Chapter 5: Perception and Belief

This chapter proves the first part of the third hypothesis and provides arguments in support of the fourth hypothesis of this thesis:

> *Third Hypothesis*: This new version of situation theory and the associated theorem prover is appropriate as a knowledge representation and reasoning system for theories of perception and belief.
>
> *Fourth Hypothesis*: Theories of perception and belief as defined by their embeddings in the new version of situation theory provide a better account of human reasoning than classical logic-based computational approaches to perception and belief.

Situation theoretic belief and perception theories are developed in this chapter, proving the first part of the third hypothesis. Automated reasoning in these theories (the second part of the third hypothesis) is addressed in the next chapter. The examination of the application of these theories to example problems provides the arguments in support of the fourth hypothesis.

Many problems in reasoning require both theories of perception and of belief; one wants to reason about what someone *believes* based on what that person is presumed to have *perceived*. The two example problems discussed in the section of this chapter dealing with belief both involve assumptions about what has been perceived and what someone believes as a result of this perception. In the following presentation, perception is discussed first and a situation theoretic approach to it is developed and contrasted with other major approaches. Following this is a more extensive treatment of belief, in which a situation theoretic belief theory is developed and two example problems are investigated.

A Logic of Perception

Perception and reports of perception pose several problems for a formal account of their “logic”. Barwise presents a proposal for several “principles” for a logic of perception and shows how these follow as theorems of situation theory, but present dif-

Let φ be some NI sentence, ~φ is the verb-phrase negation of φ, φ(t) is an NI sentence φ with a constituent (verb or noun) 't', φ(a) and φ(a/b) are the same except all occurrences of 'a' in φ(a) are replaced by 'b' in φ(a/b):

Perception Principle 1:	*Veridicality.*	If *a* sees φ, then φ.
Perception Principle 2:	*Substitutivity.*	If *a* sees $\varphi(t_1)$ and $t_1 = t_2$ then *a* sees $\varphi(t_1/t_2)$.
Perception Principle 3:	*Existential Scope*.	From "*a* sees some *x* such that φ(*x*)" one can derive "there is an *x* such that *a* sees φ(*x*)."
Perception Principle 4:	*Negation.*	If *a* sees ~ φ, then ~(*a* sees φ).
Perception Principle 5:	*Disjunction.*	If *a* sees (φ or ψ) then *a* sees φ or *a* sees ψ.
Perception Principle 6:	*Conjunction.*	If *a* sees (φ and ψ) then *a* sees φ and *a* sees ψ.
Perception Principle 7:	*Logical Equivalence Substitutivity.*	If φ and ψ are logically equivalent, then if *a* sees φ then *a* sees ψ.

Exhibit 5. 1: Principles of Perception.

ficulties for more traditional logical accounts.[1] The logic of naked infinitive perception statements (generally referred to here as simply "the logic of perception") is explored here as an "application" of the formalism developed in the previous chapters.

This discussion is limited to analyzing the meaning of a particular limited class of perception statements, those involving naked infinitives (NI statements). In NI perception sentences, "see" is not followed immediately by the word 'that', and the verb of the "perceived" embedded sentence is in its naked infinitive form. In the following discussion, only NI perception sentences are used as the "perceived" sentence.

The principles which Barwise presents and for which he argues are in Exhibit 5. 1 on page 138.

Barwise presents three non-situation theoretic "seemingly plausible" semantic accounts of NI perception statements, a situation theoretic account, and four linguistic puzzles by which he demonstrates the inadequacy of the non-situation theoretic accounts. The non-situation theoretic accounts are "naive realist logic of perception", "propositional theories of perception", and "naive adverbial theories of perception and ad hoc semantics".

1. pp. 12-15 in [Barwise 1981]. All of [Barwise 1981] is more-or-less devoted to "the logic of NI [naked infinitive] perception statements".

Naive realism theory of perception

In the naive[2] realist approach, the idea is that perception is "a direct confrontation between the perceiver *a* and the perceived object *b*; say '*a* sees *b*'."[3] In this approach, there is no direct way of representing the perception of an event - only that some object has the property of participating in a kind of event. The example is:

Whitehead saw Russell wink

which the naive-realist must represent in a manner similar to:

$(wSr) \wedge T(r)$

where $T(x)$ means "*x* has the property of winking" and xSy means "*x* saw *y*". This formula is not a satisfying expression of the meaning of the sentence, since it is also the translation of:

Whitehead saw Russell and Russell winked

where no notion that Whitehead saw Russell's wink is expressed. The situation theoretic account avoids this problem by having a way to speak of events directly (the situation in which the event occurred).

Propositional theory of perception

The propositional theories of perception are based on a different theory about the act of perception. The idea here is that one never sees an object directly, but rather sees that the object has some property: "... we never simply see a tomato, say, but rather we see that something is a tomato, or that something is red and roundish. ...seeing *is* a way of knowing or believing."[4] Thus, the objects of seeing are *propositions*. Barwise claims that this is the direction taken by Hintikka, Thomason and Niiniluoto, leading to a possible-worlds theory of perception. However, this approach doesn't fit with the principles of perception.

The example which Barwise gives shows that a "modal" argument produces an obvi-

2. Barwise uses the term "naive" to contrast with his own approach. He characterizes his approach as realist, but not (as) naive.
3. p. 21 in [Barwise 1981].
4. p. 22 in [Barwise 1981].

ously incorrect conclusion. In his example, Barwise identifies Perception Principle 7, logical equivalence, as providing "the false step" in the modal proof (specifically, the assumption that "*p* or not *p*" is always true and then extending a formula into a conjunction with another formula of this form and claiming that the extended formula is logically equivalent to the initial formula). A modal logician might argue that the problem is in the Perception Principle 5, disjunction, but this would be an attempt to alter natural language semantics to fit a mathematical system (modal logic). The situation theoretic account avoids this problem by having a more limited notion of logical equivalence. For the example, the relevant limitation is that "*p* or not *p*" is not always "true" (supported by a given situation) in situation theory.

Adverbial theory of perception

The adverbial theory of perception takes the position that the object of perception should be treated as an adverb modifying how the agent is seeing. Thus, *John sees Mary run* is interpreted as meaning John sees in a "Mary run" way. This might be represented formally as

$$S_{\text{Mary run}}(\text{John})$$

where $S_{\text{Mary run}}$ is a predicate symbol. Instead of having one "sees" predicate there is now an infinite number of S_{φ} predicates, one for each sentence φ. This has the opposite problem with logical equivalence to that of the previous approach. If φ and ψ are logically equivalent but syntactically distinct, then S_{φ} and S_{ψ} are distinct predicates. So, no logical equivalence substitutions are allowed in this approach. Certainly this approach won't make inappropriate inferences based on logical equivalence, but it won't make the appropriate inferences either. For example, in viewing two blocks *s* and *t* if *a* sees *s* is on *t*, it should be possible to infer that *a* sees *t* is under *s*. The adverbial approach doesn't support this. The situation theoretic approach doesn't have this problem since it does allow for logical equivalence substitutions. The adverbial approach doesn't justify *any* of the principles A through F presented above.

Situation theoretic theory of perception

The situation theoretic approach is to interpret "*a* sees φ" as an assertion that:

a sees some situation *s* where *s* supports φ.

More formally, the sentence "*a* sees φ" is interpreted as:

$$\forall \boldsymbol{f}\ (\exists\, d \models \mathit{discourse}(\text{“}a \text{ sees } \varphi\text{”})[\boldsymbol{f}\,]) \rightarrow \exists\, s,t\ (t \models \text{sees}(\boldsymbol{L}, a, s) \wedge s \models \mathit{content}(\varphi, \mathit{loc}) \wedge \boldsymbol{L} \text{ temporally\_equal } \mathit{loc})[\boldsymbol{f}\,],$$

where $\boldsymbol{f}$ is an *anchor* for the parameters of the infons in the antecedent and consequent, $\boldsymbol{L}$ is a parameter for the location of the 'sees' infon. $\boldsymbol{L}$ (the location of the described situation) and any other parameters in the consequent must all occur in the *discourse* infon. This formulation makes explicit the three situations of the interpretation of this sentence; the discourse situation *d*, the described situation *t*, and the "seen" situation *s*. This interpretation cannot be stated using the involvement relation and situation types, due to the explicitly referenced "seen" situation *s.*

A Murder: A Puzzle in the Logic of Perception

Barwise discusses four problems relating to perception and how these problems can be formulated in the various approaches mentioned above. One of these problems is examined here. The puzzle involves a murder:

> Bob has killed Fred with a knife. Mary testifies: "Bob and I entered the room at the same time, by different doors. Fred, facing my door, saw me enter. I saw Bob enter, but Fred did not see Bob enter."

This can be summarized by:

m saw B(b)
f saw F(m)
~(f saw B(b))

where B(b) is "Bob entering through the door in back of Fred", F(m) is "Mary entering through the door in front of Fred", 'm' is Mary and 'f' is Fred.

Barwise posits a modal logician K who undertakes to show that Mary's testimony is inconsistent and therefore should be ignored. K accepts the principles A through F, and the principle of logical equivalence substitutivity. K's reasoning is as follows:

1)	m saw B(b)	[given]
2)	f saw F(m)	[given]
3)	B(b)	[step 1 and Perception Principle 1, page 138]
4)	~(f saw ~B(b))	[step 3 and the contrapositive of Perception Principle 1, page 138]
5)	$F(m) \Leftrightarrow ((F(m) \wedge B(b)) \vee (F(m) \wedge \sim B(b)))$	[axioms of FOL]
6)	f saw $((F(m) \wedge B(b)) \vee (F(m) \wedge \sim B(b)))$	[step 2 and 5 and Perception Principle 7, page 138]
7)	(f saw $(F(m) \wedge B(b))) \vee$ (f saw $(F(m) \wedge \sim B(b)))$	[step 6 and Perception Principle 5, page 138]
8)	(f saw $(F(m) \wedge \sim B(b))) \Rightarrow (F(m) \wedge \sim B(b))$	[Perception Principle 1, page 138]
9)	$\sim(F(m) \wedge \sim B(b))$	[step 3 and axioms of FOL]
10)	~(f saw $(F(m) \wedge \sim B(b)))$	[step 9 and contrapositive of Perception Principle 1, page 138]
11)	(f saw $(F(m) \wedge B(b)))$	[step 10 and 7 and axioms of FOL]
12)	f saw B(b)	[step 11, Perception Principle 6, page 138, and and-elimination inference rule of FOL]

This contradicts Mary's testimony that "~(f saw B(b))", so her testimony is inconsistent. Barwise identifies step 6, the use of the logical equivalence, as the false step in this line of reasoning.[5]

The situation theoretic approach translates the problem as:

$s_d \models see(m, s_m) \wedge s_m \models B(b)$
$s_d \models see(f, s_f) \wedge s_f \models F(m)$
$\sim(s_d \models see(f, s_f) \wedge s_f \models B(b))$

where s_d is the overall situation being described, s_m is the situation seen by Mary, and s_f is the situation seen by Fred. The location argument has been eliminated to simplify the presentation.

5. p. 24 in [Barwise 1981].

Many of the steps of K's proof hold for the ST version, except for step 5 introducing the logical equivalence. In ST, F(m) is not logically equivalent to ((F(m) $\wedge$ B(b)) $\vee$ (F(m) $\wedge$ ~B(b))), since B(b) $\vee$ ~B(b) is not supported by all situations. Thus, a situation may support F(m) but not support ((F(m) $\wedge$ B(b)) $\vee$ (F(m) $\wedge$ ~B(b))) if that situation does not determine B(b). By this analysis, Mary's testimony is perfectly consistent - $s_m \models$ B(b), $s_f \models$ F(m), and $s_f \not\models$ B(b).

The "believes" relation

An agent's beliefs are represented by parametrized ST propositions. In [Barwise&Perry 1983] "represented beliefs" were represented via a *schema*.[6] This is translatable into modern ST terms as a (parametrized) situation type defined by a (parametrized) infon, which is a possibly compound (specifically, a disjunctive infon). That is, a belief has the form of " agent A *believes that* there exists situation s_0 such that situation s_0 supports infon P". This limitation to existential support propositions (a situation type) is overly strict — some beliefs are about the infons supported by particular situations. The model of beliefs used here is that the thing believed is a *proposition* rather than a situation type. Thus, the *believes* relation takes 3 arguments, the agent, the location (time) of the belief, and the proposition which defines the contents of the belief: $\langle\langle$believes, *Agent, Location, Belief* $\rangle\rangle$. This is a modal infon - an infon which has an argument which takes propositions (which generally are support relations between situations and infons) as its value.

As a notational convenience, located belief infons are written "*A @ L* bel *B*". This is read as "agent *A* at location *L* believes *B*". Beliefs can be nested. An example of this is $agent_0$ believing that it believes *P*: $s_1 \models agent_0$ @ $\mathbf{l_1}$ bel ($s_1 \models$ ($agent_0$ @ $\mathbf{l_1}$ bel P)). To simplify the following discussion, the location argument is generally suppressed.

Support Postulate 5. 1 *Confirmation of belief* : $s \models \langle\langle$believes, *A*, *L*, *B* $\rangle\rangle$ iff *s* is a situation wherein agent *A* believes at location *L* that there exists some situation *t* such that $t \models B$. To restate this: $s \models A$ @ *L* bel *B* iff situation *s* encompasses both location

6. pp. 241-253 in [Barwise&Perry 1983].

L and agent *A*, and agent *A* at *L* believes *B*.

Support Postulate 5. 2 *Denial of belief* : $s \models \langle\langle \text{believes}, A, L, B ; - \rangle\rangle$ iff *s* is a situation wherein agent *A* at location *L* does not believe that there exists some situation *t* such that $t \models B$. To restate this: $s \models -(A @ L \text{ bel } B)$ iff situation *s* encompasses both location *L* and agent *A*, and agent *A* at *L* does not believe *B*.

Belief principles

There are several basic principles about beliefs. In general, these principles should be assumed to be fallible. They are not necessary constraints. They are more properly considered nomic constraints, akin to laws of (intelligent) nature. However, a simplification adopted here is to assume that the logic of "belief" is defined with all of these principles as axioms. These principles[7] are given in Exhibit 5. 2, page 146.

The principles have been translated into situation theoretic terms, thus the supports relation appears in their statement. The "thing" being believed is represented by a *classical* first order logic formula. To help distinguish between classical formulae and infon formulae, the different conditional operators have been represented using different symbols; '$\rightarrow$' represents the classical conditional and '$\Rightarrow$' represents the infon conditional.

These principles are derived from the classic S5 modal logic axioms. The situation theoretic versions of these principles has a strong difference from the classical modal logic axioms in that the situation theoretic versions of these axioms are *logically independent*, whereas the classical versions are *not* independent. (The situation theoretic *distribution of belief* principle is not independent, but is derived from the *closure* and *knowledge* principles.) Classically, modal axiom T implies modal axiom D. However, the situation theoretic knowledge belief principle does *not* imply the situation theoretic belief consistency principle. The knowledge principle allows one to infer that if situation *T* supports that *A* is believed by *S*, then *T* does not support that

7. p. 36-38 in [Konolige 1986].

S believes the negation (dual) of *A*. The given consistency principle allows a stronger statement to be made, that *T* supports that if *A* is believed by *S*, then that situation supports that the negation of *A* is *not* believed by *S*. Thus, knowledge and consistency are independent principles in this situational belief theory, but they are not independent in classical modal logic.

There are at least three different ways to interpret the introspection axioms in situation theory, the weak, middle, and strong formulations. The middle formulation is given in the table. The strong formulation is similar, but does not use the existential quantification:

$T \models (S \text{ bel } A) \Rightarrow (S \text{ bel } (T \models S \text{ bel } A))$

$T \models \neg(S \text{ bel } A) \Rightarrow (S \text{ bel } (T \models \neg(S \text{ bel } A)))$

The above axioms for introspection are easier to reason with than the ones in Exhibit 5. 2, but they are less plausible. They claim that if S believes A in situation T, then S believes 'S believes A in situation T' in situation T. From the persistence of infons, this introduces all situations of which T is a part as objects about which S has beliefs. This profusion of situations seems unwarranted.

The weak formulation is weaker than that given for introspection in Exhibit 5. 2. In this formulation the quantification is moved into the nested belief:

$T \models (S \text{ bel } A) \Rightarrow (S \text{ bel } \exists U (U \models S \text{ bel } A))$

$T \models \neg(S \text{ bel } A) \Rightarrow (S \text{ bel } \exists U (U \models \neg(S \text{ bel } A)))$

The logical closure principle is interesting in that it relates logical consequence ("derives") between classical propositions to the infon conditional. This is an extension of the deduction theorem for infon logic. The rest of the principles are candidates for axiom *schema* additions to infon logic. If all of the principles are accepted, then the belief operator has the formal properties of an S5 modal logic-like extension of infon logic. These principles are properly considered schemas since they have variables which range over classical propositions, and infon logic variables may only range over terms of infon logic.

Principle	Description	Traditional Modal Axiom
$(A \rightarrow B)$ implies $T \models (S$ **bel** $A \Rightarrow S$ **bel** $B)$	*Logical closure*	**K**
$(T \models (S$ **bel** $A)) \rightarrow A$	*Knowledge*	**T**
$T \models ((S$ **bel** $A \rightarrow B) \Rightarrow (S$ **bel** $A) \Rightarrow S$ **bel** $B)$	*Distribution of belief* (from K and T)	
$T \models ((S$ **bel** $A) \Rightarrow -(S$ **bel** $-A))$	*Consistency*	**D**
$T \models (S$ **bel** $A) \Rightarrow \exists U\,(S$ **bel** $(U \models S$ **bel** $A))$	*Positive Introspection*	**4**
$T \models -(S$ **bel** $A) \Rightarrow \exists U\,(S$ **bel** $(U \models -(S$ **bel** $A)))$	*Negative Introspection*	**5**

T and U are situations. S is an agent. A and B are classical formulae.

Exhibit 5. 2: Belief Principles

A complete theory of belief would include a nonmonotonic theory of belief, where the principles are used as defeasible inference rules. This is not attempted here.

Applying the Theory of Belief

There are two examples which are used to explore the application of the theory of belief given above. The first example is a story about a poker game, where two people who see one or both of the hands come to different conclusions. The challenge is to account for the conclusions at which they arrive. The second example is the two-person version of the “wise men” puzzle. In this puzzle it is common knowledge between two men that at least one of them has a white dot on their forehead and that each can only see the other man’s forehead (not his own). One of them says he doesn’t know if he has a white dot. After hearing this the other one figures out that he, himself, must have a white dot. The challenge here is to provide a line of reasoning for the second wise man. These example problems have been adopted in this work as a benchmark of a minimal ability to deal with multiple agents, perception and belief.

The Poker Game

This example is from Allan Gibbard[8], and is discussed at length by Barwise[9] and Stalnaker[10]:

> Sly Pete and Mr. Stone are playing poker on a Mississippi riverboat. It is now up to Pete to call or fold. My henchman Zack sees Stone's hand, which is quite good, and signals its contents to Pete. My henchman Jack sees both hands, and sees that Pete's hand is rather low, so that Stone's the winning hand. At this point the room is cleared. A few minutes later Zack slips me a note which says "if Pete called, he won," and Jack slips me a note which says "if Pete called, he lost..." I conclude that Pete folded.

This example is introduced by Gibbard to demonstrate that conditional statements (e.g. "if Pete called, he won") do not have any "propositional content". Stalnaker and Barwise continue the discussion of propositional content. Stalnaker modifies Gibbard's position by saying that "open conditionals" (a kind of conditional which Jack and Zack's statements exemplify) do have a propositional content, but it is "highly context dependent". The context to which Stalnaker here refers is that of the speaker and listener. The propositional content which Barwise attributes to nearly *any* kind of sentence is "context dependent" - as interpreted in this thesis it is a claim about an infon being supported by a situation. This approach can be used to represent the conditionals of the example.

The formal analysis pursued here of this example explores some of the ways in situations, perceptions, and beliefs are present in the example, and how these things interact in the reasoning about Jack's and Zack's conclusions. Thus only a limited set of relevant facts are formalized. Also, space and time details are suppressed.

8. Originally from p. 231 of [Gibbard 1981] and discussed on pp. 231-234. This description is as given on p. 112 of [Barwise 1986]. Barwise states that he is using the version as given on pp. 108-109 of [Stalnaker 1984].
9. pp. 112-113 and pp. 131-132 in [Barwise 1986].
10. pp. 108-110 in [Stalnaker 1984].

Defining Terms and Relations

Let *t* be a situation which contains Zack and Jack, *s* the situation of the poker game, which includes Sly Pete and Mr. Stone, '*pete*' be Sly Pete, '*stone*' be Mr. Stone, and 'bel(A, P)' be the belief infon that person A believes proposition P. Let 'calls(A)' mean "A called". Let 'won(A)' mean "A won", and 'loses(A)' mean that "A lost".

Zack's belief that "if Pete called, he won" is represented as:

$t \models \mathbf{bel}(zack, s \models \text{calls}(pete) \Rightarrow \text{wins}(pete))$.

Jack's belief that "if Pete called, he lost" is represented as:

$t \models \mathbf{bel}(jack, s \models \text{calls}(pete) \Rightarrow \text{loses}(pete))$.

Let 'hand(A, H)' mean "A's hand of playing cards is H", 'players(A, B)' mean "A and B are the players in a two-handed game of poker", and 'player(A)' mean "A is a player in a game of poker". Let 'knows_poker(A)' mean "A knows what is common knowledge among poker players (e.g., rules and habits of play)".

Let 'sit(S)' mean "S is a situation". Let s_1 be the part of the poker game situation which Zack sees, which includes Mr. Stone's hand but not Sly Pete's hand. Let s_2 be the part of the poker game situation which Jack sees, which includes both Mr. Stone's hand and Sly Pete's hand but not the event of Zack telling Sly Pete what Mr. Stone's hand is. Situation s_1 is strictly a part of s_2, and s_2 is part of s.

Formalizing the Story

The story is given again below with the formalization of each part of the story placed immediately after that part, parts of the story which have no associated formalization are given in parentheses:

“Sly Pete and Mr. Stone are playing poker on a Mississippi riverboat.”

t |= bel(*zack*, s |= players(*pete*, *stone*))

t |= bel(*jack*, s |= players(*pete*, *stone*))

(“It is now up to Pete to call or fold.”)

“My henchman Zack sees Stone’s hand...”

t |= bel(*zack*, part_of(s_1, s))

t |= sees(*zack*, s_1)

s_1 |= hand(*stone*, *sh*)

(“...which is quite good,...”)

“...and signals its contents to Pete.”

t |= bel(*zack*, s |= bel(*pete*, s |= hand(*stone*, *sh*)))

“My henchman Jack sees both hands,...”

t |= bel(*jack*, part_of(s_2, s))

t |= sees(*jack*, s_2)

s_2 |= hand(*stone*, *sh*) $\wedge$ hand(*pete*, *ph*)

“...and sees that Pete’s hand is rather low, so that Stone’s the winning hand.”

better(*sh*, *ph*)

[Domain Rule 3, see below]

(“At this point the room is cleared. A few minutes later...”)

“Zack slips me a note which says ‘if Pete called, he won,’...”

t |= bel(*zack*, s |= calls(*pete*) $\Rightarrow$ wins(*pete*)).

“...and Jack slips me a note which says ‘if Pete called, he lost...’.”

t |= bel(*jack*, s |= calls(*pete*) $\Rightarrow$ loses(*pete*)).

t |= knows_poker(*zack*)
t |= bel(*zack*, part_of(s_1, s))
t |= sees(*zack*, s_1)
t |= bel(*zack*, s |= bel(*pete*, s |= hand(*stone*, *sh*)))
t |= bel(*zack*, s |= players(*pete*, *stone*))

t |= knows_poker(*jack*)
t |= bel(*jack*, part_of(s_2, s))
t |= sees(*jack*, s_2)
t |= bel(*jack*, s |= players(*pete*, *stone*))

s_1 |= hand(*stone*, *sh*)
s_2 |= hand(*stone*, *sh*) Ÿ hand(*pete*, *ph*)
better(*sh*, *ph*)

Exhibit 5. 3: Poker Game Formalization. Given Facts.

("I conclude that Pete folded.")

These given formulae are presented in Exhibit 5. 4 on page 153.

Formalizing Knowledge About Poker

There are some domain rules about poker which are used to arrive at the conclusions of Zack's and Jack's beliefs. Because these rules involve several quantifiers and nesting of the supports relation and the belief relation, they are hard to read when presented directly. Their presentation is made modular by using named, schematic formulae. The names of these schematic formulae are in **bold** face. The rules and their defined subformulas are:

The major schema for defining these rules is '**everybody_who_knows_poker_believes**(***X***)'. This schema states that for all situations *s* if *s* supports that *a* knows poker, then *a* believes ***X*** in *s*:

everybody_who_knows_poker_believes(***X***) $=_{df}$
$\forall\ s\ (\text{sit}(s) \rightarrow s \models \forall\ a\ (\text{knows\_poker}(a) \Rightarrow \text{bel}(a, \boldsymbol{X})))$

The first rule simply states that everyone knows that if the set of all players in a

poker game is {A, B}, then A is a player in the game and B is a player in the game. This rule could be more clearly stated if there was a representation for sets:

Rule 1: **everybody_who_knows_poker_believes(**

each_person_in_game_is_a_player)

each_person_in_game_is_a_player $=_{df}$

$\forall\, t\,(\text{sit}(t) \rightarrow t \models \forall\, p_1, p_2\,(\text{players}(p_1, p_2) \Rightarrow \text{player}(p_1) \wedge \text{player}(p_2)))$.

The second rule states that everyone knows that every player knows her own hand (the poker game is presumed to be draw poker - the example from Gibbard doesn't say and some other kinds of poker (such as stud poker) wouldn't have this property):

Rule 2: **everybody_who_knows_poker_believes(**
every_player_knows_her_hand)
every_player_knows_her_hand $=_{df}$
$\forall\, t\,(\text{sit}(t) \rightarrow t \models \forall\, p\,(\text{player}(p) \Rightarrow \exists\, x\, \text{bel}(p, t \models \text{hand}(p, x))))$.

The third rule states that everyone knows that if a person knows both of the hands in the game and that the hand for a particular player P is better, then that person knows that if P calls then P wins and if the other player Q calls then Q loses:

Rule 3: **everybody_who_knows_poker_believes(**
knowing_better_hand_implies_knowing_results)

knowing_better_hand_implies_knowing_results $=_{df}$
$\forall\, t, u\,(\text{sit}(t) \wedge \text{sit}(u) \rightarrow$
$\forall\, p, px, py$ **(knows_hand_is_better**$(t, u, p, px, py) \rightarrow$
knows_call_results$(t, u, p, px, py))))))$

knows_hand_is_better(*T, U, P, PX, PY*) $=_{df}$
$(\boldsymbol{U} \models \text{players}(\boldsymbol{PX}, \boldsymbol{PY}) \vee \text{players}(\boldsymbol{PY}, \boldsymbol{PX}))$
$\wedge \exists\, x, y\,(\text{better}(x, y) \wedge t \models$ **knows_both_hands**$(\boldsymbol{U}, \boldsymbol{P}, \boldsymbol{PX}, \boldsymbol{PY}, x, y))$

knows_call_results(*T, U, P, PX, PY*) $=_{df}$

T |= **knows_wins(*U, P, PX*)** $\wedge$ **knows_loses(*U, P, PY*)**

knows_wins(T, P, Q) $=_{df}$ bel(P, T |= (calls(Q) $\Rightarrow$ wins(Q))).

knows_loses(T, P, Q) $=_{df}$ bel(P, T |= (calls(Q) $\Rightarrow$ loses(Q))).

The fourth rule is a weaker version of the third rule. It states that everyone knows that if a person knows both of the hands for the game, then either she knows that if she calls then she wins or she knows that if she calls then she loses. The essential difference between the third and fourth rules is that to use the third rule "everyone" must know how the hands for the player compare, while to use the fourth rule one only needs to know that some person knows both hands without one needing to know what those hands are:

Rule 4: **everybody_who_knows_poker_believes(**
knowing_both_hands_implies_knowing_result).

knowing_both_hands_implies_knowing_result $=_{df}$
$\forall\, t$ (sit(t) $\rightarrow t$ |= $\forall\, p_1,\ p_2$ (players(p_1, p_2) $\vee$ players(p_2, p_1) $\Rightarrow$
($\exists\, x, y$ **knows_both_hands**(t, p_1, p_1, p_2, x, y) $\Rightarrow$
knows_wins(t, p_1, p_1) $\vee$ **knows_loses**(t, p_1, p_1)))).

knows_both_hands(T, P, Q_1, Q_2, X, Y) $=_{df}$ bel(P, T |= hand(Q_1, X) $\wedge$ hand(Q_2, Y)).

The fifth rule states that everyone knows that if a person calls, then it is not the case that she believes that if she calls then she loses:

Rule 5: **everybody_who_knows_poker_believes(**
no_caller_believes_she_will_lose).

no_caller_believes_she_will_lose $=_{df}$
$\forall\ t$ (sit(t) $\rightarrow t$ |= $\forall\, x$ (calls(x) $\Rightarrow$ – **knows_loses**(t, x, x))).

The summary of rules 1 through 3 is presented in Exhibit 5. 4 on page 153, and rules 4 and 5 are presented in Exhibit 5. 5 on page 154.

everybody_who_knows_poker_believes(***X***) $=_{df}$

$\forall\ s\ (\text{sit}(s) \rightarrow s \models \forall\ a\ (\text{knows\_poker}(a) \Rightarrow \text{bel}(a, \boldsymbol{X})))$

Rule 1: **everybody_who_knows_poker_believes**(**each_person_in_game_is_a_player**)

each_person_in_game_is_a_player $=_{df}$

$\forall\ t\ (\text{sit}(t) \rightarrow t \models \forall\ p_1, p_2\ (\text{players}(p_1, p_2) \Rightarrow \text{player}(p_1) \wedge \text{player}(p_2))).$

Rule 2: **everybody_who_knows_poker_believes**(**every_player_knows_her_hand**)

every_player_knows_her_hand $=_{df}$

$\forall\ t\ (\text{sit}(t) \rightarrow t \models \forall\ p\ (\text{player}(p) \Rightarrow \exists\ x\ \text{bel}(p, t \models \text{hand}(p, x)))).$

Rule 3: **everybody_who_knows_poker_believes**(

knowing_better_hand_implies_knowing_results)

knowing_better_hand_implies_knowing_results $=_{df}$

$\forall\ t, u\ (\text{sit}(t) \wedge \text{sit}(u) \rightarrow$

$\forall\ p, px, py$ **(knows_hand_is_better(*t, u, p, px, py*)**

$\rightarrow$ **knows_call_results**(***t, u, p,*** *px, py*)**)))))**

knows_hand_is_better(***T, U, P, PX***, ***PY***) $=_{df}$

(*U* $\models$ **players(*PX, PY*)** $\vee$ **players(*PY*, *PX*))**

$\wedge\ \exists\ x, y\ (\text{better}(x, y) \wedge t \models$ **knows_both_hands**(***U***, ***P***, ***PX***, ***PY***, *x*, *y*))

knows_call_results(***T, U, P, PX***, ***PY***) $=_{df}$

T $\models$ **knows_wins**(***U, P, PX***) $\wedge$ **knows_loses**(***U, P, PY***)

knows_wins(***T***, ***P***, ***Q***) $=_{df}$ $\text{bel}(\boldsymbol{P}, \boldsymbol{T} \models (\text{calls}(\boldsymbol{Q}) \Rightarrow \text{wins}(\boldsymbol{Q})))$.

knows_loses(***T***, ***P***, ***Q***) $=_{df}$ $\text{bel}(\boldsymbol{P}, \boldsymbol{T} \models (\text{calls}(\boldsymbol{Q}) \Rightarrow \text{loses}(\boldsymbol{Q})))$.

Exhibit 5. 4: Poker Game Formalization.
Domain Rules 1, 2, and 3.

Proving the Henchmen's Conclusions

Using this formalization, Zack's and Jack's conclusions can be derived using the formalization given above of the Poker Game, and various principles of perception, belief, and support. These additional principles are:

seeing is believing If someone sees a situation (or "scene"), then they believe the things which that situation supports.

Rule 4: **everybody_who_knows_poker_believes**(
knowing_both_hands_implies_knowing_result).

knowing_both_hands_implies_knowing_result $=_{df}$
$\forall\, t\,(\text{sit}(t) \rightarrow t \models \forall\, p_1,\ p_2\,(\text{players}(p_1, p_2) \vee \text{players}(p_2, p_1) \Rightarrow$
$(\exists\, x, y$ **knows_both_hands**(t, p_1, p_1, p_2, x, y)
$\Rightarrow$ **knows_wins**$(t, p_1, p_1) \vee$ **knows_loses**$(t,\ p_1, p_1))))$.

knows_both_hands$(\boldsymbol{T}, \boldsymbol{P}, \boldsymbol{Q}_1, \boldsymbol{Q}_2, \boldsymbol{X}, \boldsymbol{Y}) =_{df} \text{bel}(\boldsymbol{P}, \boldsymbol{T} \models \text{hand}(\boldsymbol{Q}_1, \boldsymbol{X}) \wedge \text{hand}(\boldsymbol{Q}_2, \boldsymbol{Y}))$.

Rule 5: **everybody_who_knows_poker_believes**(**no_caller_believes_she_will_lose**).

no_caller_believes_she_will_lose $=_{df}$
$\forall\ t\ (\text{sit}(t) \rightarrow t \models \forall\, x\,(\text{calls}(x) \Rightarrow -$ **knows_loses**$(t, x, x)))$.

Exhibit 5. 5: Poker Game Formalization.
Domain Rules 4 and 5.

persistence	If an infon is true in situation *s* and *s* is part of *t*, then that infon is true in *t.*
belief veridicality	If someone believes P, then P is true (only belief-as-knowledge is dealt with in this example).
logical closure of belief	If proposition P derives Q in classical logic, then if someone believes P they must also believe Q.
logical closure of support	If infon P derives Q in infon logic, then if *s* supports P it must also support Q.

Jack's conclusion is much simpler to derive than Zack's, so it is presented first:

1. $s \models \text{players}(pete, stone)$
2. $t \models \text{knows\_poker}(jack)$
3. $t \models \text{bel}(jack, \text{part\_of}(s_2, s))$
4. $t \models \text{sees}(jack, s_2)$
5. $s_2 \models \text{hand}(stone, sh) \wedge \text{hand}(pete, ph)$
6. $\text{better}(sh, ph)$
7. $t \models \text{bel}(jack, s_2 \models \text{hand}(stone, sh) \wedge \text{hand}(pete, ph))$ [seeing_is_believing and steps 4 and 5]
8. $t \models \text{bel}(jack,\ s \models \text{hand}(stone, sh) \wedge \text{hand}(pete, ph))$ [persistence and steps 3 and 7]
9. $t \models \text{bel}(jack, s \models \text{calls}(pete) \Rightarrow \text{loses}(pete))$ [Rule 3 and steps 1, 2,

6, and 8]

QED.

Zack's conclusion is derived as sketched below:

Suppose: t |= bel(*zack*, s |= calls(*pete*))

1. t |= knows_poker(*zack*)
2. t |= bel(*zack*, part_of(s_1, s))
3. t |= sees(*zack*, s_1)
4. t |= bel(*zack*, s |= bel(*pete*, s |= hand(*stone*, *sh*)))
5. s_1 |= hand(*stone*, *sh*) $\wedge$ players(*pete*, *stone*)
6. t |= bel(*zack*, s_1 |= hand(*stone*, *sh*) $\wedge$ players(*pete*, *stone*)) [seeing_is_believing and steps 4 and 5]
7. t |= bel(*zack*, s |= hand(*stone*, *sh*) $\wedge$ players(*pete*, *stone*)) [persistence and steps 3 and 6]
8. t |= bel(*zack*, s |= $\exists$ x bel(*pete*, s |= hand(*pete*,*x*))) [Rule 2]
9. t |= bel(*zack*, s |= – bel(*pete*, s |= calls(*pete*) $\Rightarrow$ loses(*pete*))) [Rule 5 and supposition]
10. t |= bel(*zack*, s |= bel(*pete*, s |= calls(*pete*) $\Rightarrow$ loses(*pete*)) $\vee$ bel(*pete*, s |= calls(*pete*) $\Rightarrow$ wins(*pete*))) [Rule 4 and steps 7 and 8]
11. t |= bel(*zack*, s |= bel(*pete*, s |= calls(*pete*) $\Rightarrow$ wins(*pete*))) [steps 9 and 10]
12. t |= bel(*zack*, s |= calls(*pete*) $\Rightarrow$ wins(*pete*)) [belief veridicality and step 11]
13. t |= bel(*zack*, s |= wins(*pete*)) [supposition and step 12]

Since supposing t |= bel(*zack*, s |= calls(*pete*)) derives t |= bel(*zack*, s |= wins(*pete*)) and the deduction theorem applies across belief and the support relation, then t |= bel(*zack*, s |= calls(*pete*) fi wins(*pete*)).

QED.

FELIX generates more detailed proofs similar to those above. These proofs generated by FELIX are discussed in the next chapter, after the extension of FELIX to handle perception, belief and the support relation via multiple-intensional context reasoning is presented.

The Two Wise Men

There are several versions of the "wise men" problem. The two man version is the simplest, although still quite challenging. Two problems are considered here. The first problem involves one wise man reasoning about another wise man's beliefs. This is the most common form of this problem. The second version involves a wise man reasoning about his own beliefs. This latter version is interesting because it uses more of the belief principles than the first one does. Konolige presents an extended analysis of the two-man form of this problem, focusing on a formal proof using his belief logic.[11] Frisch and Scherl give the following typical version of the two man problem[12]:

> ...there are two wise men named A and B. (1) A knows that if A does not have a white spot, B will know that A does not have a white spot. (2) A knows that B knows that either A or B has a white spot. B says that he does not know whether he has a white spot, and (3) A thereby knows that B does not know whether he has a white spot. The problem is to prove that (4) A knows that he has a white spot.

They present the formalization of this problem using two modal operators for belief, $[]_A$ and $[]_B$, one operator for each wise man. They formalize the numbered statements in the above quotation as follows[13]:

Given:	(1)	$[]_A(\neg\ White(A) \rightarrow []_B(\neg\ White(A)))$
	(2)	$[]_A([]_B(White(A) \vee White(B))$
	(3)	$[]_A(\neg\ []_B(White(B)))$
Prove:	(4)	$[]_A(White(A))$

Formula 4 is the theorem to prove given formulae 1, 2, and 3.

Konolige's formalization is similar to that of Frisch and Scherl, with minor naming changes. Konolige uses '[S]P' to indicate "S believes P". An interesting feature of

11. pp.57-61 in [Konolige 1986].
12. [Frisch&Scherl 1991], p. 198. This in turn is from [Genesereth&Nilsson 1987], p. 215-216.
13. p.198 in [Frisch&Scherl 1991].

Konolige's proof technique is that he uses "views" to reason about the beliefs of the agents in a fashion similar to the use of "intensional contexts" in FELIX. Konolige's formalization is as follows:[14]

Given:	(1k)	$W(A) \wedge W(B)$
	(2k)	$[S](W(A) \vee W(B))$
	(3k)	$[S][S'](W(A) \vee W(B))$
	(4k)	$W(S) \rightarrow [S']W(S)$
	(5k)	$\neg W(S) \rightarrow [S']\neg W(S)$[15]
	(6k)	$[S](W(S) \rightarrow [S']W(S))$
	(7k)	$[S](\neg W(S) \rightarrow [S']\neg W(S))$[16]
	(8k)	$[B]\neg [A]W(A)$
Prove:	(9k)	$[B]W(B)$

The two wise men, A and B, are in the reverse of the roles given them by Frisch and Scherl. 'S' can be either wise man.

Konolige only uses formulae 1k, 3k, 4k, 7k, and 8k in his proof. Formula 1 of Frisch and Scherl's version corresponds to 7k of Konolige's version, formula 2 corresponds to 3k, and formula 3 corresponds to 8k. There is no corresponding formula in Frisch and Scherl to formula 4k of Konolige. This formula appears to be superfluous in the proof which Konolige constructs. He uses it to establish that B know's A's spot to be white, but this is not actually used in the steps which lead to the proof (by contradiction). Thus, these two formalisms fundamentally agree on the basic formulae needed to establish the theorem.

To translate Frisch and Scherl's formulation into the belief logic presented in this thesis, the situation of the wise men must be identified. Call it *s.* The above formulae can be translated as follows:

Given:	(1s)	$s \models bel(a, s \models (\neg white(a) \text{ fi } bel(b, s \models \neg white(a))))$
	(2s)	$s \models bel(a, s \models bel(b, s \models white(a) \vee white(b)))$
	(3s)	$s \models bel(a, s \models \neg bel(b, s \models white(b)))$
Prove:	(4s)	$s \models bel(a, s \models white(a))$

14. pp. 58-59 in [Konolige 1986].
15. The second '¬ ' is missing in [Konolige 1986].
16. The second '¬ ' is missing in [Konolige 1986].

Prove:

$s \models bel(a, s \models white(a))$

Given:

$s \models bel(a, s \models (\neg white(a) \Rightarrow bel(b, s \models \neg white(a))))$

$s \models bel(a, s \models bel(b, s \models white(a) \vee white(b)))$

$s \models bel(a, s \models \neg bel(b, s \models white(b)))$

$s \models bel(a, s \models white(a) \vee \neg white(a))$

Exhibit 5. 6: Two Wise Men Problem: Theorem Statement

The situated version of this formulation is *not* a theorem. That is, the formula to be proved, 4s, does not follow from the given formulae, 1s, 2s, and 3s. According to this analysis, the missing piece of information is that A believes that the puzzle situation determines whether or not A has a white dot on his forehead. This can be stated as:

Additional Given:

(5s) $s \models bel(a, s \models white(a) \vee \neg white(a))$

The summary of the formalization of the two wise man problem is given in Exhibit 5. 6 on page 158.

The proof of this theorem relies on a *reductio ad absurdam* argument: if A were to believe that the situation doesn't support A having a white dot, it would lead to a contradiction in A's beliefs. That contradiction being that the situation supports that B doesn't believe he has a white dot, and that the situation does *not* support that B doesn't believe he has a white dot. Since supposing that the situation doesn't support his having a white dot leads A to contradictory beliefs, he can believe the negation of the supposition - that the situation *does* support his having a white dot.

The *reductio ad absurdam* argument takes place in the intensional context of A's beliefs. This is because the formula which is to be proved via this argument is a belief of A, and its negation which is being "supposed" is a belief of A. *Reductio ad absurdam* reasoning is valid in the intensional context of A's beliefs since a belief intensional context uses classical logic.[17]

17. *Reductio ad absurdam* reasoning is *not* valid in infon logic.

Prove:

$s \models \text{bel}(b, s \models \text{white}(a))$

Given:

$s \models \text{bel}(b, s \models \text{white}(a) \vee \text{white}(b))$

$s \models \text{bel}(b, s \models \text{white}(a)) \vee \text{bel}(b, s \models \neg \text{white}(a))$

$s \models \neg \text{bel}(b, s \models \text{white}(b))$

Exhibit 5. 7: Two Wise Men Introspection Problem: Theorem Statement

This proof of the two wise men puzzle requires only one principle of belief, that beliefs are closed under classical logic. Since this problem uses so few of the principles of belief it is not very satisfying as a demonstration of reasoning with these principles. There is a closely related puzzle which involves most of the principles of belief, however: Given a similar setup to the previous "two wise men" puzzle, prove that the situation supports that B believes that A has a white dot. This is presented in Exhibit 5. 7 on page 159. The setup for this problem is that we are given that the situation supports that B believes that the situation supports that either A or B has a white dot, that the situation supports that B believes that the situation supports that A has a white dot or that B believes that the situation supports that A does *not* have a white dot, and that the situation supports that B does *not* believe that the situation supports that B has a white dot. To prove this theorem one uses *reductio ad absurdam* reasoning in conjunction with four of the belief principles: closure under classical logic, veridicality, positive introspection, and negative introspection. It is the use of the introspection principles which gives the problem its name.

Barwise 1981 "Scenes and Other Situations" by Jon Barwise, in *The Journal of Philosophy*, 1981, 369-97. First reprinted in *The Philosopher's Annual* V, Boyer, Grim, and Sanders (eds.), (1982): 67-96. Also reprinted in *The Situation in Logic* by Jon Barwise, Center for the Study of Language and Information: Stanford University, 1988.

Barwise 1986 "Conditionals and Conditional Information" by Jon Barwise, in *On Conditionals* by Traugott, et. al. (eds.). Cambridge University Press, 1986. Reprinted on p. 97-135 in *The Situation in Logic* by Jon Barwise, Center for the Study of Language and Information: Stanford University, 1988.

Barwise&Perry 1983 *Situations and Attitudes* by Jon Barwise and John Perry, MIT Press: Cambridge, 1983.

Frisch&Scherl 1991 "A General Framework for Modal Deduction" by Alan M. Frisch and Richard B. Scherl, pp. 196-207, in *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference (KR91)* edited by James Allen, Richard Fikes, and Erik Sandewall. Morgan Kaufmann Publishers, Inc:San Mateo, CA. 1991.

Genesereth&Nilsson 1987 *Logical Foundations of Artificial Intelligence* by Michael Genesereth and Nils Nilsson. Morgan Kaufmann Publishers, Inc:Los Altos, CA. 1987.

Gibbard 1981 "Two Recent Theories of Conditionals" by Allan Gibbard in: *Ifs: Conditionals, Belief, Decision, Chance and Time* edited by W. L. Harper, R. Stalnaker, and G. Pearce. Dordrecht: Reidel. 1981.

Konolige 1986 *A Deduction Model of Belief* by Kurt Konolige, Los Altos:Morgan Kaufmann Publishers, Inc. 1986.

Stalnaker 1984 *Inquiry* by Robert Stalnaker. Cambridge, Massachusetts: MIT Press. 1984.