# Step-logic and the Three-wise-men Problem*

**Jennifer J. Elgot-Drapkin**
Department of Computer Science and Engineering
College of Engineering and Applied Sciences
Arizona State University
Tempe, AZ 85287-5406
drapkin@enws92.eas.asu.edu

## Abstract

The kind of resource limitation that is most evident in commonsense reasoners is the passage of time while the reasoner reasons. There is not necessarily any fixed and final set of consequences with which such a reasoning agent ends up. In formalizing commonsense reasoners, then, one must be able to take into account that time is passing as the reasoner is reasoning. The reasoner can then make use of such information in subsequent deductions. *Step-logic* is such a formalism. It was developed in [Elgot-Drapkin, 1988] to model the *on-going process* of deduction. Conclusions are drawn step-by-step. There is no "final" state of reasoning; the emphasis is on intermediate conclusions. In this paper we use step-logic to model the *Three-wise-men Problem.* Although others have formalized this problem, they have ignored the time aspect that is inherent in the problem: a correct assessment of the situation is made by recognizing that the reasoning process takes time and determining that the other wise men *would have concluded such and such by now.* This is an important aspect of the problem that needs to be addressed.

## Background

Commonsense reasoners have limited reasoning capabilities because they must deal with a world about which they have incomplete knowledge. Traditional logics are not suitable for modeling beliefs of commonsense reasoners because they suffer from the problem of logical omniscience: if an agent has $\alpha_1, \ldots, \alpha_n$ in its belief set, and if $\beta$, a wff of the agent's language, is logically entailed by $\alpha_1, \ldots, \alpha_n$, then the agent will also believe $\beta$.

The literature contains a number of approaches to limited reasoning. However, the oversimplification of a "final" state of reasoning is maintained; the limitation amounts to a reduced set of consequences, but all consequences are deduced instantaneously. In contrast, we are interested in the ever-changing set of (tentative) conclusions as the reasoning progresses. Konolige [Konolige, 1984] studies agents with fairly arbitrary rules of inference, but ignores the effort involved in actually performing the deductions. Similarly, Levesque [Levesque, 1984] and Fagin and Halpern [Fagin and Halpern, 1988] provide formal treatments of limited reasoning, but again the conclusions are drawn instantaneously, without making the intermediate steps of reasoning explicit. Lakemeyer [Lakemeyer, 1986] extends Levesque's and Fagin and Halpern's approaches to include quantifiers, but again does not address the issue with which we are concerned. Vardi [Vardi, 1986] deals with limitations on omniscience, again without taking into account the intermediate steps of deduction. Although these approaches all model *limited* reasoning, the process is still in terms of the standard mold of *static* reasoning. We do indeed have a restricted view of what counts as a theorem, but the logic still focuses on the final state of reasoning. The effort involved in actually performing the deductions is not taken into consideration.

We contend that the kind of resource limitation that is most evident in commonsense reasoners is the passage of time while the reasoner reasons. There is not necessarily any fixed and final set of consequences with which such a reasoning agent ends up. In a sense, this is a problem of modeling time. See [Allen, 1984, McDermott, 1982]. Yet these treatments deal with reasoning *about* time, as opposed to reasoning *in* time. Reasoning in time refers to the fact that, as the reasoner reasons, time passes, and this passage of time itself must be recognized by the reasoner. *Step-logic* is proposed as an alternative to the approaches to limited reasoning just discussed, where it is *not* the final set of conclusions in which one is interested, but rather the ever-changing set of conclusions drawn along the way. That is, step-logic is designed to model reasoning that focuses on the *on-going process* of deduction; there is no final state of reasoning.

There are many examples of situations in which the effort or time spent making deductions is crucial. Consider Little Nell who has been tied to the railroad tracks. A train is quickly approaching. Dudley must save her. (See [Haas, 1985, McDermott, 1982].) It is not appropriate for Dudley to spend hours figuring out a plan to save Nell; she will no longer need saving by then. Thus if we are to model Dudley's reasoning, we must have a mechanism that takes into account the passage of time as the agent is reasoning.

The *Three-wise-men Problem* (described in Section ) is

*Our thanks to Don Perlis, Kevin Gary, and Laurie Ihrig for helpful comments.

another example in which the effort involved in making deductions is critical. In this paper we show how step-logic is a useful model for the reasoning involved in this problem. In other formalizations of the *Three-wise-men Problem* this aspect has been ignored. (See [Konolige, 1984, Kraus and Lehmann, 1987, Konolige, 1990].)

## Step-logic

In [Drapkin and Perlis, 1986, Elgot-Drapkin, 1988] we defined a family of eight step-logics—$SL_0, SL_1, \ldots, SL_7$—arranged in increasing sophistication, each designed to model the reasoning of a reasoning agent. Each differs in the capabilities that the agent has. In an $SL_0$ step-logic, for instance, the reasoner has no knowledge of the passage of time as it is reasoning, it cannot introspect on its beliefs, and it is unable to retract former beliefs. ($SL_0$ is not very useful for modeling commonsense reasoners.) In an $SL_7$ step-logic, by contrast, the agent is capable of all three of these aspects that are so critical to commonsense reasoning. Most commonsense reasoners seem to need the full capabilities of an $SL_7$ step-logic.

A step-logic is characterized by a language, observations, and inference rules. We emphasize that step-logic is *deterministic* in that at each step $i$ *all* possible conclusions from one application of the rules of inference applied to the previous steps are drawn (and therefore are among the wffs at step $i$). However, for real-time effectiveness and cognitive plausibility, at each step we want only a finite number of conclusions to be drawn.

Intuitively, we view an agent as an inference mechanism that may be given external inputs or observations. Inferred wffs are called beliefs; these may include certain observations.

Let $\mathcal{L}$ be a first-order or propositional language, and let $\mathcal{W}$ be the set of wffs of $\mathcal{L}$.

**Definition 1** *An* observation-function *is a function* $OBS : \mathbf{N} \rightarrow \mathcal{P}(\mathcal{W})$, *where* $\mathcal{P}(\mathcal{W})$ *is the power set of* $\mathcal{W}$, *and where for each* $i \in \mathbf{N}$, *the set* $OBS(i)$ *is finite.*

**Definition 2** *A* history *is a finite tuple of pairs of finite subsets of* $\mathcal{W}$. $\mathcal{H}$ *is the set of histories.*

**Definition 3** *An* inference-function *is a function* $INF : \mathcal{H} \rightarrow \mathcal{P}(\mathcal{W})$, *where for each* $h \in \mathcal{H}$, $INF(h)$ *is finite.*

Intuitively, a history is a conceivable temporal sequence of belief-set/observation-set pairs. The history is a *finite* tuple; it represents the temporal sequence up to a certain point in time. The inference-function extends the temporal sequence of belief sets by one more step beyond the history.

**Definition 4** *An* $SL_n$-theory over a language $\mathcal{L}$ *is a triple,* $< \mathcal{L}, OBS, INF >$, *where* $\mathcal{L}$ *is a first-order or propositional language,* $OBS$ *is an observation-function, and* $INF$ *is an inference-function. We use the notation,* $SL_n(OBS, INF)$, *for such a theory (the language* $\mathcal{L}$ *is implicit in the definitions of* $OBS$ *and* $INF$*).*

For more background on step-logic, see [Elgot-Drapkin, 1988, Elgot-Drapkin and Perlis, 1990].

## The Problem

We present a variation of this classic problem which was first introduced to the AI literature by McCarthy in [McCarthy, 1978]. This version best illustrates the type of reasoning that is so characteristic of commonsense reasoners.

A king wishes to know whether his three advisors are as wise as they claim to be. Three chairs are lined up, all facing the same direction, with one behind the other. The wise men are instructed to sit down. The wise man in the back (wise man #3) can see the backs of the other two men. The man in the middle (wise man #2) can only see the one wise man in front of him (wise man #1); and the wise man in front (wise man #1) can see neither wise man #3 nor wise man #2. The king informs the wise men that he has three cards, all of which are either black or white, at least one of which is white. He places one card, face up, behind each of the three wise men. Each wise man must determine the color of his own card and announce what it is as soon as he knows. The first to correctly announce the color of his own card will be aptly rewarded. All know that this will happen. The room is silent; then, after several minutes, wise man #1 says "My card is white!".

We assume in this puzzle that the wise men do not lie, that they all have the same reasoning capabilities, and that they can all think at the same speed. We then can postulate that the following reasoning took place. Each wise man knows there is at least one white card. If the cards of wise man #2 and wise man #1 were black, then wise man #3 would have been able to announce immediately that his card was white. They all realize this (they are all truly wise). Since wise man #3 kept silent, either wise man #2's card is white, or wise man #1's is. At this point wise man #2 would be able to determine, if wise man #1's were black, that his card was white. They all realize this. Since wise man #2 also remains silent, wise man #1 knows his card must be white.

It is clear that it is important to be able to reason in the following manner:

> If such and such were true at that time, then so and so *would have realized it by this time.*

So, for instance, if wise man #2 is able to determine that wise man #3 would have already been able to figure out that wise man #3's card is white, and wise man #2 has heard nothing, then wise man #2 knows that wise man #3 does *not* know the color of his card. Step-logic is particularly well-suited to this type of deduction since it focuses on the actual individual deductive steps. Others have studied this problem (e.g. see [Konolige, 1984, Kraus and Lehmann, 1987, Konolige, 1990]) from the perspective of a final state of reasoning, and thus are not able to address this temporal aspect of the problem: assessing what others have been able to conclude *so far*. Elgot-Drapkin [Elgot-Drapkin, 1991a] provides a solution based on step-logic to a version of this problem in which there are only two men.

$OBS_{W_3}$ is defined as follows.

$$OBS_{W_3}(i) = \begin{cases} \left\{ \begin{array}{l} (\forall j)K_2(j, (\forall i)(\forall x)(\forall y)[K_3(i, x \rightarrow y) \rightarrow \\ \qquad (K_3(i, x) \rightarrow K_3(s(i), y))]) \\ (\forall j)K_2(j, K_3(s(0), (B_1 \wedge B_2) \rightarrow W_3)) \\ (\forall j)K_2(j, (B_1 \wedge B_2) \rightarrow K_3(s(0), B_1 \wedge B_2)) \\ (\forall j)K_2(j, \neg(B_1 \wedge B_2) \rightarrow (B_1 \rightarrow W_2)) \\ (\forall j)K_2(j, (\forall i)[\neg U(s(i), W_3) \rightarrow \neg K_3(i, W_3)]) \\ (\forall i)(\forall x)[\neg K_1(s(i), U(i, x)) \rightarrow \neg U(i, x)] \\ (\forall i)[\neg U(i, W_3) \rightarrow K_2(s(i), \neg U(i, W_3))] \\ (\forall i)(\forall x)(\forall y)[K_2(i, x \rightarrow y) \rightarrow (K_2(i, x) \rightarrow K_2(s(i), y))] \\ (\forall i)(\forall x)(\forall x')(\forall y)(\forall y') \\ \quad [(K_2(i, \neg(x \wedge x') \rightarrow (y \wedge y')) \wedge K_2(i, \neg(x \wedge x'))) \rightarrow \\ \quad K_2(s(i), y \wedge y')] \\ (\forall j)(\forall k)(\forall z)(\forall z')(\forall w) \\ \quad [(K_2(j, (\forall i)(\forall x)(\forall y)[K_3(i, x \rightarrow y) \rightarrow \\ \qquad (K_3(i, x) \rightarrow K_3(s(i), y))]) \wedge \\ \quad K_2(j, K_3(k, (z \wedge z') \rightarrow w))) \rightarrow \\ \qquad K_2(s(j), K_3(k, z \wedge z') \rightarrow K_3(s(k), w))] \\ (\forall j)(\forall k) \\ \quad [(K_2(j, (\forall i)[\neg U(s(i), W_3) \rightarrow \neg K_3(i, W_3)]) \wedge \\ \quad K_2(j, \neg U(s(k), W_3))) \rightarrow K_2(s(j), \neg K_3(k, W_3)] \\ (\forall i)(\forall x)(\forall y)[(K_2(i, x \rightarrow y) \wedge K_2(i, \neg y)) \rightarrow K_2(s(i), \neg x)] \\ (\forall i)(\forall x)(\forall x')(\forall y) \\ \quad [(K_2(i, (x \wedge x') \rightarrow y) \wedge K_2(i, \neg y)) \rightarrow K_2(s(i), \neg(x \wedge x'))] \\ (\forall i)[B_1 \rightarrow K_2(i, B_1)] \\ (\neg B_1 \rightarrow W_1) \\ (\forall i)[\neg U(s(i), W_2) \rightarrow \neg K_2(i, W_2)] \end{array} \right\} & \text{if } i = 1 \\ \emptyset & \text{otherwise} \end{cases}$$

Figure 1: $OBS_{W_3}$ for the Three-wise-men Problem

## Formulation

The step-logic used to model the *Three-wise-men problem* is defined in Figures 1 and 2. The problem is modeled from wise man #1's point of view. The observation-function contains all the axioms that wise man #1 needs to solve the problem, and the inference-function provides the allowable rules of inference.

We use an $SL_5$ theory. An $SL_5$ theory gives the reasoner knowledge of its own beliefs as well as knowledge of the passage of time.[1] The language of $SL_5$ is first-order, having binary predicate symbols $K_j$ and $U$, and function symbol $s$. $K_j(i, \text{`}\alpha\text{'})$ expresses the fact that "$\alpha$ is known[2] by agent $j$ at step $i$". Note that this gives the agent the expressive power to introspect on his own beliefs as well as the beliefs of others. $U(i, \text{`}x\text{'})$ expresses the fact that an utterance of $x$ is made at step $i$.[3] $s(i)$ is the successor function (where $s^k(0)$ is used as an abbreviation for $\underbrace{s(s(\cdots(s}_{k}(0))\cdots))$ ). $W_i$ and $B_i$ express the facts that $i$'s card is white, and $i$'s card is black, respectively.

In the particular version of step-logic that is used, the formulas that the agent has at step $i$ (the *i-theorems*) are precisely all those that can be deduced from step $i-1$ using one application of the applicable rules of inference. As previously stated, the agent is to have only a finite number of theorems (conclusions, beliefs, or simply wffs) at any given step. We write:

$$\begin{array}{rl} i: & \ldots, \alpha \\ i+1: & \ldots, \beta \end{array}$$

to mean that $\alpha$ is an $i$-theorem, and $\beta$ is an $i+1$-theorem. There is no implicit assumption that $\alpha$ is present at step $i+1$. That is, wffs are not assumed to be inherited or retained in passing from one step to the next, unless explicitly stated in an inference rule. Rule 8 in Figure 2, however, does provide an unrestricted form of inheritance.[4]

We note several points about the axioms which wise man #1 requires. (Refer to Figure 1.) Wise man #1 knows the following:

[1] For more details on $SL_n$ theories, see [Drapkin and Perlis, 1986, Elgot-Drapkin, 1988].

[2] known, believed, or concluded. The distinctions between these (see [Gettier, 1963, Perlis, 1986, Perlis, 1988]) are not addressed here.

[3] For simplicity, in the remainder of the paper we drop the quotes around the second argument of predicates $U$ and $K_j$.

[4] Although many commonsense reasoning problems require former conclusions to be withdrawn (based on new evidence), this particular formulation of the *Three-wise-men Problem* does not require any conclusions to be retracted. We can thus use an unrestricted form of inheritance.

The inference rules given here correspond to an inference-function, $INF_{W_3}$. For any given history, $INF_{W_3}$ returns the set of all immediate consequences of Rules 1–8 applied to the last step in that history.

**Rule 1:** $\dfrac{i:\ldots}{i+1:\ldots,\alpha}$ if $\alpha \in OBS(i+1)$

**Rule 2:** $\dfrac{i:\ldots,\alpha,(\alpha \rightarrow \beta)}{i+1:\ldots,\beta}$ Modus ponens

**Rule 3:** $\dfrac{i:\ldots,P_1\overline{a},\ldots,P_n\overline{a},(\forall \overline{x})[(P_1\overline{x} \wedge \ldots \wedge P_n\overline{x}) \rightarrow Q\overline{x}]}{i+1:\ldots,Q\overline{a}}$ Extended modus ponens

**Rule 4:** $\dfrac{i:\ldots,\neg\beta,(\alpha \rightarrow \beta)}{i+1:\ldots,\neg\alpha}$ Modus tolens

**Rule 5:** $\dfrac{i:\ldots,\neg Q\overline{a},(\forall \overline{x})(P\overline{x} \rightarrow Q\overline{x})}{i+1:\ldots,\neg P\overline{a}}$ Extended modus tolens

**Rule 6:** $\dfrac{i:\ldots}{i+1:\ldots,\neg K_1(s^i(0),U(s^{i-1}(0),W_j))}$ if $U(s^{i-1}(0),W_j) \notin \vdash_i$, $j=2,3,\ i>1$

**Rule 7:** $\dfrac{i:\ldots,(\forall j)K_2(j,\alpha)}{i+1:\ldots,K_2(s^i(0),\alpha)}$ Instantiation

**Rule 8:** $\dfrac{i:\ldots,\alpha}{i+1:\ldots,\alpha}$ Inheritance

Figure 2: $INF_{W_3}$ for the Three-wise-men Problem

1. Wise man #2 knows (at every step) that wise man #3 uses the rule of *modus ponens.*
2. Wise man #2 uses the rules of *modus ponens* and *modus tolens.*
3. Wise man #2 knows (at every step) that if both my card and his card are black, then wise man #3 would know this fact at step 1.
4. Wise man #2 knows (at every step) that if it's not the case that both my card and his are black, then if mine is black, then his is white.[5]
5. Wise man #2 knows (at every step) that if there's no utterance of $W_3$ at a given step, then wise man #3 did not know $W_3$ at the previous step. (Wise man #2 knows (at every step) that there will be an utterance of $W_3$ the step after wise man #3 has proven that his card is white.)
6. If I don't know about a given utterance, then it has not been made at the previous step.
7. If there's no utterance of $W_3$ at a given step, then wise man #2 will know this at the next step.[6]
8. If my card is black, then wise man #2 knows this (at every step).
9. If there is no utterance of $W_2$ at a given step, then wise man #2 doesn't know at the previous step that his card is white. (There would be an utterance of $W_2$ the step after wise man #2 knows his card is white.)

Note the following concerning the inference rules:

1. Rule 6 is a rule of introspection. Wise man #1 can introspect on what utterances have been made.[7]
2. The rule for extended *modus ponens* allows an arbitrary number of variables.
3. Rule 7 is a rule of instantiation. If wise man #1 knows that wise man #2 knows $\alpha$ at *each* step then, in particular, wise man #1 will know at step $i+1$ that wise man #2 knew $\alpha$ at step $i$.
4. The rule of inheritance is quite general: *everything* is inherited from one step to the next.[8]

[5]In other words, if wise man #2 knows that at least one of our cards is white, then my card being black would mean that his is white. Indeed, this axiom gives wise man #2 quite a bit of information, perhaps too much. (He should be able to deduce some of this himself.) This is discussed in more detail in [Elgot-Drapkin, 1988, Elgot-Drapkin, 1991b].

[6]Interestingly, it is not necessary for wise man #1 to know there was no utterance; wise man #1 only needs to know that wise man #2 will know there was no utterance.

[7]We limit the number of wffs on which the agent can introspect in order to keep the set of beliefs at any given step finite.

[8]For other commonsense reasoning problems, a far more restrictive version of inheritance is necessary.

## Solution

The solution to the problem is given in Figure 3. The step number is listed on the left. The reason (inference rule used) for each deduction is listed on the right. To allow for ease of reading, only the wffs in which we are interested are shown at each step. In addition, none of the inherited wffs are shown. This means that a rule appears to be operating on a step other than the previous one; the wffs involved have, in fact, actually been inherited to the appropriate step.

In step 1 all the initial axioms ($OBS_{W_3}(1)$) have been inferred through the use of Rule 1.[9] Nothing of interest is inferred in steps 2 through 4. In step 5, wise man #1 is able to negatively introspect and determine that no utterance of $W_3$ was made at step 3. Note the time delay: wise man #1 is able to prove *at step 5* that he did not know *at step 4* of an utterance made *at step 3*.[10] The remaining wffs shown in step 5 were all inferred through the use of Rule 7, the rule of instantiation. Wise man #1 needs to know that wise man #2 knows these particular facts at step 4. The reasoning continues from step to step. Note that at step 11, wise man #1 has been able to deduce that wise man #2 knows that if wise man #1's card is black, then his is white. From this step on, we essentially have the *Two-wise-men problem*. (See [Elgot-Drapkin, 1991a].) In step 17 wise man #1 is finally able to deduce that his card is white.

We see that step-logic is a useful vehicle for formulating and solving a problem of this kind in which the time that something occurs is important. Wise man #1 does indeed determine "if wise man #2 or wise man #3 knew the color of his card, he would have announced it by now." Wise man #1 then reasons backwards from here to determine that his card must not be black, and hence must be white.

Several points of contrast can be drawn between this version and the two-wise-men version.

1. In the two-wise-men version, wise man #1 needs only to know about a *single* rule of inference used by wise man #2. In this version wise man #1 needs to know *several* rules used by wise man #2: *modus ponens*, extended *modus ponens*, and *modus tolens*. Because wise man #1 reasons within first-order logic, these three rules required the use of six axioms.
2. In the two-wise-men version, it is sufficient for wise man #1 to know that wise man #2 has certain beliefs *at step 1*. In the three-wise-men version, this is not sufficient—wise man #1 must know that wise man #2 *always* holds these beliefs.
3. What wise man #2 needs to know about wise man #3 is analogous to what wise man #1 needs to know about wise man #2 in the two-wise-men version. So, for instance, wise man #2 must know that wise man #3 uses the rule of *modus ponens* (and this is the only rule of wise man #3's about which wise man #2 must know). Also wise man #2 needs only to know that wise man #3 has certain beliefs *at step 1*.

Many formulations of the *Three-wise-men problem* have involved the use of common knowledge or common belief (see [Konolige, 1984] and [Kraus and Lehmann, 1987] in particular). For instance, a possible axiom might be $C(W_1 \vee W_2 \vee W_3)$: it is common knowledge that at least one card is white. Adding the common knowledge concept here introduces unnecessary complications due, to a large degree, to the fact that the problem is modeled *from wise man #1's point of view*, rather than using a meta-language that describes the reasoning of all three (as [Konolige, 1984, Kraus and Lehmann, 1987] have both done). This is more in the spirit of step-logics, where the idea is to allow the reasoner itself enough power (with no outside "oracle" intervention) to solve the problem. Thus we model the agent directly, rather than using a meta-theory as a model.

## Conclusions

We have shown that step-logic is a powerful formalism for modeling the on-going process of deduction. There is no final state of reasoning; it is the intermediate steps in the reasoning process that are of importance. We have given a solution using step-logic to the *Three-wise-men problem*. Although others have formalized this problem, they have ignored the time aspect that we feel is so critical. In order to correctly assess the situation, one must be able to recognize that the reasoning process itself takes time to complete. Before wise man #1 can deduce that his card is white, he must know that wise men #2 and #3 would have deduced *by now* the color of their cards.

## References


[Allen, 1984] J. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23:123–154, 1984.

[Drapkin and Perlis, 1986] J. Drapkin and D. Perlis. Step-logics: An alternative approach to limited reasoning. In *Proceedings of the European Conf. on Artificial Intelligence*, pages 160–163, 1986. Brighton, England.

[Elgot-Drapkin and Perlis, 1990] J. Elgot-Drapkin and D. Perlis. Reasoning situated in time I: Basic concepts. *Journal of Experimental and Theoretical Artificial Intelligence*, 2(1):75–98, 1990.

[Elgot-Drapkin, 1988] J. Elgot-Drapkin. *Step-logic: Reasoning Situated in Time*. PhD thesis, Department of Computer Science, University of Maryland, College Park, Maryland, 1988.

[Elgot-Drapkin, 1991a] J. Elgot-Drapkin. A real-time solution to the wise-men problem. In *Proceedings of the AAAI Spring Symposium on Logical Formalizations of Commonsense Reasoning*, 1991. Stanford, CA.

[Elgot-Drapkin, 1991b] J. Elgot-Drapkin. Reasoning situated in time II: The three-wise-men problem. Forthcoming, 1991.


[9] To save space we have not repeated them in the figure. See Figure 1 for the individual axioms.

[10] For a detailed description of this phenomenon, see [Elgot-Drapkin, 1988].

| | | | |
|---|---|---|---|
| 0: | | $\emptyset$ | |
| 1: | (a)–(p) | All wffs in $OBS_{W_3}(1)$ | (R1) |
| 2: | | (no new deductions of interest) | |
| 3: | | (no new deductions of interest) | |
| 4: | | (no new deductions of interest) | |
| 5: | (a) | $\neg K_1(s^4(0), U(s^3(0), W_3))$ | (R6) |
| | (b) | $K_2(s^4(0), (\forall i)(\forall x)(\forall y)$ $[K_3(i, x \rightarrow y) \rightarrow (K_3(i, x) \rightarrow K_3(s(i), y))])$ | (R7,1a) |
| | (c) | $K_2(s^4(0), K_3(s(0), (B_1 \wedge B_2) \rightarrow W_3))$ | (R7,1b) |
| | (d) | $K_2(s^4(0), (\forall i)[\neg U(s(i), W_3) \rightarrow \neg K_3(i, W_3)])$ | (R7,1e) |
| 6: | (a) | $\neg U(s^3(0), W_3)$ | (R3,5a,1f) |
| | (b) | $K_2(s^5(0), K_3(s(0), B_1 \wedge B_2) \rightarrow K_3(s^2(0), W_3))$ | (R3,5b,5c,1j) |
| 7: | (a) | $K_2(s^4(0), \neg U(s^3(0), W_3))$ | (R3,6a,1g) |
| | (b) | $K_2(s^6(0), (B_1 \wedge B_2) \rightarrow K_3(s(0), B_1 \wedge B_2))$ | (R7,1c) |
| 8: | (a) | $K_2(s^5(0), \neg K_3(s^2(0), W_3))$ | (R3,7a,5d,1k) |
| | (b) | $K_2(s^7(0), \neg(B_1 \wedge B_2) \rightarrow (B_1 \rightarrow W_2))$ | (R7,1d) |
| 9: | | $K_2(s^6(0), \neg K_3(s(0), B_1 \wedge B_2))$ | (R3,8a,6b,1l) |
| 10: | | $K_2(s^7(0), \neg(B_1 \wedge B_2))$ | (R3,9,7b,1m) |
| 11: | | $K_2(s^8(0), B_1 \rightarrow W_2)$ | (R3,10,8b,1i) |
| 12: | (a) | $(K_2(s^8(0), B_1) \rightarrow K_2(s^9(0), W_2))$ | (R3,11,1h) |
| | (b) | $\neg K_1(s^{11}(0), U(s^{10}(0), W_2))$ | (R6) |
| 13: | | $\neg U(s^{10}(0), W_2)$ | (R3,12b,1f) |
| 14: | | $\neg K_2(s^9(0), W_2)$ | (R3,13,1p) |
| 15: | | $\neg K_2(s^8(0), B_1)$ | (R4,14,12a) |
| 16: | | $\neg B_1$ | (R5,15,1n) |
| 17: | | $W_1$ | (R2,16,1o) |

Figure 3: Solution to the Three-wise-men Problem


[Fagin and Halpern, 1988] R. Fagin and Y. Halpern, J. Belief, awareness, and limited reasoning. *Artificial Intelligence*, 34(1):39–76, 1988.

[Gettier, 1963] E. Gettier. Is justified true belief knowledge? *Analysis*, 23:121–123, 1963.

[Ginsberg, 1991] M. Ginsberg. The computational value of nonmonotonic reasoning. In *Proceedings of the Second International Conference on Principles of Knowledge Representation and Reasoning*, April 1991.

[Haas, 1985] A. Haas. Possible events, actual events, and robots. *Computational Intelligence*, 1(2):59–70, 1985.

[Konolige, 1984] K. Konolige. Belief and incompleteness. Technical Report 319, SRI International, 1984.

[Konolige, 1990] K. Konolige. Explanatory belief ascription. In R. Parikh, editor, *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Third Conference*, pages 85–96. Morgan Kaufmann, 1990. Pacific Grove, CA.

[Kraus and Lehmann, 1987] S. Kraus and D. Lehmann. Knowledge, belief and time. Technical Report 87-4, Department of Computer Science, Hebrew University, Jerusalem 91904, Israel, April 1987.

[Lakemeyer, 1986] G. Lakemeyer. Steps towards a first-order logic of explicit and implicit belief. In J. Halpern, editor, *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 325–340. Morgan Kaufmann, 1986. Monterey, CA.

[Levesque, 1984] H. Levesque. A logic of implicit and explicit belief. In *Proceedings of the 3rd National Conf. on Artificial Intelligence*, pages 198–202, 1984. Austin, TX.

[McCarthy, 1978] J. McCarthy. Formalization of two puzzles involving knowledge. Unpublished note, Stanford University, 1978.

[McDermott, 1982] D. McDermott. A temporal logic for reasoning about processes and plans. *Cognitive Science*, 6:101–155, 1982.

[Perlis, 1986] D. Perlis. On the consistency of commonsense reasoning. *Computational Intelligence*, 2:180–190, 1986.

[Perlis, 1988] D. Perlis. Languages with self reference II: Knowledge, belief, and modality. *Artificial Intelligence*, 34:179–212, 1988.

[Vardi, 1986] M. Vardi. On epistemic logic and logical omniscience. In J. Halpern, editor, *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning about Knowledge*, pages 293–305. Morgan Kaufmann, 1986. Monterey, CA.