-link #80 from { 73 } by inversion_from_contradictory_nodes_74_and_73
        support-link #57 from { 51 } by inversion_from_contradictory_nodes_55_and_51   DEFEATED
defeatees: { link 17 for node 17 , link 28 for node 17 }
# 54
~((the velocity of b1 is (0.5 0.0)) throughout (0 , 10])              DEFEATED
Inferred by:
        support-link #81 from { 68 , 17 } by inversion_from_contradictory_nodes_56_and_17
 DEFEATED
        support-link #77 from { 68 , 51 } by inversion_from_contradictory_nodes_55_and_51
 DEFEATED
        support-link #55 from { 27 } by inversion_from_contradictory_nodes_53_and_27   DEFEATED
defeatees: { link 46 for node 45 }
# 75
~((the velocity of b1 is (0.4 0.3)) throughout (10 , 20])              DEFEATED
Inferred by:
        support-link #82 from { 45 , 17 } by inversion_from_contradictory_nodes_56_and_17   DEFEATED
        support-link #78 from { 51 , 45 } by inversion_from_contradictory_nodes_54_and_45   DEFEATED
defeatees: { link 69 for node 68 }
        vvvvvvvvvvvvvvvvvvvvvvvvvvvv
        The undefeated-degree-of-support of #<Node 55> has increased to 0.980
        vvvvvvvvvvvvvvvvvvvvvvvvvvvv
        #<Node 74> has become defeated.
        vvvvvvvvvvvvvvvvvvvvvvvvvvvv
        #<Node 75> has become defeated.
        vvvvvvvvvvvvvvvvvvvvvvvvvvvv
        #<Node 51> has become defeated.
        vvvvvvvvvvvvvvvvvvvvvvvvvvvv
        ========================================
        Lowering the undefeated-degree-of-support of ((the position of b1 is (5.0 3.0)) at 20)
        retracts the previous answer to #<Query 2: (? x)(? y)((the position of b1 is (x y)) at 20)>
        ========================================
                # 73
                interest: (((the velocity of b1 is (x5 x6)) at 10) & (~((x5 x6) = (0.0 0.0)) at 10))
                For interest 17 by &-at-intro
```

This interest is discharged by node 85
74
interest: (((the velocity of b1 is (x5 x6)) at 0) & (~((x5 x6) = (0.0 0.0)) at 0))
For interest 8 by &-at-intro
This interest is discharged by node 79
75
interest: ((the velocity of b1 is (x5 x6)) at 0)
For interest 74 by adjunction
This interest is discharged by node 3
76
interest: (~((0.5 0.0) = (0.0 0.0)) at 0)
For interest 74 by adjunction
This interest is discharged by node 78
77
interest: ~((0.5 0.0) = (0.0 0.0))
For interest 76 by pair-nonidentity-at-time
This interest is discharged by node 77

77
~((0.5 0.0) = (0.0 0.0))
Inferred by:
support-link #84 from { } by pair-nonidentity
This discharges interests (77 88)
78
(~((0.5 0.0) = (0.0 0.0)) at 0)
Inferred by:
support-link #85 from { 77 } by pair-nonidentity-at-time
This node is inferred by discharging interest #76
79
(((the velocity of b1 is (0.5 0.0)) at 0) & (~((0.5 0.0) = (0.0 0.0)) at 0))
Inferred by:
support-link #86 from { 3 , 78 } by adjunction
This node is inferred by discharging interest #74
80
(((the velocity of b1 is (0.5 0.0)) & ~((0.5 0.0) = (0.0 0.0))) at 0)
Inferred by:
support-link #87 from { 79 } by &-at-intro
This node is inferred by discharging interest #8
81
(((the position of b1 is (0.0 3.0)) at 0) ⊗ ((the position of b1 is (0.0 3.0)) at 10))
Inferred by:
support-link #88 from { 22 , 1 , 80 } by *causal-undercutter*
defeatees: { link 27 for node 27 }
This node is inferred by discharging interest #13
82
(((the position of b1 is (0.0 3.0)) at 0) ⊗ ((the position of b1 is (0.0 3.0)) at 20)) DEFEATED
Inferred by:
support-link #89 from { 22 , 1 , 80 } by *causal-undercutter*
defeatees: { link 17 for node 17 }
This node is inferred by discharging interests (6 6)
82
interest: ((the velocity of b1 is (x5 x6)) at 10)
For interest 73 by adjunction
This interest is discharged by node 83
83
((the velocity of b1 is (0.5 0.0)) at 10)
Inferred by:
support-link #90 from { 3 } by *temporal-projection*

```
 This discharges interest 82
                              # 85
                              interest: (~((0.5 0.0) = (0.0 0.0)) at 10)
                              For interest 73 by adjunction
                              This interest is discharged by node 84
                              # 88
                              interest: ~((0.5 0.0) = (0.0 0.0))
                              For interest 85 by pair-nonidentity-at-time
                              This interest is discharged by node 77
 # 84
 (~((0.5 0.0) = (0.0 0.0)) at 10)
 Inferred by:
          support-link #93 from { 77 } by pair-nonidentity-at-time
 This node is inferred by discharging interest #85
 # 85
 (((the velocity of b1 is (0.5 0.0)) at 10) & (~((0.5 0.0) = (0.0 0.0)) at 10))
 Inferred by:
          support-link #94 from { 83 , 84 } by adjunction
 This node is inferred by discharging interest #73
 # 86
 (((the velocity of b1 is (0.5 0.0)) & ~((0.5 0.0) = (0.0 0.0))) at 10)
 Inferred by:
          support-link #95 from { 85 } by &-at-intro
 This node is inferred by discharging interest #17
 # 87
 (((the position of b1 is (5.0 3.0)) at 10) ⊗ ((the position of b1 is (5.0 3.0)) at 20))
 Inferred by:
          support-link #96 from { 22 , 50 , 86 } by *causal-undercutter*
 defeatees: { link 52 for node 51 }
 This node is inferred by discharging interest #41
 # 88
 (((the position of b1 is (0.0 3.0)) at 10) ⊗ ((the position of b1 is (0.0 3.0)) at 20))          DEFEATED
 Inferred by:
          support-link #97 from { 22 , 27 , 86 } by *causal-undercutter*   DEFEATED
 defeatees: { link 28 for node 17 }
 This node is inferred by discharging interest #18
 # 82
 (((the position of b1 is (0.0 3.0)) at 0) ⊗ ((the position of b1 is (0.0 3.0)) at 20))          DEFEATED
 Inferred by:
          support-link #98 from { 22 , 27 , 86 } by *causal-undercutter*   DEFEATED
          support-link #89 from { 22 , 1 , 80 } by *causal-undercutter*
 defeatees: { link 17 for node 17 }
 This node is inferred by discharging interests (6 6)
          vvvvvvvvvvvvvvvvvvvvvvvvvvvvv
          #<Node 88> has become defeated.
          vvvvvvvvvvvvvvvvvvvvvvvvvvvvv
          #<Node 82> has become defeated.
          vvvvvvvvvvvvvvvvvvvvvvvvvvvvv
============================================================


================== ULTIMATE EPISTEMIC INTERESTS ==================
 Interest in (? x)(? y)((the position of b1 is (x y)) at 20)
 is answered by node 73:  ((the position of b1 is (9.0 6.0)) at 20)
-----------------------------------------------------
 Interest in (? ((b1 and b2 collide) at 10))
 is answered by node 65:  ((b1 and b2 collide) at 10)
```

Apparently the Extended Prediction Problem poses no new problems for OSCAR's ability to reason causally

13. Conclusions and Comparisons

An agent operating in a complex changing environment must be able to acquire new information perceptually, project those observations forwards in time to draw conclusions about times when observations are not occurring, and reason about how the world will change as a result of changes either observed or wrought by the agent's own behavior. This paper has proposed defeasible reasons that license such reasoning, and described their implementation in OSCAR. PERCEPTION, and the associated principles PERCEPTUAL-RELIABILITY, DISCOUNTED-PERCEPTION, and PERCEPTUAL-UNRELIABILITY, enable an agent to acquire information perceptually while taking account of the fact that perception is less than totally reliable. TEMPORAL-PROJECTION and PROBABILISTIC-DEFEAT-FOR-TEMPORAL-PROJECTION enable an agent to draw conclusions about times when observations are not occurring on the basis of observations at other times. CAUSAL-IMPLICATION and CAUSAL-UNDERCUTTER enable an agent to make use of causal information in forming expectations regarding how the world will evolve over time. The use of TEMPORAL-PROJECTION is pervasive, because observations occur at particular times and making one observation tends to preclude making another, but an agent's reasoning about the world will frequently depend upon the outcomes of a number of distinct observations. This is made more efficient by the introduction of temporal indexicals, which allow an agent to store and recall temporal conclusions rather than making new inferences.

Most work in AI that has been aimed at similar reasoning has proceeded within the framework of the situation calculus.[16] The situation calculus may be quite useful as a semantical tool, but it has always seemed to me to be needlessly complicated for use in actual reasoning. Automated reasoners have as much difficulty as human reasoners in dealing with axiomatizations of domains expressed in the situation calculus. It is noteworthy that human reasoners are able to reason about complex domains using much simpler formalizations. One aim of this paper has been to show that automated reasoners can do so as well.

The formal principles that have been fashioned in this paper were motivated by philosophical analyses, in the style of traditional philosophical epistemology, of human reasoning about perception, time, and causation. A second aim of this paper has been to illustrate the fecundity of such epistemological analysis in addressing problems in artificial intelligence.

The fundamental tool that makes this analysis possible is the OSCAR system of defeasible reasoning. This is the only implemented system of defeasible reasoning capable of reasoning successfully in a language like first-order logic where logical consistency is not decidable. Once again, this system of defeasible reasoning was motivated by an epistemological analysis of human defeasible reasoning. The problems that must be solved in the construction of such a reasoner are logical and computational, but solutions to these problems were suggested by considering how they are solved in human reasoning.[17] I take the success of the analyses of perceptual, temporal, and causal reasoning proposed in this paper to be a strong argument for the use of this system of defeasible reasoning in the design and construction of rational agents.

To the best of my knowledge, there is no work in AI with which to compare the

---

[16] McCarthy and Hayes [1969].

[17] A sketch of these problems and how they are solved in OSCAR is presented in my [1996a]. A more detailed account can be found in my [1995].

analysis of perceptual reasoning proposed here.[18] This is a topic that has not previously received careful attention in AI. On the other hand, there has been a lot of work addressed at temporal projection, causal reasoning, the Frame Problem, and the Yale Shooting Problem. As remarked in section ten, the Frame Problem led Sandewall [1970] to propose reasoning about change defeasibly and adopting some sort of defeasible inference scheme to the effect that it is reasonable to believe that something doesn't change unless you are forced to conclude otherwise. In recent literature, this has come to be called "the commonsense law of inertia". This motivated much of the work in AI on nonmonotonic logic. I was led to a similar principle in my [1974] from a completely different direction—by thinking about the reidentification of objects (see section four above). The principles of temporal projection formulated in this paper are an attempt to make these common ideas precise within the framework of OSCAR.

The basic idea behind the treatment I have proposed for the combination of the Frame Problem and the Yale Shooting Problem is to think of the world as unfolding temporally, with changes occurring only when they are forced to occur by what has already happened. This is also the idea behind Shoham's logic of chronological ignorance [1987], and a series of papers stemming from Gelfond and Lifschitz [1993].[19] There is also a similarity to the notion of progressing a database discussed in Lin and Reiter [1994 and 1995]. Shoham proposed to capture this idea by modifying the logic of defeasible reasoning. The resulting theory was complex, and has never been implemented. The work stemming from Gelfond and Lifschitz [1993] has proceeded by applying circumscription to axiomatizations within (extensions or modifications of) the situation calculus.[20] The resulting theories are unwieldy because of their reliance on the formalism of the situation calculus. But even more important, circumscription generates second-order theories, and so the proposed solutions are in principle not implementable in general. Shanahan [1996] describes an implemented solution to special cases of the Yale Shooting Problem that occur in robot navigation. However, being based upon circumscription, his proposed solution is not implementable in any but special cases. Kartha and Lifschitz [1995] also describe implemented solutions to special cases, based upon circumscription. It is noteworthy that their implementation works by attempting to reduce the circumscription to a set of first-order frame axioms, which is just what the appeal to nonmonotonic logic and defeasible reasoning was originally intended to avoid. It is also worth noting that none of these approaches to temporal projection can solve the perceptual updating problem, because they have no way of decaying the strengths of conclusions. In addition, they fall prey to the projectibility problems discussed in the text. By way of contrast, the proposals made in this paper (1) are based upon a simple formalism, (2) are logically much simpler than the proposals based upon circumscription, and (3) are implemented in general. The current proposals result in reasoning sufficiently simple that we can expect an artificial agent to perform it in real time. For example, the reasoning underlying the solution to the Yale Shooting Problem was performed in .05 seconds on a Macintosh, and that underlying the Extended Prediction Problem (the most complex problem to which these proposals have been applied) required about one second on a Macintosh.

The implementation of the reasoning described here should free automated planning from reliance on STRIPS-like representations of actions. STRIPS operators handle reasoning about change by precompiling the results of all possible actions under all possible circumstances. That works efficiently in narrowly circumscribed domains, but is impractical in most realistic domains. The difficulty has been that the only obvious alternative is to reason explicitly about change, and no one has known how to do that efficiently. The

---

[18] There is a lot of relevant work in philosophical epistemology. The basic ideas of the present analysis originate in my [1967], [1971], and [1974]. Competing philosophical theories are discussed at length in my [1987].

[19] This work includes Kartha and Lifschitz [1995], Shanahan [1995] and [1996],

[20] Examples of such extensions and modifications are the event calculus of Kowalski and Sergot [1986] and the extension of that in Shanahan [1990].

principles described here should take us at least part way along the path to a solution to this problem. That is the subject of current research in the OSCAR Project.