Centrum voor Wiskunde en Informatica
REPORT*RAPPORT*

Honesty in partial logic

W. van der Hoek, J. Jaspars and E. Thijsse

Computer Science/Department of Software Technology

CS-R9512 1995

Report CS-R9512
ISSN 0169-118X

CWI
P.O. Box 94079
1090 GB Amsterdam
The Netherlands

Honesty in Partial Logic

Wiebe van der Hoek[1], Jan Jaspars[2] and Elias Thijsse[3]

[1] *Dept. of Computer Science, Utrecht University, P.O. Box 80089 3508 TB Utrecht, The Netherlands*
wiebe@cs.ruu.nl

[2] *CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands*
jaspars@cwi.nl

[3] *ITK, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands*
thysse@kub.nl

Abstract

We propose an epistemic logic in which knowledge is fully introspective and implies truth, although truth need not imply epistemic possibility. The logic is presented in sequential format and is interpreted in a natural class of partial models, called balloon models. We examine the notions of honesty and circumscription in this logic: What is the state of an agent that 'only knows φ' and which *honest* φ enable such circumscription? Redefining *stable sets* enables us to provide suitable syntactic and semantic criteria for honesty. The rough syntactic definition of honesty is the existence of a minimal stable expansion, so the problem resides in the ordering relation underlying minimality. We discuss three different proposals for this ordering, together with their semantic counterparts, and show their effects on the induced notions of honesty.

AMS Subject Classification (1991): 03B20,03B45, 03B50, 03C90, 14E30, 68T30.
CR Subject Classification (1991): F.4.1, H.2.1, I.2.0, I.2.4.
Keywords and Phrases: knowledge representation, circumscription, honesty, modal logic, partial models, stable sets
Note: Wiebe van der Hoek was partially supported by ESPRIT Basic Research Action No. 6156 (DRUMS). Jan Jaspars was sponsored by CEC-project LRE-62-051 (FraCaS). This paper was presented at the fourth international conference on the principles of knowledge representation and reasoning (KR'94) in May 1994, Bonn, Germany. A short version of this paper has been published in the corresponding proceedings [HJT94].

1. Introduction

In this paper we argue that honesty in knowledge representation calls for a *partial* approach, for reasons of adequacy and efficiency. Let us informally explain the matter, starting by indicating that there are good reasons to embed epistemic logic in a partial modal context[1]. The first reason deals with *lack of truth*: if an agent considers no world possible in which p is true, it may simply mean that he does not reckon with p, rather than that he knows that $\neg p$ is true. In other words, the *absence* of truth of a formula in the agent's alternatives, should not imply the *presence* of the negated formula in his knowledge-state. Secondly, it offers a more natural way to deal with *growth of information:* in a classical possible world semantics this grow of information can only be modelled by elimination of possible worlds. Hence, *certainty grows* parallel with new information. In the partial approach the agent may constructively add worlds to his alternatives, or fill them up with new gained evidence. In this

[1] [BP83] shows how partial possible world semantics may present a fine analysis of 'propositional attitudes'; [Jas94] provides an extensive motivation for using partial modal logic when studying logics of knowledge.

way, disjunctive information (generally giving rise to the construction of several worlds) is generating *uncertainty*. Thirdly, the notion of *active disbelief* that partial epistemic logic can distinguish, makes the principle of negative introspection more natural than in the classical case: if we now have $\neg\Box\varphi$, it means that the agent has constructed himself a possibility in which $\neg\varphi$ is true, and hence it seems natural to expect that he than also knows this: $\Box\neg\Box\varphi$.

Let us now give some technical arguments for laying the foundations for epistemic logic in a partial semantics: it allows for a greater flexibility with respect to the epistemic background logic. For the case study presented in this paper this is revealed in adopting the veridicality principle of knowledge $\Box\varphi \Rightarrow \varphi$, without being forced to accept its contrapositive $\varphi \Rightarrow \Diamond\varphi$ ('if something is true, you must consider it possible'). Moreover, knowledge will be fully introspective: both positive introspection and negative introspection as well as their contrapositives are properties our logic embraces. In all, the logic resembles a weak variant of the classical system **S5**, although its epistemic properties are stronger than so-called *weak* **S5**. Apart from fitting our intuitions about (strong) knowledge, this logic enables us to simplify the partial models needed, essentially omitting most of the relational structure.

Having motivated our choice for partial semantics for epistemic logic, we now argue that the treatment of *honesty* makes the need for partiality even more eminent. The notion of honesty was introduced by Halpern and Moses ([HM85]), who presented a thorough study of honesty for the 'classical case'. Honesty is the quality of a proposition which can be said to be *only* known, i.e. knowing that fact and its consequences, but not knowing more than that. For example, you may only know p, without knowing anything at all about, say, q. Also, you may only know that either p or q will be the case, which implies you do not know which one of the two is true. These are examples of honest knowledge. By contrast, you cannot sincerely claim to only know (that you know) *whether* p is true, for then you would either know p or know that $\neg p$, both options being logically stronger than what is supposed to be known. Hence, the formula $(\Box p \vee \Box\neg p)$ is called *dishonest*.

Partiality and honesty may seem totally unrelated themes, but in fact we argue that they are closely related. Let us reinspect the case in which you only know p. This, we claim, does not involve any knowledge about q. In particular it does not even imply that you know the *possibility* that q. In a straightforward total semantics ignorance leads to wide knowledge of possibilities, which, however, contradicts the initial idea of *only* knowing some honest formula. The proliferation problem simply does not occur in our partial semantics, since facts unrelated to some honest formula can be left undefined. In other words, when considering the consequences of saying that one only knows φ, classical approaches try to *maximize possibilities*, whereas a partial treatment *economizes* them. Thus, not only is the relational structure of our partial models simple: using them we also gain efficiency, which is reflected in the small size of the characterizing models, whereas a classical possible world model tends to 'explode' when representing ignorance.

Apart from the fact that our partial semantics both allows for a greater flexibility of the underlying epistemic background logic, and elegantly solves the proliferation problem by economizing possibilities, let us try and indicate how we think our approach to honesty improves upon 'classical' treatments. A first difference deals with the *logical omniscience* of a — possibly totally ignorant — agent, and is related to the proliferation problem. In the classical approach of [HM85] for instance, the formula $\top$ is perfectly honest, doing justice to

the fact that it makes sense to claim that one 'knows nothing'. We agree with the designation of $\top$ being honest, but we depart from [HM85] when determining the *consequences* of such an observation. Let us write, for any honest formula φ and epistemic formula ψ,

$$\varphi \mid\!\sim \psi$$

for 'the agent knows ψ, if he claims to only know φ'. Then, in the framework of [HM85] one obtains $\top \mid\!\sim \psi$ when ψ is an **S5**-tautology and one also obtains $\top \mid\!\sim \Diamond s \wedge \Diamond \neg s$, for any atom s. However, we prefer the ignorant agent who only knows $\top$, to really let him know no more than that; in our set-up, we have $\top \mid\!\sim \psi$ iff ψ is a tautology. Now, there are relatively few tautologies in partial logic: all of them essentially contain $\top$. More generally, we agree with the analysis of [HM85] that objective (propositional) formulas should be rendered honest, but we depart from [HM85] when it comes to the $\mid\!\sim$ -consequences of such formulas. Generalizing the situation we gave above, one may say that our $\mid\!\sim$ gives a serious account of *relevance* : if φ and ψ have no proposition symbol in common, and do not contain $\top$, then $\varphi \mid\!\not\sim \psi$. Thus, e.g. $p \mid\!\not\sim (q \vee \neg q)$.

Moving up in the hierarchy from objective formulas to epistemic statements, our classification of honest formulas is starting to diverge from the classical analysis. We agree with that analysis that disjunctive epistemic assertions often are problematic with regard to honesty. The argument runs like this: the formula $\varphi = \Box p \vee \Box q$ is dishonest; only knowing φ implies not knowing any stronger formula, in particular, this gives not knowing p and not knowing q, but the latter two conclusions are easily seen to be inconsistent with φ. However, where the framework of [HM85] does indeed label the disjunctive epistemic formula $(\Box p \vee \Box q)$ dishonest, we think their set-up still yields some counterintuitive results. For instance, their definitions are such that the disjunctive epistemic formula $\psi = (\Diamond p \vee \Diamond q)$ is still honest! But here, we think a similar argument as for $(\Box p \vee \Box q)$ can be given to account for ψ's dishonesty: if you only know that you either consider p to be possible, or q (to be possible), then you must know which of the two.

In the next section we introduce the epistemic logic, presenting its language, semantics and inference system, and prove its completeness. Then, in section 3, we study ways to minimize knowledge for this logic, discussing different notions of honesty. For each of these notions, we present deductive, model-theoretic and syntactic characterizations.

2. The Logic

In this section we introduce a partial modal logic **L** of which we will investigate the notions of stability, honesty and several disjunction properties in subsequent sections. We present our logic following a common pattern: we first give its language and semantics (section 2.1), then we provide a deductive system for **L** (2.2) and finally prove a completeness result (2.3) connecting syntax (deduction) and semantics (consequence).

2.1 Language and Semantics

Definition 2.1 Let $\mathcal{P}$ be a non-empty countable set of propositional variables. The *language* $\mathcal{L}$ is the smallest superset of $\mathcal{P}$ such that

$$\varphi, \psi \in \mathcal{L} \Rightarrow \neg\varphi, (\varphi \wedge \psi), \bot, \Box\varphi \in \mathcal{L}.$$

$\mathcal{L}_0$ is the subset of $\mathcal{L}$ of all formulas which do not contain $\Box$-operators. For any $\Gamma \subseteq \mathcal{L}$, we write Γ_0 for $\Gamma \cap \mathcal{L}_0$ and $\overline{\Gamma}$ for $\{\varphi \in \mathcal{L} \mid \varphi \notin \Gamma\}$. Moreover, for any $\Gamma \subseteq \mathcal{L}$ and any $\odot \in \{\neg, \Box, \Diamond\}$, we define $\odot\Gamma = \{\odot\gamma \mid \gamma \in \Gamma\}$ and $\odot^-\Gamma = \{\gamma \mid \odot\gamma \in \Gamma\}$.

Here, the intended meaning of $\Box\varphi$ is that 'φ is known'. We write $\top$ for $\neg\bot$, $\varphi \vee \psi$ for $\neg(\neg\varphi \wedge \neg\psi)$ and $\Diamond\varphi$ for $\neg\Box\neg\varphi$. It is important to note that in our set-up, $\Diamond\varphi$ does not just mean that $\neg\varphi$ is not known, but that the agent considers some epistemic alternative to be possible, in which φ has a meaning: it is true!

Given a set of formulas Γ, we may consider its *objective kernel* (Γ_0), the *knowledge* it encodes ($\Box^-\Gamma$) and its *possibilities* ($\Diamond^-\Gamma$). These sets induce the following orderings.

DEFINITION 2.2 Let Γ and Δ be sets of formulas of $\mathcal{L}$. Then:

- $\Gamma \subseteq_0 \Delta \Leftrightarrow \Gamma_0 \subseteq \Delta_0$
- $\Gamma \subseteq_\Box \Delta \Leftrightarrow \Box^-\Gamma \subseteq \Box^-\Delta$
- $\Gamma \subseteq_\Diamond \Delta \Leftrightarrow \Diamond^-\Gamma \subseteq \Diamond^-\Delta$

Each of these orders can be linked to an equivalence relation, e.g. $\Gamma =_0 \Delta \Leftrightarrow \Gamma_0 = \Delta_0$. Let $\Re$ be some subset of $\wp(\mathcal{L})$, the power set of $\mathcal{L}$. For any $\star \in \{0, \Box, \Diamond\}$, we say that $\Gamma \in \Re$ is $\subseteq_\star$-*minimal in* $\Re$, if for all $\Delta \in \Re$, $\Gamma \subseteq_\star \Delta$, and similarly for $\subseteq$-minimality in $\Re$. Note that these minimal sets are the first (smallest) elements with respect to the corresponding ordering, rather than the elements without a predecessor.

We now give a formal interpretation of the language $\mathcal{L}$. The mathematical structure for such an interpretation is a Kripke model with partial worlds. Since we are only interested in models for our epistemic logic here, we do not have to consider arbitrary partial Kripke models.[2] Instead, we restrict attention to what we call 'balloon models', which are somewhat reminiscent of the well-known **KD45**-Kripke models. The basic entities in our balloon models are partial worlds, which are linked to partial valuations.

DEFINITION 2.3 A *partial valuation* V is a partial function which assigns truth-values to a given set of propositional variables $\mathcal{P}$. The collection of all partial valuations is denoted by VAL. The *domain of* V is defined as $Dom(V) = \{\, p \in \mathcal{P} \mid V(p) \in \{0,1\} \,\}$. $V' \in \mathsf{VAL}$ is said to be an *extension* of $V \in \mathsf{VAL}$ if $V(p) = V'(p)$ for all $p \in Dom(V)$. We abbreviate the extension relation by $V \sqsubseteq V'$.

DEFINITION 2.4 A *balloon model* is a triple $M = \langle W, g, V\rangle$ with W a non-empty finite set of worlds, called the *balloon*, g the *root* or *generator* of the model, and V a global valuation function $V : W \cup \{g\} \to \mathsf{VAL}$, such that $V(w) \sqsubseteq V(g)$ for certain $w \in W$. We also write M_g for such a model: notice that any combination of $w \in W$ and $V : W \to \mathsf{VAL}$ gives rise to a balloon model $M_w = \langle W, w, V\rangle$.

The *truth* and *falsity* of a formula $\varphi \in \mathcal{L}$ in a balloon model $M = \langle W, g, V\rangle$, written as $M \models \varphi$ and $M \nvDash \varphi$, respectively, are defined by induction:

[2] For a general approach, see [Thi92] or [JT93].

$M \not\models \bot$		$M \dashv \bot$	
$M \models p$	$\Leftrightarrow V(g)(p) = 1$ $(p \in \mathcal{P})$	$M \dashv p$	$\Leftrightarrow V(g)(p) = 0$ $(p \in \mathcal{P})$
$M \models \neg\varphi$	$\Leftrightarrow M \dashv \varphi$	$M \dashv \neg\varphi$	$\Leftrightarrow M \models \varphi$
$M \models \varphi \wedge \psi$	$\Leftrightarrow M \models \varphi$ and $M \models \varphi$	$M \dashv \varphi \wedge \psi$	$\Leftrightarrow M \dashv \varphi$ or $M \dashv \psi$
$M \models \Box\varphi$	$\Leftrightarrow M_w \models \varphi$ for all $w \in W$	$M \dashv \Box\varphi$	$\Leftrightarrow M_w \dashv \varphi$ for some $w \in W$

Note the special role played in the truth definition by the root of the model. Although we usually display the models with the root outside of the balloon, the recursive $\Box$-clauses show that this need not be the case. (Alternatively, we can duplicate the root, the new root being outside of the balloon.) Also note that the truth-definitions yield the intended effect for $\Diamond$-formulas: we have that $M \models \Diamond\varphi \Leftrightarrow M_w \models \varphi$ for some $w \in W$. In particular, our partial semantics makes $\Box\varphi \vee \neg\Box\varphi$, and hence $\Box\neg\varphi \vee \Diamond\varphi$ invalid. This reflects the idea that, in our opinion, $\Diamond\varphi$-formulas should express some positive evidence about φ, not just lack of knowledge of $\neg\varphi$.

For any model $M = \langle W, g, V\rangle$ we define the *theory* $Th(M)$ of M by $Th(M) = \{\varphi \in \mathcal{L} \mid M \models \varphi\}$, the *knowledge* $\kappa(M)$ in M by $\kappa(M) = \{\varphi \in \mathcal{L} \mid M \models \Box\varphi\}$ and the *possibilities* $\pi(M)$ by $\pi(M) = \{\varphi \in \mathcal{L} \mid M \models \Diamond\varphi\}$. Note that $\kappa M = \Box^- Th(M)$ and $\pi(M) = \Diamond^- Th(M)$. Let Γ and Δ be sets of formulas. We write $\Gamma \models \Delta$, if all balloon models which verify all members of Γ also verify at least one of the elements of Δ, i.e. $\forall M : \Gamma \subseteq Th(M) \Rightarrow \Delta \cap Th(M) \neq \emptyset$. Finally, we write $M \models \Gamma$ for $\Gamma \subseteq Th(M)$.

The following proposition presents two counterparts of a persistence result from partial propositional logic for balloon models.

PROPOSITION 2.5 Let $M = \langle W, g, V\rangle$ and $M' = \langle W', g', V'\rangle$ be two balloon models. Then, for all $w \in W \cup \{g\}$ and all $w' \in W' \cup \{g'\}$:

$$V(w) \sqsubseteq V'(w') \Leftrightarrow \forall \pi \in \mathcal{L}_0 : (M_w \models \pi \Rightarrow M'_{w'} \models \pi)$$

For every balloon model $M = \langle W, g, V\rangle$ and $w, w' \in W \cup \{g\}$:

$$V(w) \sqsubseteq V(w') \Leftrightarrow \forall \varphi \in \mathcal{L} : (M_w \models \varphi \Rightarrow M_{w'} \models \varphi)$$

Proofs for these two simple observations can be obtained immediately by the persistence of $\mathcal{L}_0$ with respect to the extension relation over VAL [Bla86]. We call the former property *propositional persistence* and the latter *internal persistence*. The way of ordering possible worlds in a balloon model by only considering their local truth-assignment is too limited to give a sufficient and necessary condition of informational expansion of balloon models. The following notion will settle this equivalence.

DEFINITION 2.6 For two balloon models $M = \langle W, g, V\rangle$ and $M' = \langle W', g', V'\rangle$ we say that M' is a *bisimulation extension* of M, if

- $V(g) \sqsubseteq V'(g)$,
- $\forall w \in W\ \exists w' \in W'$ such that $V(w) \sqsubseteq V'(w')$,
- $\forall w' \in W'\ \exists w \in W$ such that $V(w) \sqsubseteq V'(w')$.

This definition is in fact a special case of a more complex definition for arbitrary partial Kripke models (which have a genuine accessibility relation), which in its turn is linked to the general notion of bisimulation in process algebra and modal logic. [3]

THEOREM 2.7 M' is a bisimulation extension of $M \Leftrightarrow \forall \varphi \in \mathcal{L} : M \models \varphi \Rightarrow M' \models \varphi$.

Proof: The left to right direction of the theorem can be proved by a simple induction over the construction of $\mathcal{L}$-formulas. The $\mathcal{L}_0$-part is in fact the same as the propositional persistence result in proposition 2.5, which is applicable through the first requirement in definition 2.6. Persistence of formulas of the form $\Box\varphi$ and $\neg\Box\varphi$ are immediate consequences of the third and second requirement in definition 2.6, respectively.

From right to left, suppose that $M' = \langle W', g', V' \rangle$ is not a bisimulation extension of $M = \langle W, g, V \rangle$. By definition 2.6, we have one of the following cases:

- $V(g) \not\subseteq V'(g')$. Applying proposition 2.5 we find a formula $\pi \in \mathcal{L}_0$ such that $M \models \pi$ and $M' \not\models \pi$.
- $\exists w \in W \, \forall w' \in W' : V(w) \not\subseteq V'(w')$. For such w proposition 2.5 gives us, for each $w' \in W'$, a formula $\pi_{w'} \in \mathcal{L}_0$ for which $M_w \models \pi_{w'}$, but $M'_{w'} \not\models \pi_{w'}$. But then $M \models \Diamond \bigwedge_{w' \in W'} \pi_{w'}$, whereas $M' \not\models \Diamond \bigwedge_{w' \in W'} \pi_{w'}$.
- $\exists w' \in W' \, \forall w \in W : V(w) \not\subseteq V'(w')$. For such w' proposition 2.5 gives us, for each $w \in W$, formulas $\pi_w \in \mathcal{L}_0$ such that $M_w \models \pi_w, M'_{w'} \not\models \pi_w$. But then $M \models \Box \bigvee_{w \in W} \pi_w$, whereas $M' \not\models \Box \bigvee_{w \in W} \pi_w$.

Summarizing, we see that if M' is not a bisimulation extension of M, we always find a formula φ for which $M \models \varphi$, $M' \not\models \varphi$. ∎

2.2 Deductions in **L**

We now formally define the deductive machinery of our logic. The sequent $\Gamma \vdash \Delta$, with $\Gamma, \Delta \subseteq \mathcal{L}$, should be understood as: 'the disjunction of the members of Δ follows from the conjunction of the formulas in Γ'. In such a sequent, Γ and Δ are considered to be sequences rather than sets. Instead of $\Gamma \cup \{\varphi\}$ and $\Gamma \cup \Delta$ we write Γ, φ and Γ, Δ respectively.

We use a sequential axiomatization of partial logic for several reasons.[4] It is short and clearly marks the difference with classical logic in a deductive fashion. Moreover, it smoothes the meta-theory of partial logic (see also [JT93] and [Jas94]). The latter advantage becomes clear in this paper in subsections 2.3 and 3.3.

DEFINITION 2.8 To start with, we distinguish the following *structural rules*:

- $\Gamma \cap \Delta \neq \emptyset \Rightarrow \Gamma \vdash \Delta$ START
- $\dfrac{\Gamma \vdash \Delta \quad \Delta \subseteq \Delta' \quad \Gamma \subseteq \Gamma'}{\Gamma' \vdash \Delta'}$ MON
- $\dfrac{\Gamma \vdash \varphi, \Delta \quad \Gamma', \varphi \vdash \Delta'}{\Gamma, \Gamma' \vdash \Delta, \Delta'}$ CUT

[3] See e.g. [Sti87]. In [Jas94] this general notion has been used to transfer other information relations from partial propositional logic to partial modal logic.

[4] Cf. [Thi90] for a different, natural deduction type presentation.

The following rules explain how the logical constants are introduced on the left (L-TRUE) and right hand side (R-TRUE) of the '$\vdash$'-sign, respectively, possibly accompanied with a negation sign (L-FALSE or R-FALSE). First we give the propositional rules for the connectives.

- $\dfrac{\Gamma \vdash \varphi, \Delta}{\Gamma, \neg\varphi \vdash \Delta}$ L-TRUE$\neg$
- $\dfrac{\Gamma, \varphi \vdash \Delta}{\Gamma, \neg\neg\varphi \vdash \Delta}$ L-FALSE $\neg$
- $\dfrac{\Gamma \vdash \varphi, \Delta}{\Gamma \vdash \neg\neg\varphi, \Delta}$ R-FALSE $\neg$
- $\Gamma \vdash \neg\bot, \Delta$ R-FALSE $\bot$
- $\dfrac{\Gamma, \varphi, \psi \vdash \Delta}{\Gamma, \varphi \wedge \psi \vdash \Delta}$ L-TRUE $\wedge$
- $\dfrac{\Gamma \vdash \varphi, \Delta \quad \Gamma' \vdash \psi, \Delta'}{\Gamma, \Gamma' \vdash \varphi \wedge \psi, \Delta, \Delta'}$ R-TRUE $\wedge$
- $\dfrac{\Gamma, \neg\varphi \vdash \Delta \quad \Gamma', \neg\psi \vdash \Delta'}{\Gamma, \Gamma', \neg(\varphi \wedge \psi) \vdash \Delta, \Delta'}$ L-FALSE $\wedge$
- $\dfrac{\Gamma \vdash \neg\varphi, \neg\psi, \Delta}{\Gamma \vdash \neg(\varphi \wedge \psi), \Delta}$ R-FALSE $\wedge$

We add the following 'epistemic' rules to this propositional basis.

- $\dfrac{\Gamma, \varphi \vdash \Delta}{\Gamma, \Box\varphi \vdash \Delta}$ L-TRUE $\Box$
- $\dfrac{\Gamma \vdash \varphi, \neg\Delta}{\Box\Gamma \vdash \Box\varphi, \neg\Box\Delta}$ R-TRUE $\Box$
- $\dfrac{\Gamma, \neg\varphi \vdash \neg\Delta}{\Box\Gamma, \neg\Box\varphi \vdash \neg\Box\Delta}$ L-FALSE $\Box$
- $\dfrac{\Gamma \vdash \Box\varphi, \Delta}{\Gamma \vdash \Box\Box\varphi, \Delta}$ $4_\Box$
- $\dfrac{\Gamma, \Box\varphi \vdash \Delta}{\Gamma, \neg\Box\neg\Box\varphi \vdash \Delta}$ $5_\Diamond$
- $\dfrac{\Gamma \vdash \neg\Box\varphi, \Delta}{\Gamma \vdash \Box\neg\Box\varphi, \Delta}$ $5_\Box$

Obvious gaps in this system **L** are L-TRUE $\bot$ ($\Gamma, \bot \vdash \Delta$) and R-TRUE $\neg$:

$$\frac{\Gamma, \varphi \vdash \Delta}{\Gamma \vdash \neg\varphi, \Delta}.$$

It turns out that the former rule is derivable in **L**. The latter rule cannot be derived in **L**. Adding rule R-TRUE $\neg$ to system **L** would make **L** coincide with the classical modal system **S5**.

DEFINITION 2.9 The rules above are called **L**-*rules*. A sequence $\Delta \subseteq \mathcal{L}$ is said to be **L**-*derivable* from another sequence $\Gamma \subseteq \mathcal{L}$, $\Gamma \vdash_{\mathbf{L}} \Delta$, if $\Gamma \vdash \Delta$ can be derived by a finite number of applications of **L**-rules. We usually drop the subscript '**L**' in the sequel. Then, two formulas $\varphi, \psi \in \mathcal{L}$ are said to be equivalent, $\varphi \dashv\vdash \psi$, if $\varphi \vdash \psi$ and $\psi \vdash \varphi$.

Derivable sequents are at least valid on balloon models:

LEMMA 2.10 *(Soundness)* For all $\Gamma, \Delta \subseteq \mathcal{L}$: $\Gamma \vdash \Delta \Rightarrow \Gamma \models \Delta$.

Proof: We prove soundness of L-TRUE $\Box$ and R-TRUE $\Box$. To start with L-TRUE $\Box$, suppose that $\Gamma, \varphi \models \Delta$. This means that for arbitrary balloon models M we have $M \models \Gamma \cup \{\varphi\} \Rightarrow \exists\delta \in \Delta, M \models \delta$ (*).

To prove $\Gamma, \Box\varphi \models \Delta$, suppose that $N = \langle W, g, V\rangle$ is an arbitrary balloon model for which $N \models \Gamma \cup \{\Box\varphi\}$. This means that both $N_g \models \Gamma$ and $N_g \models \Box\varphi$. By definition of balloon model, there is some $w \in W$ with $V(w) \sqsubseteq V(g)$. Since $N_g \models \Box\varphi$, we have for this w that $N_w \models \varphi$. Now by internal persistence we conclude that $N_g \models \varphi$. Thus we have $N_g \models \Gamma \cup \{\varphi\}$, and applying (*), we get $N_g \models \delta$, for some $\delta \in \Delta$.

To prove soundness of R-TRUE $\Box$, suppose that $\Gamma \models \varphi, \neg\Delta$. Let M be a model $\langle W, g, V\rangle$ such that $M \models \Box\Gamma$. So $M_w \models \Gamma$ for all $w \in W$. Now suppose $M \not\models \Box\varphi$, then, for some $u \in W$ we have $M_u \not\models \varphi$. Since $\Gamma \models \varphi, \neg\Delta$, we have $M_u \models \neg\delta$ for certain $\delta \in \Delta$, and hence $M \models \neg\Box\delta$. Therefore $M \models \Box\varphi, \neg\Box\Delta$ for an arbitrary balloon model M verifying $\Box\Gamma$, hence $\Box\Gamma \models \Box\varphi, \neg\Box\Delta$. ∎

Let us pause for a moment and reflect on our logic. Claims below that some sequents are not derivable, are now easily verified semantically, as is justified by lemma 2.10.

- The first thing to note about the logic is that it is indeed partial, which is mirrored by the fact that we do not have the *law of excluded middle*: $\not\vdash \varphi, \neg\varphi$. In fact, as is shown in [Thi92], there is not any theorem of **L** in the $\{\bot, \top\}$-free language. This can immediately be seen from the surface of the sequential presentation of **L**. The only way to obtain an empty left hand argument in a sequent is to use R-FALSE $\bot$. The non-derivability of the law of excluded middle can also be demonstrated directly by a counterexample, applying the soundness result in lemma 2.10. Any balloon model with an empty valuation[5] on its root is a counterexample for $p \vee \neg p$.

- Moreover, we do not have *contraposition*: $\Gamma \vdash \Delta \not\Rightarrow \neg\Delta \vdash \neg\Gamma$.

- Although **L** lacks contraposition and does not have any $\{\bot, \top\}$-free theorems, much of the structure of classical logic is preserved. Well-known principles such as the laws of De Morgan and double negation, and properties such as associativity, idempotence and commutativity of the disjunction and conjunction are also valid in **L**.

- For the defined symbols one easily proves the following *derived rules:*

$$\Gamma, \neg\top \vdash \Delta \quad \text{L-FALSE } \top \qquad\qquad \Gamma \vdash \top, \Delta \quad \text{R-TRUE } \top$$

$$\frac{\Gamma, \varphi \vdash \Delta \quad \Gamma', \psi \vdash \Delta'}{\Gamma, \Gamma', \varphi \vee \psi \vdash \Delta, \Delta'} \quad \text{L-TRUE } \vee \qquad \frac{\Gamma \vdash \varphi, \psi, \Delta}{\Gamma \vdash \varphi \vee \psi, \Delta} \quad \text{R-TRUE } \vee$$

$$\frac{\Gamma, \neg\varphi, \neg\psi \vdash \Delta}{\Gamma, \neg(\varphi \vee \psi) \vdash \Delta} \quad \text{L-FALSE } \vee \qquad \frac{\Gamma \vdash \neg\varphi, \Delta \quad \Gamma' \vdash \neg\psi, \Delta'}{\Gamma, \Gamma' \vdash \neg(\varphi \vee \psi), \Delta, \Delta'} \quad \text{R-FALSE } \vee$$

$$\frac{\Gamma, \varphi \vdash \Delta}{\Box\Gamma, \Diamond\varphi \vdash \Diamond\Delta} \quad \text{L-TRUE } \Diamond \qquad \frac{\Gamma \vdash \neg\varphi, \Delta}{\Box\Gamma \vdash \neg\Diamond\varphi, \Diamond\Delta} \quad \text{R-FALSE } \Diamond$$

- For the *epistemic part*, we have the following:

  Conjoining knowledge: $\Box\varphi \wedge \Box\psi \dashv\vdash \Box(\varphi \wedge \psi)$.

  Disjoining epistemic possibility: $\Diamond(\varphi \vee \psi) \dashv\vdash \Diamond\varphi \vee \Diamond\psi$.

[5] A valuation V is empty in world w of $Dom(V(w)) = \emptyset$.

Positive introspection: $\Box\varphi \vdash \Box\Box\varphi \quad \Diamond\Diamond\varphi \vdash \Diamond\varphi$

Negative introspection: $\neg\Box\varphi \vdash \Box\neg\Box\varphi \quad \Diamond\Box\varphi \vdash \Box\varphi$

Veridicality: $\Box\varphi \vdash \varphi \quad \varphi \nvdash \Diamond\varphi$!

Consistency of knowledge $\vdash \neg\Box\bot$

Note that, although we *do* have veridicality of knowledge ('known facts are true') we got rid of its contrapositive ('true facts are considered to be possible'). A counterexample can be given by the simple balloon model existing of a root g and one world $W = \{w\}$ such that $V(g)(p) = 1$ and an empty valuation at w. Then $\langle W, g, V\rangle$ is a balloon model, because $V(w) \sqsubseteq V(g)$. It verifies p but not $\Diamond p$.

Negative introspection is now better motivated than in classical **S5**: if some fact is considered possible by the agent, it is explicitly present in his set of alternatives, so he knows that particular possibility. This should be contrasted to the classical case where merely not knowing the opposite is supposed to involve knowledge of the possibility. In the sequel, we will denote a property like positive introspection by '$\Box \Rightarrow \Box\Box$' or '$\Diamond\Diamond \Rightarrow \Diamond$'.

To see the system **L** at work, we will provide a proof of the property that nestings of modal operators are in fact superfluous (theorem 2.14). Later on, this property will be used in our completeness proof. Let us first define the *modal depth* $md(\varphi)$ of a formula φ by: $md(p) = md(\bot) = 0$ $(p \in \mathcal{P})$; $md(\neg\varphi) = md(\varphi)$; $md(\varphi \wedge \psi) = \max(md(\varphi), md(\psi))$; $md(\Box\varphi) = 1 + md(\varphi)$. By using the propositional equivalences as stated above and treating formulas like $\Box\alpha$ and $\Diamond\beta$ as literals, we obtain the following *normal forms* in **L** :

PROPOSITION 2.11 Every $\varphi \in \mathcal{L}$ is equivalent to a formula of the form

$$\bigvee_{i=1}^{n} \bigwedge_{j=1}^{m} \varphi_{i,j}$$

where each $\varphi_{i,j}$ is of the form $\Box\alpha$ with $md(\alpha) < md(\varphi)$, $\Diamond\beta$ with $md(\beta) < md(\varphi)$, $\neg p$ or p. If $n = 0$, we interpret the disjunction as $\bot$, and if $n > 0$ but $m = 0$ as $\top$. We call the format displayed above a *semi-disjunctive normal form* of φ. There also exists a *semi-conjunctive normal form:*

$$\bigwedge_{i=1}^{n} \bigvee_{j=1}^{m} \varphi_{i,j}$$

where the formulas $\varphi_{i,j}$ are of the same form as above.

The following lemma is the heart of theorem 2.14: it explains how nestings of modal operators are removed.

LEMMA 2.12 We have:

$\Box(\Box\alpha \vee \psi) \dashv\vdash \Box\alpha \vee \Box\psi$ and $\Diamond(\Box\alpha \wedge \psi) \dashv\vdash \Box\alpha \wedge \Diamond\psi$.

Proof: We only prove the first equivalence, the second is similar.

1.	$\Box\alpha \vdash \Box\alpha$	START
2.	$\psi \vdash \psi$	START
3.	$\Box\alpha \vee \psi \vdash \Box\alpha, \psi$	L-TRUE $\vee$ (1,2)
4.	$\Box(\Box\alpha \vee \psi) \vdash \Diamond\Box\alpha, \Box\psi$	R-TRUE $\Box$ (3)
5.	$\Diamond\Box\alpha \vdash \Box\alpha$	$\Diamond\Box \Rightarrow \Box$
6.	$\Box(\Box\alpha \vee \psi) \vdash \Box\alpha, \Box\psi$	CUT (4,5)
7.	$\Box(\Box\alpha \vee \psi) \vdash \Box\alpha \vee \Box\psi$	R-TRUE $\vee$ (6)

1.	$\psi \vdash \Box\alpha, \psi$	START
2.	$\psi \vdash \Box\alpha \vee \psi$	R-TRUE $\vee$ (1)
3.	$\Box\psi \vdash \Box(\Box\alpha \vee \psi)$	R-TRUE $\Box$ (2)
4.	$\Box\alpha \vdash \Box\alpha, \psi$	START
5.	$\Box\alpha \vdash \Box\alpha \vee \psi$	R-TRUE $\vee$ (4)
6.	$\Box\Box\alpha \vdash \Box(\Box\alpha \vee \psi)$	R-TRUE $\Box$ (5)
7.	$\Box\alpha \vdash \Box\Box\alpha$	$\Box \Rightarrow \Box\Box$
8.	$\Box\alpha \vdash \Box(\Box\alpha \vee \psi)$	CUT (6,7)
9.	$\Box\alpha \vee \Box\psi \vdash \Box(\Box\alpha \vee \psi)$	L-TRUE $\vee$ (3,8)

■

COROLLARY 2.13

$$\Box(\bigvee_{i=1}^{n} \Box\alpha_i \vee \bigvee_{j=1}^{m} \Diamond\beta_j \vee \pi) \dashv\vdash \bigvee_{i=1}^{n} \Box\alpha_i \vee \bigvee_{j=1}^{m} \Diamond\beta_j \vee \Box\pi$$

$$\Diamond(\bigwedge_{i=1}^{n} \Box\alpha_i \wedge \bigwedge_{j=1}^{m} \Diamond\beta_j \wedge \pi) \dashv\vdash \bigwedge_{i=1}^{n} \Box\alpha_i \wedge \bigwedge_{j=1}^{m} \Diamond\beta_j \wedge \Diamond\pi$$

By means of these preliminaries we now easily establish:

THEOREM 2.14 Every $\varphi \in \mathcal{L}$ is equivalent to a formula φ' with $md(\varphi') \leq 1$.

The proof runs by induction on the modal depth of formulas. Obviously the result for the basic step is 'for free'.

Let the modal depth of φ be larger than 1. By proposition 2.11 φ is equivalent to the semi-disjunctive normal form

$$\bigvee_{i=1}^{n}(\bigwedge_{j=1}^{m} \Box\alpha_{i,j} \wedge \bigwedge_{k=1}^{\ell} \Diamond\beta_{i,k} \wedge \pi_i)$$

with $\pi \in \mathcal{L}_0$, $md(\alpha_{i,j}) < md(\varphi)$ and $md(\beta_{i,k}) < md(\varphi)$.

The induction hypothesis applies to all the formulas $\alpha_{i,j}$ and $\beta_{i,k}$. This means that these formulas can be assumed to have a modal depth at most 1. If this result can also be obtained for $\Box\alpha_{i,j}$ and $\Diamond\beta_{i,k}$, then the result has been shown for the formula φ. This result can be obtained quite easily by using lemma 2.12, together with $\Box(\alpha \wedge \alpha') \dashv\vdash (\Box\alpha \wedge \Box\alpha')$ and $\Diamond(\beta \vee \beta') \dashv\vdash (\Diamond\beta \vee \Diamond\beta')$. If α has modal depth 1 it must be equivalent to the semi-conjunctive normal form

$$\alpha \dashv\vdash \bigwedge_{i=1}^{n'}(\bigvee_{j=1}^{m'} \Box\sigma_{i,j} \vee \bigvee_{k=1}^{\ell'} \Diamond\eta_{i,k} \vee \pi_i) \text{ with } \pi_i, \sigma_{i,j}, \eta_{i,k} \in \mathcal{L}_0,$$

while each β is equivalent to the semi-disjunctive normal form:

$$\beta \dashv\vdash \bigvee_{i=1}^{n''}(\bigwedge_{j=1}^{m''} \Box\epsilon_{i,j} \wedge \bigwedge_{k=1}^{\ell''} \Diamond\lambda_{i,k} \wedge \varrho_i) \text{ with } \varrho_i, \epsilon_{i,j}, \lambda_{i,k} \in \mathcal{L}_0.$$

Corollary 2.13 yields, after applying R-TRUE $\Box$ and $\Box$-distribution over $\wedge$, for each α and β in the semi-disjunctive normal form of φ:

$$\Box\alpha \dashv\vdash \bigwedge_{i=1}^{n'}(\bigvee_{j=1}^{m'} \Box\sigma_{i,j} \vee \bigvee_{k=1}^{\ell'} \Diamond\eta_{i,k} \vee \Box\pi_i), \text{ and}$$

$$\Diamond\beta \dashv\vdash \bigvee_{i=1}^{n''}(\bigwedge_{j=1}^{m''} \Box\epsilon_{i,j} \wedge \bigwedge_{k=1}^{\ell''} \Diamond\lambda_{i,k} \wedge \Diamond\pi_i).$$

Clearly the latter formulas have a modal depth not larger than 1. ∎

Combined with proposition 2.11, this theorem also implies that every formula has a semi-disjunctive and a semi-conjunctive normal form of at most modal depth 1.

2.3 Completeness

The aim of this section is to prove that the logic **L** is complete for the class of balloon models. By definition 2.4 our models are *finite*; as a consequence, not each consistent set will be satisfiable (e.g. $\{\Diamond(p_1 \wedge \cdots \wedge p_{n-1} \wedge \neg p_n) \mid n \in I\!N\}$ has only infinite models). We *can* guarantee satisfiability of *finite* sets. However, this requirement can be eased a little: what we can prove is that $\Gamma \models \Delta \Rightarrow \Gamma \vdash \Delta$ for those Γ and Δ for which the set of atoms in $\Gamma \cup \Delta$ is finite. To avoid cumbersome notation, from now on we simply assume that $\mathcal{P}$ itself is finite. We first show that this assumption implies that **L** is *logically finite*.

PROPOSITION 2.15 **L** is logically finite: there are only finitely many non-equivalent formulas.

Proof: From the proof of theorem 2.14 we can infer that every formula in $\mathcal{L}$ is equivalent to a semi-disjunctive normal form of modal degree ≤ 1. Since $\mathcal{P}$ is assumed to be finite, modulo logical equivalence there are only finitely many distinct formulas in $\mathcal{L}_0$. Thus there are only finitely many logically different choices for the $\alpha_{i,j}$, $\beta_{i,k}$ and π_i in the semi-disjunctive normal form displayed on page 10. Therefore there are only finitely many non-equivalent formulas. ∎

For 'Henkin-style' completeness proofs in classical (modal) logic the notion of maximal consistency is used [HC84]. A maximally consistent set of formulas is a consistent set which cannot be extended consistently. Maximally consistent sets facilitate completeness proofs since they form the syntactic counterparts of possible worlds. These sets enable the construction of a *canonical model*. The corresponding completeness proof amounts to demonstrating that the canonical model is a counterexample for any non-sequent of the associated logic.

Partial systems require a more flexible and general notion to mimic worlds in this way. The requirement of maximality disappears, which relates to the failure of the law of excluded middle in (most) partial systems. Sequentially formulated systems enable a short and general definition of 'sets-as-worlds'. In partial logic and other non-classical logic, the relevant concept is known as *saturation*.

DEFINITION 2.16 Let Σ, Δ and Ω be subsets of $\mathcal{L}$. Then:

- Σ is *consistent* iff $\Sigma \nvdash \emptyset$.
- $\Sigma \subseteq \mathcal{L}$ is *saturated* iff for every Δ: $\Sigma \vdash \Delta \Rightarrow \Sigma \cap \Delta \neq \emptyset$. $\mathcal{SAT}$ is the collection of all saturated sets (in **L**).

- Ω is a *saturator* of Σ iff $\Omega \cap \Delta \neq \emptyset$ for all Δ such that $\Sigma \vdash \Delta$. In such a case we write $\Sigma \lhd \Omega$.

$\Lambda \subseteq \mathcal{L}$ being a saturator of a set $\Gamma \subseteq \mathcal{L}$ boils down to non-derivability from Γ of the formulas not in Λ:

OBSERVATION 2.17

$$\Gamma \lhd \Lambda \;\Leftrightarrow\; \Gamma \nvdash \overline{\Lambda}.$$

The $\Rightarrow$-direction is a direct consequence of the definition of saturators ($\overline{\Lambda} \cap \Lambda = \emptyset$). The converse follows from the monotonicity rule. Suppose $\Gamma \not\lhd \Lambda$, then there exists Δ such that $\Gamma \vdash \Delta$ and $\Delta \cap \Lambda = \emptyset$. This means that $\Delta \subseteq \overline{\Lambda}$, and therefore, $\Gamma \vdash \overline{\Lambda}$.

The original definition of saturation goes back to [Acz68] and [Tho68], where it has been used to prove completeness for first order intuitionistic logic. Usually, saturation is presented as a combination of three properties: consistency, deductive closure and disjunctive saturation (i.e., $\Sigma \vdash \varphi \vee \psi \Rightarrow \Sigma \vdash \varphi$ or $\Sigma \vdash \psi$). These properties follow immediately by taking $\sharp\Delta = 0, 1$ and 2, respectively, in our sequential definition above.

The following three lemmas describe the 'Lindenbaum-part' of the Henkin construction. The standard Lindenbaum lemma says that every consistent set is a subset of a maximally consistent set. In completeness proofs of partial and constructive logics this lemma is not sufficient. In this kind of systems, we are often confronted with certain upper bounds in the syntactic construction of saturated sets. Fortunately, saturated sets are not necessarily maximally consistent and therefore the generalization of Lindenbaum's lemma given in lemma 2.19 turns out to be satisfactory for partial systems. It says that if such an upper bound is a saturator of a set of formulas, then a saturated set can be found which extends the original set and which respects the upper bound.

In order to prove this general Lindenbaum result, we give a short lemma which justifies the construction steps for saturated extensions within saturators.

LEMMA 2.18
If $\Gamma \lhd \Lambda$ and $\Gamma \vdash \Delta$ for certain finite set $\Delta \subseteq \mathcal{L}$, then there exists a $\delta \in \Delta$ such that $\Gamma \cup \{\delta\} \lhd \Lambda$.

Proof: Let $\Gamma \lhd \Lambda$ and $\Gamma \vdash \Delta$ with Δ finite, and suppose that $\Gamma \cup \{\delta\} \not\lhd \Lambda$ for all $\delta \in \Delta$. This means that for all $\delta \in \Delta$ there exists $\Sigma_\delta \subseteq \mathcal{L}$ such that

$$\Gamma, \delta \vdash \Sigma_\delta \quad \text{and} \quad \Sigma_\delta \cap \Lambda = \emptyset.$$

Let $\Sigma := \bigcup_{\delta \in \Delta} \Sigma_\delta$. MON yields $\Gamma, \delta \vdash \Sigma$ for all $\delta \in \Delta$. Application of CUT to this last sequent and the assumption $\Gamma \vdash \Delta$ yields $\Gamma \vdash \Delta \setminus \{\delta\}, \Sigma$. Repetition of CUT-application for all δ's eliminates the complete Δ from the last sequent. In all, $\Gamma \vdash \Sigma$. Because $\Gamma \lhd \Lambda$ we conclude $\Sigma \cap \Lambda \neq \emptyset$. This contradicts that $\Sigma_\delta \cap \Lambda = \emptyset$ for all $\delta \in \Delta$. ∎

LEMMA 2.19 *(saturation lemma)*
If $\Gamma \lhd \Lambda$, then there exists a saturated set Γ^* such that $\Gamma \subseteq \Gamma^* \subseteq \Lambda$.

Proof: Let $\Gamma \trianglelefteq \Lambda$ and let $\{\varphi_i\}_{i\in\mathbb{N}}$ be an enumeration of Λ. We define the following sequence of subsets of $\mathcal{L}_S$

$$\Gamma_0 := \Gamma$$

$$\Gamma_{n+1} := \begin{cases} \Gamma_n \cup \{\varphi_n\} & \text{if } \Gamma_n \cup \{\varphi_n\} \trianglelefteq \Lambda \\ \Gamma_n & \text{otherwise.} \end{cases}$$

We define Γ^* to be the limit of this sequence:

$$\Gamma^* := \bigcup_{n\in\mathbb{N}} \Gamma_n.$$

$\Gamma \subseteq \Gamma^* \subseteq \Lambda$ is immediately clear from the definition of Γ^* above. Another direct consequence of the construction above is $\Gamma_n \trianglelefteq \Lambda$ for all $n \in \mathbb{N}$. What is left to show is $\Gamma^* \in \mathcal{SAT}$.

Suppose $\Gamma^* \vdash \Delta$. We need to prove $\Gamma^* \cap \Delta \neq \emptyset$. The assumption set can be reduced to a finite sequence $\gamma_1, .., \gamma_m$ in Γ^* such that $\gamma_1, .., \gamma_m \vdash \Delta$. Because every member of Γ^* is a member of some Γ_i, this means that there exists Γ_k such that $\{\gamma_1, .., \gamma_m\} \subseteq \Gamma_k$,[6] and thus $\Gamma_k \vdash \Delta$ according MON. Since $\Gamma_k \trianglelefteq \Lambda$, we also have $\Delta \cap \Lambda \neq \emptyset$. Since $\Delta \subseteq \mathcal{L}$ has been picked arbitrarily as an conclusion set of Γ^* we have $\Gamma^* \trianglelefteq \Lambda$. This conclusion, combined with lemma 2.18, guarantees the existence of a formula $\delta \in \Delta$ such that

$$\Gamma^* \cup \{\delta\} \trianglelefteq \Lambda.$$

This result also ensures that $\Gamma_n \cup \{\delta\} \trianglelefteq \Lambda$ for all $n \in \mathbb{N}$, because all these sets are subsets of the limit set Γ^*.[7] Obviously, $\delta \in \Lambda$, which means that there exists $\ell \in \mathbb{N}$ such that $\varphi_\ell = \delta$. Because $\Gamma_\ell \cup \{\varphi_\ell\} \trianglelefteq \Lambda$, we know that $\delta \in \Gamma_{\ell+1}$ by the inductive definition of the sequence $\{\Gamma_n\}_{n\in\mathbb{N}}$. We conclude $\delta \in \Gamma^*$, and so $\Gamma^* \cap \Delta \neq \emptyset$. This establishes the desired result: $\Gamma^* \in \mathcal{SAT}$. ∎

The converse of this lemma is a trivial result. In other words, $\Gamma \trianglelefteq \Lambda$ is a precise criterion for Λ to contain a saturated extension of Γ. Note that the construction closely resembles the standard proof procedure for the Lindenbaum lemma. In fact, the classical Lindenbaum lemma is the special case where $\Lambda = \mathcal{L}$.[8]

Given the result in observation 2.17 the saturation lemma is equivalent to a more frequently used version, which we call the separation lemma.

LEMMA 2.20 *(separation lemma)*
If $\Sigma \nvdash \Delta$ then there exists a saturated set Γ such that $\Sigma \subseteq \Gamma$ and $\Delta \cap \Gamma = \emptyset$.

Proof: If $\Gamma \nvdash \Delta$ then $\Gamma \trianglelefteq \overline{\Delta}$ (observation 2.17). Lemma 2.19 shows that $\exists \Sigma \in \mathcal{SAT} : \Gamma \subseteq \Sigma \;\&\; \Sigma \subseteq \overline{\Delta}$. This last conjunct means that $\Sigma \cap \Delta = \emptyset$.[9] ∎

The reason to present the last two lemmas separately is purely pragmatic. The saturation lemma is convenient when we are restricted by upper bounds in the construction of a saturated set, and the separation lemma is useful when non-derivability comes on stage.

[6] Take for example $k = \max_{i\in\{1,..,n\}} \Gamma_{n_i}$ where $\{\Gamma_{n_i}\}_{i=1}^m$ is a subsequence of $\{\Gamma_n\}_{n\in\mathbb{N}}$ with $\gamma_i \in \Gamma_{n_i}$.

[7] By MON we have $\Sigma \trianglelefteq \Theta \;\&\; \Sigma' \subseteq \Sigma \Rightarrow \Sigma' \trianglelefteq \Theta$.

[8] Γ is consistent iff $\Gamma \trianglelefteq \mathcal{L}$.

[9] To see that the separation lemma also implies the saturation lemma, suppose that $\Gamma \trianglelefteq \Lambda$, which means $\Gamma \nvdash \overline{\Lambda}$ according to observation 2.17. Substituting $\overline{\Lambda}$ for Δ in the separation lemma gives us the desired result immediately.

DEFINITION 2.21 *(Canonical Model)*
Let Γ be a saturated set. We define the canonical model for Γ as $\mathcal{M}_\Gamma = \langle \mathcal{W}_\Gamma, \Gamma, \mathcal{V} \rangle$, where

- $\mathcal{W}_\Gamma = \{\Sigma \mid \Sigma \text{ is saturated and } \Box^-\Gamma \subseteq \Sigma \subseteq \Diamond^-\Gamma\}$
- For all $\Sigma \in \mathcal{W}_\Gamma \cup \{\Gamma\}$ and $p \in \mathcal{P}$: $\mathcal{V}(\Sigma)(p) = \begin{cases} 1 & \text{if } p \in \Sigma \\ 0 & \text{if } \neg p \in \Sigma \end{cases}$

LEMMA 2.22 The canonical model $\mathcal{M}_\Gamma$ is a balloon model.

Proof: (Cf. Definition 2.4)

1. $\mathcal{W}_\Gamma$ is finite by proposition 2.15.[10]
2. $\mathcal{V}$ is well-defined since saturated sets are consistent.
3. The root Γ is an extension of some world in the balloon, i.e. for some $\Sigma \in \mathcal{W}_\Gamma$ it holds that $\mathcal{V}(\Sigma) \sqsubseteq \mathcal{V}(\Gamma)$. To see this, we claim that $\Box^-\Gamma \trianglelefteq \Omega = \Diamond^-\Gamma \cap \Gamma$; then we are done, since then (by the saturation lemma) there is a saturated Σ such that $\Box^-\Gamma \subseteq \Sigma \subseteq \Diamond^-\Gamma \cap \Gamma$. Therefore $\Sigma \in \mathcal{W}_\Gamma$ and $\Sigma \subseteq \Gamma$, and so $\mathcal{V}(\Sigma) \sqsubseteq \mathcal{V}(\Gamma)$. The proof of the claim about Ω is as follows.

   By induction on finite $\Lambda \subseteq \mathcal{L}$ we prove that

   $$\Box^-\Gamma \vdash \Lambda \Rightarrow \Lambda \cap \Gamma \cap \Diamond^-\Gamma \neq \emptyset.$$

   Because $\Box^-\Gamma \nvdash \emptyset$,[11] the implication above holds trivially for $\Lambda = \emptyset$. So suppose $\Lambda = \{\lambda_1, \ldots, \lambda_n\}$, $(n \geq 1)$, and $\Box^-\Gamma \vdash \Lambda$. Then, by $n-1$ applications of R-TRUE $\vee$ we have, for each $(i \leq n)$, $\Box^-\Gamma \vdash (\lambda_1 \vee \cdots \vee \lambda_{i-1} \vee \lambda_{i+1} \vee \cdots \vee \lambda_n), \lambda_i$ and hence, by using R-TRUE $\Box$, we obtain

   $$\forall i \leq n : \Gamma \vdash \Box(\lambda_1 \vee \cdots \vee \lambda_{i-1} \vee \lambda_{i+1} \vee \cdots \vee \lambda_n), \Diamond\lambda_i$$

   Since Γ is saturated, we have two possibilities:

   - For some $i \leq n$ $\Box(\lambda_1 \vee \cdots \vee \lambda_{i-1} \vee \lambda_{i+1} \vee \cdots \vee \lambda_n) \in \Gamma$. Then $\Box^-\Gamma \vdash \Lambda \setminus \{\lambda_i\}$ and, by the induction hypothesis, $\Lambda \setminus \{\lambda_i\} \cap \Gamma \cap \Diamond^-\Gamma \neq \emptyset$, and hence $\Lambda \cap \Gamma \cap \Diamond^-\Gamma \neq \emptyset$.
   - For all $i \leq n, \Diamond\lambda_i \in \Gamma$. Then $\Lambda \subseteq \Diamond^-\Gamma$ (a), and, since $\Box^-\Gamma \vdash \Lambda$, by the L-TRUE $\Box$-rule, we have $\Gamma \vdash \Lambda$ and hence, by saturation of Γ, $\Gamma \cap \Lambda \neq \emptyset$ (b). Combining (a) and (b), we obtain $\Lambda \cap \Gamma \cap \Diamond^-\Gamma \neq \emptyset$.

■

[10] Notice this is virtually the only place where the specific (introspection) rules of **L** are used in the completeness proof. I.e. these rules license the special form of our balloon models.

[11] $\Box^-\Gamma \vdash \emptyset \Rightarrow \Box^-\Gamma \vdash \bot \Rightarrow \Gamma \vdash \Box\bot \Rightarrow \Gamma \vdash \bot$.

LEMMA 2.23 *(truth lemma)*
For all formulas $\varphi \in \mathcal{L}$, and all sets $\Gamma \in \mathcal{SAT}$ and each canonical model $\mathcal{M}_\Gamma$:

$$\mathcal{M}_\Gamma \models \varphi \Leftrightarrow \varphi \in \Gamma \qquad \mathcal{M}_\Gamma \models\!\!\!| \; \varphi \Leftrightarrow \neg\varphi \in \Gamma$$

Proof: By induction on φ: we only give the $\Box$-step. So we assume that $\varphi = \Box\psi$, while the induction hypothesis (IH) says that the lemma holds for ψ.
First we show the equivalence for $\models$:

($\Rightarrow$) If $\mathcal{M}_\Gamma \models \Box\psi$, then, by the truth definition of $\Box$, for all $\Delta \in \mathcal{W}_\Gamma : \mathcal{M}_\Delta \models \psi$. By IH, we conclude that for all $\Delta \in \mathcal{W}_\Gamma, \psi \in \Delta$. Now consider

$$\Gamma \vdash \Box\psi, \Diamond\overline{\Diamond^-\Gamma} \qquad (*)$$

After observing that $\Diamond\overline{\Diamond^-\Gamma} = \{\Diamond\gamma \mid \Diamond\gamma \notin \Gamma\}$, we claim that $(*)$ holds: for, suppose not, then by R-TRUE $\Box$ and MON we also have $\Box^-\Gamma \nvdash \psi, \overline{\Diamond^-\Gamma}$ and we use the separation lemma to find a Δ for which $\Box^-\Gamma \subseteq \Delta \subseteq \overline{\Diamond^-\Gamma}$, and $\psi \notin \Delta$, contradicting IH. Thus, since $(*)$ holds, we may use saturation of Γ to conclude that either $\Box\psi \in \Gamma$ or $\Gamma \cap \Diamond\overline{\Diamond^-\Gamma} \neq \emptyset$. Since the latter is impossible, we conclude that $\Box\psi \in \Gamma$.

($\Leftarrow$) Suppose $\Box\psi \in \Gamma$, and choose $\Delta \in \mathcal{W}_\Gamma$, which means that $\Delta \in \mathcal{SAT}$ and $\Box^-\Gamma \subseteq \Delta \subseteq \Diamond^-\Gamma$. We immediately find $\psi \in \Delta$ and, by IH, $\mathcal{M}_\Delta \models \psi$ so that $\mathcal{M}_\Gamma \models \Box\psi$.

Next the steps for $\models\!\!\!|$ are:

($\Rightarrow$) If $\mathcal{M}_\Gamma \models\!\!\!| \; \Box\psi$, then, by the falsity condition for $\Box$, for some $\Delta \in \mathcal{W}_\Gamma : \mathcal{M}_\Delta \models\!\!\!| \; \psi$, and, using IH, $\neg\psi \in \Delta$. Since $\Delta \subseteq \Diamond^-\Gamma$, $\Diamond\neg\psi \in \Gamma$, and, since Γ is deductively closed, $\neg\Box\psi \in \Gamma$.

($\Leftarrow$) Suppose $\neg\Box\psi \in \Gamma$. We claim that $\Box^-\Gamma \cup \{\neg\psi\} \trianglelefteq \Diamond^-\Gamma$. To see this, suppose that Θ is such that $\Box^-\Gamma \cup \{\neg\psi\} \vdash \Theta$, then, by L-FALSE $\Box$, also $\Box\Box^-\Gamma, \neg\Box\psi \vdash \Diamond\Theta$ and, by monotonicity, $\Gamma, \neg\Box\psi \vdash \Diamond\Theta$. Since $\neg\Box\psi$ is already a member of Γ, this implies $\Gamma \vdash \Diamond\Theta$. Now we use saturation of Γ to find a formula $\Diamond\theta$ in $\Gamma \cap \Diamond\Theta$, so $\theta \in \Diamond^-\Gamma \cap \Theta$. Now we have proven the claim, we use the saturation lemma to obtain a saturated set Δ with $\Box^-\Gamma \cup \{\neg\psi\} \subseteq \Delta \subseteq \Diamond^-\Gamma$. Clearly, $\Delta \in W_\Gamma, \neg\psi \in \Delta$, so we apply IH to conclude $\mathcal{M}_\Delta \models\!\!\!| \; \psi$. ∎

THEOREM 2.24 *(Completeness)* For all $\Sigma, \Delta \subseteq \mathcal{L}$, $\Sigma \models \Delta \Rightarrow \Sigma \vdash \Delta$.

Proof: Suppose $\Sigma \nvdash \Delta$, then, using the separation lemma we obtain a saturated set Γ for which $\Sigma \subseteq \Gamma$ and $\Gamma \cap \Delta = \emptyset$. Clearly, by lemma 2.23, $\mathcal{M}_\Gamma \models \Sigma$ and $\mathcal{M}_\Gamma \not\models \delta$ for all $\delta \in \Delta$, hence $\Sigma \not\models \Delta$ ∎

3. Honesty

This section concerns both the 'syntactic' and 'semantic' view on *circumscription* and *honesty*. Circumscribing the knowledge expressed by, say, φ, is to characterize what a rational agent knows when he or she *only* knows φ (together with its logical consequences). If such circumscription is possible, φ is called 'honest' in [HM85]. Though it may seem, *prima facie*, that circumscribing φ is always possible (by taking the deductive closure of $\Box\varphi$), this need not be the case. As we noticed earlier, the formula $\varphi = \Box p \vee \Box\neg p$ cannot be circumscribed, and is hence *dishonest*.

The main issue we want to address in this section, is the problem of deciding which formulas can be rendered honest. We will in fact present several notions of honesty in section 3.1, and illustrate them by means of a number of examples. Most of the technical justification for these examples, is provided after we have given a semantic account of the various notions of honesty, and are therefore postponed until section 3.2. In section 3.3 we supply inferential tests (so-called disjunction properties) which are particularly convenient for demonstrating that a formula is dishonest.

3.1 Stable Sets

We start out by investigating the deductive view on circumscription and honesty. Which criteria does the set $C_{\Box\varphi}$ consisting of the consequences of $\Box\varphi$ have to meet to consider φ honest? The crucial notion here is that of a *stable set*.[12] Although stability can be defined in many ways, the notion itself is stable, since various definitions turn out to be equivalent.

Thinking of $C_{\Box\varphi} = \{\psi \mid \Box\varphi \vdash \psi\}$ as the 'epistemic state' of a rational agent knowing only φ, it is clear that a stable set at least has to be a *consistent theory* (Cf. definition 2.16). In addition to being a consistent theory we want a stable set to have the property that the ignorance of non-consequences is compatible with the knowledge of consequences. In [Moo85] and [Jas91b] this leads to the following requirements for a stable set with respect to a normal modal system:

- S is a theory
- $\Box S \cup \neg\Box\overline{S}$ is consistent

Though correct for normal systems, the latter requirement is too strong for the partial logic we advocate. Recall from section 2.2 that our logic does not have any $\{\top, \bot\}$-free theorems. Yet we want to allow the set $S = C_{\Box\top}$ to be stable, characterizing the epistemic state of an agent knowing nothing. However, S is unstable by the second requirement: since $\Box\top \nvdash (p \vee \neg p)$, we have that $(p \vee \neg p) \in \overline{S}$, and therefore $\{\Box\top, \neg\Box(p \vee \neg p)\}$ would be consistent, which it is not. So we propose to replace the requirement above by the more general condition that knowledge of non-consequences does not follow from the initial knowledge.

Definition 3.1 *(Stability)*
A set S of formulas is *stable* if S is a theory for which $\Box S \nvdash \Box\overline{S}$

Notice that stable sets are consistent: suppose that S is an inconsistent theory. By the (derived) rule L-TRUE $\bot$ and the theoricity of S we have $S = \mathcal{L}$ and hence $\Box\overline{S} = \emptyset$. The

[12] See [Sta], [Moo85], [HM85] for **S5** stability. [Jas91b] defines stability for arbitrary normal systems. Our text definition is from [Thi92].

inconsistency of S implies that of $\Box S$, i.e. $\Box S \vdash \emptyset$, and therefore $\Box S \vdash \Box\overline{S}$, which means that S is not stable.

The insightful but somewhat esoteric definition 3.1 can be recast in a format which is closer to Stalnaker's original formulation:

PROPOSITION 3.2
S is a stable set of formulas iff

1. S is a theory
2. if $\varphi \in S$ then $\Box\varphi \in S$ (positive introspection)
3. if $\Box\varphi \vee \Box\psi \in S$ then $\varphi \in S$ or $\psi \in S$ (modal saturation)
4. $\varphi \notin S$ for some φ (consistency)

LEMMA 3.3 *(modal saturation)*
For all consistent theories S, modal saturation is equivalent to

$$S \vdash \Box\Gamma \Rightarrow S \cap \Gamma \neq \emptyset \text{ for all } \Gamma \subseteq \mathcal{L}$$

Proof: Modal saturation is obviously implied by the above requirement. For the other direction, suppose that $S \vdash \Box\Gamma$. First note that the consistency of S implies that $\Gamma \neq \emptyset$. By the finiteness of $\mathbf{L}$ we may assume that Γ is finite, say $\{\gamma_1, \ldots, \gamma_n\}$. So $S \vdash \Box\gamma_1 \vee \cdots \vee \Box\gamma_n$, therefore (by corollary 2.13) $S \vdash \Box(\Box\gamma_1 \vee \cdots \vee \Box\gamma_n)$, and thus (by lemma 2.12) $S \vdash \Box\gamma_1 \vee \Box(\Box\gamma_2 \vee \cdots \vee \Box\gamma_n)$. Therefore, by modal saturation, $\gamma_1 \in S$ or $\Box\gamma_2 \vee \cdots \vee \Box\gamma_n$. Repeating this argument it follows that for some i, $\gamma_i \in S$. ∎

Because of this equivalence, we will also refer to the elegant property displayed in lemma 3.3 by 'modal saturation'.

Proof of proposition 3.2:

($\Rightarrow$) Let S be a stable set. Then

1. by definition, S is a theory
2. suppose $\varphi \in S$ and $\Box\varphi \notin S$ then, since $\Box\varphi \vdash \Box\Box\varphi$, S violates the $\forall$ condition in definition 3.1.
3. suppose for some Γ, we have $S \vdash \Box\Gamma$, and $S \cap \Gamma = \emptyset$, then $\Gamma \subseteq \overline{S}$, so $\Box\Gamma \subseteq \Box\overline{S}$, so that, by monotonicity, $S \vdash \Box\overline{S}$ and, by L-TRUE $\Box$, $\Box S \vdash \Box\overline{S}$, contradicting the stability of S.
4. S is consistent, so $p \wedge \neg p \notin S$.

($\Leftarrow$) Next let S satisfy the conditions (1–4) then S is a theory, thus (by 4 and *ex falso*) consistent. Suppose $\Box S \vdash \Box\overline{S}$. By (2) $\Box S \subseteq S$, so by MON $S \vdash \Box\overline{S}$. Lemma 3.3 tells us that $S \cap \overline{S} \neq \emptyset$, a contradiction. ∎

Although the characterization of stability given by proposition 3.2 is useful, sometimes a more concise requirement is convenient. Since saturated sets are the possible worlds in the canonical model, the proposition essentially means that a stable set consists of all and only formulas known in some world.

PROPOSITION 3.4 S is stable iff $S = \Box^{-}\Gamma$ for some $\Gamma \in \mathcal{SAT}$.

Proof: ($\Rightarrow$) Let S be stable, then $\Box S \nvdash \Box\overline{S}$. By the separation lemma there is a saturated set Γ such that (i) $\Box S \subseteq \Gamma$ and (ii) $\Gamma \cap \Box\overline{S} = \emptyset$. Then $\varphi \in S \Rightarrow$ (by i) $\Box\varphi \in \Gamma \Rightarrow \varphi \in \Box^{-}\Gamma$ and $\varphi \notin S \Rightarrow \Box\varphi \in \Box\overline{S} \Rightarrow$ (by ii) $\Box\varphi \notin \Gamma \Rightarrow \varphi \notin \Box^{-}\Gamma$. Hence $S = \Box^{-}\Gamma$. ($\Leftarrow$) Suppose $S = \Box^{-}\Gamma$ for some saturated set Γ, and also that $\Box S \vdash \Box\overline{S}$. Since $\Box S \subseteq \Gamma$, and using MON we have $\Gamma \vdash \Box\overline{S}$. Γ is saturated, and hence there is some $\psi \notin S$ with $\Gamma \vdash \Box\psi$. But then, since Γ is deductively closed, we have $\Box\psi \in \Gamma$ and hence $\psi \in S$, a contradiction. ∎

Having characterized stability in different ways, we are ready for a formal account of circumscription and honesty. Writing $\mathcal{ST}(\varphi)$ for $\{S \subseteq \mathcal{L} \mid \varphi \in S \;\&\; S \text{ is stable }\}$, circumscription of knowledge of φ involves finding a minimal element in $\mathcal{ST}(\varphi)$, the set of stable expansions of φ. If there is a stable set which is minimum, according to some order on sets of formulas, the knowledge is honest. What is this ordering relation? In the paradigm case of the (total) system **S5**, different stable sets are incomparable, so set inclusion does not work. This is not the case for the present (partial) system, basically because the notorious Stalnaker condition $\varphi \notin S \Rightarrow \neg\Box\varphi \in S$ does not hold for stable sets in partial logic. The invalidity of the latter condition implies that in **L** a stable set is not determined by its propositional content (the purely propositional formulas in it), although a stable set is determined by its formulas of degree 1 (i.e. with modal depth less or equal to 1), by theorem 2.14. This might suggest set inclusion as the ordering relation of the stable sets, and a definition of honesty induced by $\subseteq$: basically existence of a smallest stable expansion.

DEFINITION 3.5 *(Naïve Honesty)*
φ is called *naïvely honest* if there is a $\subseteq$-minimal element in $\mathcal{ST}(\varphi)$.

EXAMPLE 3.6 The formulas p, $p \wedge q$, $\Box p$, $\Box(p \wedge q)$, $\Diamond p$ and $\Diamond(p \wedge q)$ are naïvely honest.

Can we give other sufficient and necessary conditions for naïve honesty? To this purpose reinspect $C_{\Box\varphi} = \{\psi \mid \Box\varphi \vdash \psi\}$. Observe that

- $C_{\Box\varphi}$ is a *theory*, since $\vdash$ is transitive for single formulas;
- $C_{\Box\varphi}$ is contained in every stable set containing φ: let $\varphi \in S$ for some stable S, then by proposition 3.2(2) $\Box\varphi \in S$, so, by proposition 3.2(1) S contains all the consequences of $\Box\varphi$, i.e. $C_{\Box\varphi} \subseteq S$.

As an easy result, we now present a necessary and sufficient condition for a stable set to be $\subseteq$-minimal.

THEOREM 3.7 A set S is $\subseteq$-minimal in $\mathcal{ST}(\varphi)$ iff $S = C_{\Box\varphi}$ is stable.

Proof: ($\Rightarrow$) Suppose S is $\subseteq$-minimal for φ. By definition $\varphi \in S$, and, by the remark above, $C_{\Box\varphi} \subseteq S$. Now suppose that $S \not\subseteq C_{\Box\varphi}$, then we have a ψ with $\psi \in S$ and $\Box\varphi \nvdash \psi$. The separation lemma then provides a saturated set Γ for which $\Box\varphi \in \Gamma, \psi \notin \Gamma$. Since $\Box\psi \vdash \psi$ and Γ is a theory, we also have $\Box\psi \notin \Gamma$. By proposition 3.4, $\Box^{-}\Gamma$ is a stable set containing φ, contradicting the $\subseteq$-minimality of S.
($\Leftarrow$) If $C_{\Box\varphi}$ is a stable set, then, by the remarks above, it must be $\subseteq$-minimal. ∎

The theorem above immediately provides a necessary and sufficient condition for naïve honesty.

COROLLARY 3.8 φ is naïvely honest iff $C_{\Box\varphi}$ is stable.

Proof: Let $C_{\Box\varphi}$ be stable. By L-TRUE $\Box$, $\varphi \in C_{\Box\varphi}$. By theorem 3.7, $C_{\Box\varphi}$ is also $\subseteq$-minimal for φ, implying that φ is naïvely honest. The other direction is obvious. ∎

EXAMPLE 3.9 The objective formula $p \vee q$ is not naïvely honest. For suppose that S would be $\subseteq$-minimal in $\mathcal{ST}(p \vee q)$, then $(p \vee q) \in S$, and, by proposition 3.2(2), also $\Box(p \vee q) \in S$. Since (by R-TRUE $\Box$ and $\Diamond \Rightarrow \Box\Diamond$)), we have $\Box(p \vee q) \vdash \Box p \vee \Box\Diamond q$, we use proposition 3.2(3) to conclude that either $p \in S$ or $\Diamond q \in S$ (*). Now, let $\Sigma_1 = \{\Box p\}$ and $\Sigma_2 = \{\Box q\}$. Using completeness, we immediately see that $\Sigma_1 \nvdash \Box\Diamond q$ and $\Sigma_2 \nvdash \Box p$. The separation lemma then guarantees the existence of saturated sets Γ_1, Γ_2 for which $\Sigma_i \subseteq \Gamma_i (i = 1, 2)$, $\Box\Diamond q \notin \Gamma_1$ and $\Box p \notin \Gamma_2$. By proposition 3.4 we find two stable sets $S_i = \Box^{-}\Gamma_i, (i = 1, 2), S_i \in \mathcal{ST}(p \vee q)$, for which $\Diamond q \notin S_1$ and $p \notin S_2$. Since S is $\subseteq$-minimal in $\mathcal{ST}(p \vee q)$ we find $p \notin S, \Diamond q \notin S$, contradicting (*).

Intuition says that all *objective* (i.e. propositional) formulas should be rendered honest: it seems to be perfectly sensible to claim to only know some objective information. Together with example 3.9 we see that the definition of naïve honesty is too strong, and also too naïve. Thus, though the set inclusion ordering of stable sets is (non-trivially) possible, it produces wrong results as far as honesty is concerned. Now one alternative is to replace ordinary set inclusion by the relation of epistemic inclusion $\subseteq_{\Box}$. This, however, will not produce any new results, due to the following observation.

OBSERVATION 3.10 For all stable sets Γ, Δ: $\Gamma \subseteq_{\Box} \Delta \Leftrightarrow \Gamma \subseteq \Delta$.

Somewhat surprisingly, since its propositional content does not determine a stable set, propositional minimality of a stable expansion produces a more adequate notion of honesty. In this way, we obtain essentially the same definition of honesty as was proposed in [HM85]. However, in **L** one can generally derive less conclusions from the circumscription of a weakly honest formula than in the **S5** case. For example, $\Diamond p$ is honest in **S5**, and also weakly honest in **L**, but 'knowing only $\Diamond p$' entails different conclusions in both set-ups.

DEFINITION 3.11 *(Weak Honesty)*
φ is *weakly honest* if there is a $\subseteq_0$-minimal element in $\mathcal{ST}(\varphi)$.

OBSERVATION 3.12 All naïvely honest formulas are weakly honest.

EXAMPLE 3.13 So p is weakly honest. The disjunction $p \vee q$ is also weakly honest: more generally, for each consistent objective formula π, π itself, $\Box\pi$ and $\Diamond\pi$ are weakly honest. Other examples of weakly honest formulas are $(\Box p \wedge \Diamond q)$, and disjunctions such as $\Box p \vee \neg\Box p$ and $p \vee \neg\Box p$. The formula $\Box p \vee \Box q$ is not weakly honest, nor is $\Box p \vee \neg p$. (This will be proved in section 3.3.)

Notice that a propositionally smallest stable expansion for some formula need not be unique: $S \cap \mathcal{L}_0$ does not determine S. For example, in the case of $p \vee q$, S may or may not contain $\Diamond(p \wedge q)$.

THEOREM 3.14 A set S is $\subseteq_0$-minimal in $\mathcal{ST}(\varphi)$ iff $S \in \mathcal{ST}(\varphi)$ and $S_0 = C_{\Box\varphi} = \{\mu \in \mathcal{L}_0 \mid \Box\varphi \vdash \mu\}$.

Proof: Omitted; essentially the same as the proof of theorem 3.7. ■

COROLLARY 3.15 φ is weakly honest iff $S_0 = (C_{\Box\varphi})_0$ for some stable expansion S of φ.

As we noticed in the introduction, a similar argument that motivates that $(\Box p \vee \Box q)$ should not be considered honest, can be applied to $(\Diamond p \vee \Diamond q)$: neither can be circumscribed, intuitively spoken. This is why the current notion of honesty is too weak:

EXAMPLE 3.16 The formula $\Diamond p \vee \Diamond q$ is weakly honest.[13] This will be shown in section 3.2.

Analyzing the reason for this counter-intuitive result, we note that for weak honesty we did minimize the *objective* formulas in the stable set for φ, but not the *possibilities* contained in it. In fact, $\subseteq_0$-minimality is insufficiently restrictive: among the $\subseteq_0$-minimal stable sets, we want to single out those containing the smallest number of epistemic possibilities. This is achieved in our last notion of honesty:

DEFINITION 3.17 *(Strong Honesty)*
A formula φ is called strongly honest if S is *strongly minimal* for φ, i.e. there is a $\subseteq_\Diamond$-minimal element in $\{S \subseteq \mathcal{L} \mid S \text{ is } \subseteq_0\text{-minimal in } \mathcal{ST}(\varphi)\}$.

EXAMPLE 3.18 Now, $\Diamond p \vee \Diamond q$ is not strongly honest. As with weak honesty, for each objective formula $\pi \in \mathcal{L}_0$, the formulas π and $\Box\pi$ are strongly honest (this follows from the observation below), but now, $\Diamond(p \vee q)$ is not strongly honest (for a proof of the latter fact, see section 3.2).

OBSERVATION 3.19 All naïvely honest formulas are strongly honest, and all strongly honest formulas are weakly honest.

In order to characterize the stable sets that contain a strongly honest formula, we need a lemma that we will not prove until section 3.2, and one more definition.

LEMMA 3.20 Let S and S' be stable sets such that $\Box^- S \subseteq_0 \Box^- S'$ and $\Diamond^- S \subseteq_0 \Diamond^- S'$. Then $S \subseteq S'$.

DEFINITION 3.21 For a formula φ we define its *diamond remainder* $R^\Diamond_{\Box\varphi}$ as:

$$R^\Diamond_{\Box\varphi} = \{\Diamond\mu \in \Diamond(\mathcal{L}_0) \mid \Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\mu\}$$

In words, $R^\Diamond_{\Box\varphi}$ contains $\Diamond$-formulas with a propositional argument that are derivable from $\Box\varphi$, in disjunction with those $\Box$-formulas of which the argument is propositional and not a consequence of $\Box\varphi$.

THEOREM 3.22 A set S is strongly minimal for φ iff $S_0 = (C_{\Box\varphi})_0$ and $S \cap \Diamond\mathcal{L}_0 = R^\Diamond_{\Box\varphi}$ and $S \in \mathcal{ST}(\varphi)$.

[13] Notice that $\Diamond p \vee \Diamond q$ is also honest in the analysis of **S5** as given in [HM85].

Proof:
($\Rightarrow$) Let S be $\subseteq_\Diamond$-minimal in $\{S \subseteq \mathcal{L} \mid S$ is $\subseteq_0$ -minimal in $\mathcal{ST}(\varphi)\}$. $\subseteq_0$-minimality of S for φ implies $S_0 = (C_{\Box\varphi})_0$ (theorem 3.14). In order to show that $R^\Diamond_{\Box\varphi} \subseteq S \cap \Diamond\mathcal{L}_0$, suppose $\Diamond\mu \in R^\Diamond_{\Box\varphi}$. Then

$$\Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\mu. \tag{3.1}$$

Proposition 3.4 shows that there exists a $\Gamma \in \mathcal{SAT}$ such that $S = \Box^-\Gamma$. Because $\Box\varphi \in \Gamma$ and (3.1):

$$\Gamma \cap (\Box(\overline{C_{\Box\varphi}})_0 \cup \{\Diamond\mu\}) \neq \emptyset. \tag{3.2}$$

Theorem 3.14 and the $\subseteq_0$-minimality of S entail $\Gamma \cap \Box(\overline{C_{\Box\varphi}})_0 = \emptyset$, which means $\Diamond\mu \in \Gamma$ (3.2). Since $\Diamond\mu \vdash \Box\Diamond\mu$, we find $\Diamond\mu \in S \cap \Diamond\mathcal{L}_0$. The arbitrariness of $\Diamond\mu \in R^\Diamond_{\Box\varphi}$ now guarantees $R^\Diamond_{\Box\varphi} \subseteq S \cap \Diamond\mathcal{L}_0$.

To see that also $S \cap \Diamond\mathcal{L}_0 \subseteq R^\Diamond_{\Box\varphi}$, suppose $\Diamond\nu \notin R^\Diamond_{\Box\varphi}$ with $\nu \in \mathcal{L}_0$, then:

$$\Box\varphi \nvdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\nu. \tag{3.3}$$

Using the separation lemma, we find a $\Gamma \in \mathcal{SAT}$ for which $\Box\varphi \in \Gamma$, $\Gamma \cap \Box(\overline{C_{\Box\varphi}})_0 = \emptyset$ and $\Diamond\nu \notin \Gamma$. The first two conclusions imply that $\Box^-\Gamma$ is a $\subseteq_0$-minimal stable set for φ (Cf. theorem 3.14). The last conclusion entails $\Diamond\nu \notin S$, because S is $\subseteq_\Diamond$-minimal among the $\subseteq_0$-minimal stable sets.

($\Leftarrow$) Suppose that both $S \cap \mathcal{L}_0 = (C_{\Box\varphi})_0$ and $S \cap \Diamond\mathcal{L}_0 = R^\Diamond_{\Box\varphi}$ for some $S \in \mathcal{ST}(\varphi)$. Let S' be an arbitrary stable set for φ that is $\subseteq_0$-minimal. We have to show that

$$S \subseteq_\Diamond S'. \tag{3.4}$$

By lemma 3.20 it is sufficient to show that both $\Box^-S \subseteq_0 \Box^-S'$ and $\Diamond^-S \subseteq_0 \Diamond^-S'$. Since $S \cap \mathcal{L}_0 = (C_{\Box\varphi})_0$, by theorem 3.14 we have $S \subseteq_0 S'$ so in particular, $\Box^-S \subseteq_0 \Box^-S'$. What is left to prove is $\Diamond^-S \subseteq_0 \Diamond^-S'$.
Suppose that we have some $\mu \in \mathcal{L}_0$ with $\Diamond\mu \in S$. Since $S \cap \Diamond\mathcal{L}_0 = R^\Diamond_{\Box\varphi}$:

$$\Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\mu. \tag{3.5}$$

By $\subseteq_0$-minimality of S' for φ and theorem 3.14 we obtain $S' \cap \mathcal{L}_0 = (C_{\Box\varphi})_0$, hence, $S' \cap (\overline{C_{\Box\varphi}})_0 = \emptyset$ and $\Gamma' \cap \Box(\overline{C_{\Box\varphi}})_0 = \emptyset$ for certain $\Gamma' \in \mathcal{SAT}$ with $S' = \Box^-\Gamma'$. This implies $\Diamond\mu \in \Gamma'$ (3.5), and, (since $\Diamond\mu \vdash \Box\Diamond\mu$) also $\Diamond\mu \in \Box^-\Gamma' = S'$. ∎

COROLLARY 3.23 φ is strongly honest iff there is an $S \in \mathcal{ST}(\varphi)$ with:
$S_0 = (C_{\Box\varphi})_0$ and $(\Diamond^-S)_0 = \{\mu \mid \exists\psi_i \in \mathcal{L}_0 : \Box\varphi \vdash \Box\psi_1 \vee \cdots \vee \Box\psi_n \vee \Diamond\mu \,\&\Box\varphi \nvdash \psi_i\}$.

3.2 Minimal Models

Proposition 3.4 ties up the notion of stable set with a main semantic notion: recall that saturated sets correspond to partial worlds in the canonical model. The following corollary of proposition 3.4 relates stability directly to the knowledge in a balloon model.

COROLLARY 3.24 S is stable iff $S = \kappa(M)$ for some balloon model M.

In order to decide whether some formula is honest, we considered stable sets that were minimal in some sense. Combined with corollary 3.24 the following orders on models emerge.

DEFINITION 3.25 For any two models $M = \langle W, g, V\rangle$ and $M' = \langle W', g', V'\rangle$ we define:

- $M \sqsubseteq_\Box M' \Leftrightarrow \forall w' \in W'\ \exists w \in W : V(w) \sqsubseteq V'(w')$
- $M \sqsubseteq_\Diamond M' \Leftrightarrow \forall w \in W\ \exists w' \in W' : V(w) \sqsubseteq V'(w')$
- $M \sqsubseteq M' \Leftrightarrow (M \sqsubseteq_\Box M' \ \&\ M \sqsubseteq_\Diamond M')$
- For any $\preceq \in \{ \sqsubseteq_\Box, \sqsubseteq_\Diamond, \sqsubseteq \}$, we say that a model M is $\preceq$-minimal for φ if $\varphi \in \kappa(M)$ and for all M' with $\varphi \in \kappa(M')$ it holds that $M \preceq M'$. We then say that φ has a $\preceq$-minimal model.

The above orders are familiar from *domain theory* see e.g. [Sto77]; they are known as the Smyth, Hoare and Egli-Milner order, respectively. The orders do not specify anything about the root g of a model $M = \langle W, g, V\rangle$. Recall that $Th(M) = \{\ \varphi \in \mathcal{L} \mid M \models \varphi\ \}$, that $\kappa(M) = \Box^- Th(M)$ and $\pi(M) = \Diamond^- Th(M)$. This is how $\subseteq_\star$ and $\sqsubseteq_\star$ are related:

PROPOSITION 3.26 Consider two balloon models $M = \langle W, g, V\rangle$ and $M' = \langle W', g', V'\rangle$.

1. $M \sqsubseteq_\Box M' \ \Leftrightarrow\ Th(M) \cap \Box\mathcal{L}_0 \subseteq_\Box Th(M') \ \Leftrightarrow\ \kappa(M) \subseteq_0 \kappa(M')$
2. $M \sqsubseteq_\Diamond M' \ \Leftrightarrow\ Th(M) \cap \Diamond\mathcal{L}_0 \subseteq_\Diamond Th(M') \ \Leftrightarrow\ \pi(M) \subseteq_0 \pi(M')$
3. $M \sqsubseteq M' \ \&\ V(g) \sqsubseteq V'(g') \ \Leftrightarrow\ Th(M) \subseteq Th(M')$
4. $M \sqsubseteq M' \Leftrightarrow \kappa(M) \subseteq \kappa(M') \Leftrightarrow \pi(M) \subseteq \pi(M')$

Proof: We only prove the first item *in extenso*, the second is proven similarly, whereas the third follows from definition 2.6 and theorem 2.7. The facts in the last item can be deduced from the others, using the degree 1 normal forms from the proof of theorem 2.14 for the first equivalence.

So, suppose that $M \sqsubseteq_\Box M'$ and let μ be some propositional formula for which $M \models \Box\mu$, i.e. for all $w \in W : M_w \models \mu$. Choose any $v' \in W'$. Since $M \sqsubseteq_\Box M'$ there is a $v \in W$ such that $V(v) \sqsubseteq V'(v')$, and so, since $\mu \in \mathcal{L}_0$ we use propositional persistence (proposition 2.5) to conclude $M'_{v'} \models \mu$. Since v' was arbitrary, we have $M' \models \Box\mu$. The opposite direction is proved using contraposition: if $M \not\sqsubseteq_\Box M'$, then there is some $w' \in W'$ such that for all $w \in W$: $V(w) \not\sqsubseteq V'(w')$. So, for each $w \in W$ there is a literal $\alpha_w \in \mathcal{P} \cup \neg\mathcal{P}$ such that $M_w \models \alpha_w$ and $M'_{w'} \not\models \alpha_w$. Now if $\alpha = \bigvee_{w \in W} \alpha_w$, obviously $\alpha \in \mathcal{L}_0$ and $M_w \models \alpha$ for all

$w \in W$, so $M \models \Box\alpha$, yet $M'_{w'} \not\models \alpha$, so $M' \not\models \Box\alpha$. Therefore $\kappa(M) \cap \mathcal{L}_0 \not\subseteq \kappa(M')$, i.e. $\kappa(M) \not\subseteq_0 \kappa(M')$. ■

It is not hard to see that the restrictions to $\mathcal{L}_0$ in the proposition are necessary. In the case of $\sqsubseteq_\Box$ for instance, let $M' = \langle W', g', V' \rangle$ such that $V'(w')(p) = 0$ for all $w' \in W'$. Consider $M = \langle W, g, V \rangle$ with $W = W' \cup \{x\}$ for some $x \notin W'$, $V(x)(p) = 1$ and $V(w') = V'(w')$ for all $w' \in W'$. Although $M \sqsubseteq_\Box M'$, we have $M \models \Box\Diamond p$, but at the same time $M' \not\models \Box\Diamond p$.

Lemma 3.20 (repeated)
Let S and S' be stable sets such that $\Box^- S \subseteq_0 \Box^- S'$ and $\Diamond^- S \subseteq_0 \Diamond^- S'$. Then $S \subseteq S'$.

Proof: Let M and M' be such that $S = \kappa(M), S' = \kappa(M')$. Applying the first two items of proposition 3.26, we obtain $M \sqsubseteq_\Box M'$ and $M \sqsubseteq_\Diamond M$, hence $M \sqsubseteq M'$ and thus, by the last item of the same theorem, $S \subseteq S'$. ■

Now we can characterize our different notions of honesty in semantic terms.

THEOREM 3.27 φ is naïvely honest iff $\Box\varphi$ has a $\sqsubseteq$-minimal model.

Proof: Using corollary 3.24 and proposition 3.26, the argument is straightforward:

φ is naively honest	$\Leftrightarrow$	(definition)
$\exists S \in \mathcal{ST}(\varphi)\ \forall S' \in \mathcal{ST}(\varphi) : S \subseteq S'$	$\Leftrightarrow$	(cor. 3.24)
$\exists M \forall M' : \varphi \in \kappa(M)\ \&\ \varphi \in \kappa(M') \Rightarrow \kappa(M) \subseteq \kappa(M')$	$\Leftrightarrow$	(def. κ, prop. 3.26)
$\exists M \forall M' : M \models \Box\varphi\ \&\ M' \models \Box\varphi \Rightarrow M \sqsubseteq M'$	$\Leftrightarrow$	(def. 3.25)
$\exists M$ which is $\sqsubseteq$-minimal for $\Box\varphi$		

■

EXAMPLE 3.28 Figure 1 gives $\sqsubseteq$-minimal models for $p \wedge q$, $\Diamond(p \wedge q)$ and $\Box(p \wedge q)$, respectively. As was announced in example 3.6, these formulas are thus naïvely honest by virtue of the M' and M''. Note that the existence of an empty balloon world such as in M and M' makes a model immediately $\sqsubseteq_\Box$-minimal.

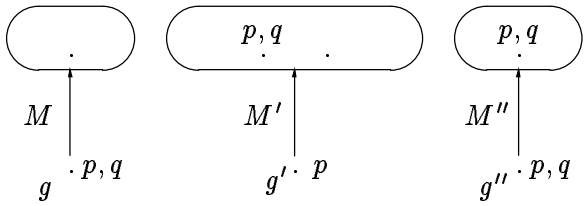

Figure 1: Three $\sqsubseteq$-minimal models

THEOREM 3.29 φ is weakly honest iff $\Box\varphi$ has a $\sqsubseteq_\Box$-minimal model.

Proof: Again a direct argument is possible:

φ is weakly honest	$\Leftrightarrow$	(def. weak honesty)
$\exists S \in \mathcal{ST}(\varphi)\ \forall S' \in \mathcal{ST}(\varphi) : S \subseteq_0 S'$	$\Leftrightarrow$	(cor. 3.24)
$\exists M \forall M' : \varphi \in \kappa(M)\ \&\ \varphi \in \kappa(M') \Rightarrow \kappa(M) \subseteq_0 \kappa(M')$	$\Leftrightarrow$	(def. κ, prop. 3.26)
$\exists M \forall M' : M \models \Box\varphi\ \&\ M' \models \Box\varphi \Rightarrow M \sqsubseteq_\Box M'$	$\Leftrightarrow$	(def. 3.25)
$\exists M$ which is $\sqsubseteq_\Box$-minimal for $\Box\varphi$		

■

EXAMPLE 3.30 The models M and M' of figure 2 are $\sqsubseteq_\Box$-minimal for $\Box(p \vee q)$ and $\Box p \vee \neg\Box p$, respectively. Moreover, both models are $\sqsubseteq_\Box$-minimal for $\Box(\Diamond p \vee \Diamond q)$, thus proving that $\Diamond p \vee \Diamond q$ is weakly honest. This justifies claims of weak honesty made in examples 3.13 and 3.16.

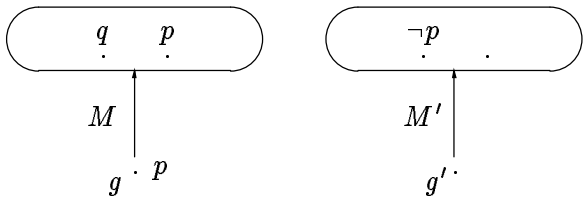

Figure 2: Two $\sqsubseteq_\Box$-minimal models

Connecting strong honesty with a semantic notion requires one more definition.

DEFINITION 3.31 A model M is called *strongly minimal for* φ if M is $\sqsubseteq_\Diamond$-minimal in the set $\{M' \mid M' \text{ is } \sqsubseteq_\Box \text{-minimal for } \varphi\}$.

Note that strongly minimal models for φ are by definition $\sqsubseteq_\Box$-minimal for φ. Also note, however, that a strongly minimal model need not be $\sqsubseteq_\Diamond$-minimal.

THEOREM 3.32 φ is strongly honest iff $\Box\varphi$ has a strongly minimal model.

Proof:
($\Rightarrow$) Assume S to be $\subseteq_\Diamond$-minimal amongst the $\subseteq_0$-minimal stable sets for φ, and let M be a model for which $S = \kappa(M)$ (corollary 3.24). We will show that M is strongly minimal for φ, i.e. $M \sqsubseteq_\Diamond M'$ for any M' that is $\sqsubseteq_\Box$-minimal for φ. Consider such an M'. Then, as in the proof of theorem 3.29, $\kappa(M')$ is $\subseteq_0$-minimal in $\mathcal{ST}(\varphi)$. So, by assumption, $\kappa(M) \subseteq_\Diamond \kappa(M')$. To draw the required conclusion $M \sqsubseteq_\Diamond M'$, it suffices, because of proposition 3.26 (2), to show that $\pi(M) \subseteq_0 \pi(M')$. This is straightforward: $\alpha \in \pi(M) \cap \mathcal{L}_0 \Rightarrow M \models \Diamond\alpha \Rightarrow M \models \Box\Diamond\alpha \Rightarrow \Diamond\alpha \in \kappa(M) \Rightarrow \Diamond\alpha \in \kappa(M') \Rightarrow M' \models \Box\Diamond\alpha \Rightarrow M' \models \Diamond\alpha \Rightarrow \alpha \in \pi(M')$.
($\Leftarrow$) Let M be strongly minimal for $\Box\varphi$, and $S = \kappa(M)$. If S' is some $\subseteq_0$-minimal stable set for φ, the proof of theorem 3.29 shows that the model M' for which $S' = \kappa(M')$ is $\sqsubseteq_\Box$-minimal for $\Box\varphi$. We have to show that $S \subseteq_\Diamond S'$. Since M was strongly minimal for $\Box\varphi$, we have that $M \sqsubseteq_\Box M'$ and $M \sqsubseteq_\Diamond M'$, i.e. $M \sqsubseteq M'$. Hence, by proposition 3.26 $\kappa(M) \subseteq \kappa(M')$, i.e. $S \subseteq S'$, and so $\Box\Diamond\psi \in Th(M)$, then $S \subseteq_\Diamond S'$. ■

EXAMPLE 3.33 We argue that $\Diamond p \vee \Diamond q$ is not strongly honest, proving the claim made in example 3.18: consider the two models M and M' that both have a balloon with two worlds, one being the empty world; moreover M has a p-world (i.e. a world in which p is true) that does not verify q, where M' has a q-world that does not verify p. Both models verify $\Box(\Diamond p \vee \Diamond q)$ and the empty world guarantees that they are $\sqsubseteq_\Box$-minimal for $\Box(\Diamond p \vee \Diamond q)$. But then we also see that there can be no model N for $\Box(\Diamond p \vee \Diamond q)$ for which both $N \sqsubseteq_\Diamond M$ and $N \sqsubseteq_\Diamond M'$: such a model N has to contain at least a p- or a q-world, if it has a p-world then $N \not\sqsubseteq_\Diamond M'$, if it has a q-world, then $N \not\sqsubseteq_\Diamond M$.

3.3 Disjunction Properties

One might want to have an even more direct condition providing honesty, without interference of the notion of stability. Here, we will provide several syntactic, or perhaps rather deductive characterizations for honesty. Inspecting the properties of saturated and stable sets, one good candidate for this is the *disjunction property*, defined below. In fact, this property is already mentioned in [HC84], be it that there it is a property of logical systems, rather than of formulas. In partial logic the property should be slightly reformulated, and adapted to the different notions of honesty.

DEFINITION 3.34 *(Disjunction Properties)*

Let $\varphi \in \mathcal{L}$. The following conditions determine when φ has the *disjunction property* (DP), the *propositional disjunction property* (PDP) or *propositional diamond disjunction property* (PDDP), respectively.

DP $\quad \forall \Sigma \subseteq \mathcal{L}: \quad \Box\varphi \vdash \Box\Sigma \Rightarrow \exists \sigma \in \Sigma : \Box\varphi \vdash \sigma$

PDP $\quad \forall \Pi \subseteq \mathcal{L}_0: \quad \Box\varphi \vdash \Box\Pi \Rightarrow \exists \pi \in \Pi : \Box\varphi \vdash \pi$

PDDP $\quad \forall \Pi \subseteq \mathcal{L}_0: \quad \Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\Pi \Rightarrow \exists \pi \in \Pi : \Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\pi$

OBSERVATION 3.35

- All disjunction properties imply consistency. Take $\Pi, \Sigma = \emptyset$ for the arguments Π, Σ in the rules of the definition above.
- Note that $\Box\varphi \vdash \psi \Leftrightarrow \Box\varphi \vdash \Box\psi$ for all $\psi \in \mathcal{L}$. Furthermore, $\Box\varphi \vdash \Box\Delta, \Diamond\psi \Leftrightarrow \Box\varphi \vdash \Box\Delta, \Box\Diamond\psi$ for every formula ψ. These are simple consequences of corollary 2.13. This observation facilitates proving the next theorem about the relation between different notions of honesty and the various disjunction properties which were presented in the definition above.

THEOREM 3.36 *(Disjunction properties and honesty)*

φ has the DP $\quad \Leftrightarrow \quad \varphi$ is naïvely honest

φ has the PDP $\quad \Leftrightarrow \quad \varphi$ is weakly honest

φ has the PDDP $\quad \Leftrightarrow \quad \varphi$ is strongly honest

Proof: We start by proving the $\Leftarrow$-direction for the stated equivalences. These are in fact almost immediate consequences of the modal saturation property of stable sets (Cf. proposition 3.2 item 3) and the various characterizations of minimal stable sets.

- Let φ be naïvely honest. This means it has a $\subseteq$-minimal stable set S. Now suppose $\Box\varphi \vdash \Box\Sigma$, then $S \vdash \Box\Sigma$. By modal saturation we know $S \cap \Sigma \neq \emptyset$. According to theorem 3.7, $S = C_{\Box\varphi}$, so for some $\sigma \in \Sigma$, $\Box\varphi \vdash \sigma$. In other words, φ has the disjunction property.
- If φ is weakly honest, there is a similarly straightforward proof that φ has the PDP, since by theorem 3.14: $\exists S \in \mathcal{ST}(\varphi) : S_0 = (C_{\Box\varphi})_0$.

- Let φ be strongly honest. Suppose $\Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\Pi$ for certain $\Pi \subseteq \mathcal{L}_0$. The second item in observation 3.35 tells us that

  $$\Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Box\Diamond\Pi.$$

  Let S be $\subseteq_\Diamond$-minimal amongst the $\subseteq_0$-minima of $\mathcal{ST}(\varphi)$. Again modal saturation shows that there exists a $\varrho \in (\overline{C_{\Box\varphi}})_0 \cup \Diamond\Pi$ such that $\varrho \in S$. On account of theorem 3.22 we know that $S_0 = (C_{\Box\varphi})_0$, so ϱ must be some $\Diamond\pi$ in $\Diamond\Pi$. We also know from 3.22 that $S \cap \Diamond\mathcal{L}_0 = R^\Diamond_{\Box\varphi}$, so $\Diamond\pi \in R^\Diamond_{\Box\varphi}$. By definition of the diamond remainder of $\Box\varphi$, we conclude that $\Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\pi$, the PDDP.

The $\Rightarrow$-direction of the proof is accounted for by the saturation lemma and the relation between stability and saturation as formulated in proposition 3.4. Then following three claims provide the desired results.

a. φ has the DP $\Rightarrow$ $\{\Box\varphi\} \trianglelefteq \Lambda_1 := \Box C_{\Box\varphi} \cup \overline{\Box\mathcal{L}}$,

b. φ has the PDP $\Rightarrow$ $\{\Box\varphi\} \trianglelefteq \Lambda_2 := \Box(C_{\Box\varphi})_0 \cup \overline{\Box\mathcal{L}_0}$,

c. φ has PDP & PDDP $\Rightarrow$ $\{\Box\varphi\} \trianglelefteq \Lambda_3 := \Box(C_{\Box\varphi})_0 \cup \Diamond R^\Diamond_{\Box\varphi} \cup \overline{(\Box\mathcal{L}_0 \cup \Diamond\mathcal{L}_0)}$.

The following arguments show that these implications are sufficient.

1. φ has DP $\Rightarrow$ (saturation lemma, claim *a* above)
   $\exists\Theta \in \mathcal{SAT} : \{\Box\varphi\} \subseteq \Theta \subseteq \Box C_{\Box\varphi} \cup \overline{\Box\mathcal{L}} \Rightarrow$ ($S = \Box^-\Theta$, proposition 3.4)
   $\exists S \in \mathcal{ST}(\varphi) : S = C_{\Box\varphi} \Rightarrow$ (corollary 3.8)
   φ is naïvely honest.

2. φ has PDP $\Rightarrow$ (saturation lemma, claim *b* above)
   $\exists\Theta \in \mathcal{SAT} : \{\Box\varphi\} \subseteq \Theta \subseteq \Box(C_{\Box\varphi})_0 \cup \overline{\Box\mathcal{L}_0} \Rightarrow$ ($S = \Box^-\Theta$, proposition 3.4)
   $\exists S \in \mathcal{ST}(\varphi) : S_0 = (C_{\Box\varphi})_0 \Rightarrow$ (theorem 3.14)
   φ is weakly honest.

3. φ has PDDP $\Rightarrow$ (saturation lemma, claim *c*)
   $\exists\Theta \in \mathcal{SAT} : \{\Box\varphi\} \subseteq \Theta \subseteq \Box(C_{\Box\varphi})_0 \cup \Diamond R^\Diamond_{\Box\varphi} \cup \overline{\Box\mathcal{L}_0 \cup \Diamond\mathcal{L}_0} \Rightarrow$ ($S = \Box^-\Theta$)
   $\exists S \in \mathcal{ST}(\varphi) : S_0 = (C_{\Box\varphi})_0$ & $\Diamond(\Diamond^- S)_0 = R^\Diamond_{\Box\varphi} \Rightarrow$ (theorem 3.22)
   φ is strongly honest.

What remains to be shown are the three claims a – c above. Recall that this boils down to showing $\Sigma \cap \Lambda_i \neq \emptyset$, for each Σ for which $\Box\varphi \vdash \Sigma$ $i \leq 3$).

a Suppose φ has the DP and $\Box\varphi \vdash \Sigma$.

- If $\Sigma \cap \overline{\Box\mathcal{L}} \neq \emptyset$, we immediately obtain $\Sigma \cap \Lambda_1 \neq \emptyset$.

- If $\Sigma \cap \overline{\Box\mathcal{L}} = \emptyset$ then $\Sigma \subseteq \Box\mathcal{L}$, which means that $\Sigma = \Box\Sigma'$ for certain $\Sigma' \subseteq \mathcal{L}$. DP guarantees the existence of a $\sigma' \in \Sigma'$ such that $\Box\varphi \vdash \sigma'$. Since this means that $\Box\sigma' \in \Box C_{\Box\varphi}$, we may conclude $\Sigma \cap \Box C_{\Box\varphi} \neq \emptyset$, hence $\Sigma \cap \Lambda_1 \neq \emptyset$.

b Suppose φ has the PDP, and $\Box\varphi \vdash \Sigma$.

If $\Sigma \subseteq \Box\mathcal{L}_0$ then, according to PDP, there exists $\Box\sigma \in \Sigma$ such that $\Box\varphi \vdash \sigma$. This means $\Sigma \cap \Box(C_{\Box\varphi})_0 \neq \emptyset$. If $\Sigma \not\subseteq \Box\mathcal{L}_0$, then $\Sigma \cap \overline{\Box\mathcal{L}_0} \neq \emptyset$. Consequently, in all cases $\Sigma \cap \Lambda_2 \neq \emptyset$.

c Suppose φ has the PDDP, and $\Box\varphi \vdash \Sigma$.

- If $\Sigma \cap \overline{(\Box\mathcal{L}_0 \cup \Diamond\mathcal{L}_0)} \neq \emptyset$ then $\Sigma \cap \Lambda_3 \neq \emptyset$.
- Suppose $\Sigma \subseteq \Box\mathcal{L}_0 \cup \Diamond\mathcal{L}_0$.
  - If $\Sigma \cap \Box(C_{\Box\varphi})_0 \neq \emptyset$, then also $\Sigma \cap \Lambda_3 \neq \emptyset$.
  - Take $\Sigma \cap \Box\mathcal{L}_0 \subseteq \Box(\overline{C_{\Box\varphi}})_0$. In this remaining case all formulas of Σ are either of the form $\Diamond\pi$ with $\pi \in \mathcal{L}_0$ or $\Box\varrho$ with $\varrho \notin (C_{\Box\varphi})_0$. Application of the rule MON yields

    $$\Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Sigma \cap \Diamond\mathcal{L}_0.$$

    According to PDDP this means that there exist $\sigma \in \Sigma \cap \Diamond\mathcal{L}_0$ such that $\Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \sigma$ and therefore $\sigma \in \Diamond R^{\Diamond}_{\Box\varphi}$. We conclude, also in this last case, $\Sigma \cap \Lambda_3 \neq \emptyset$. ∎

Since the disjunction properties are purely inferential and strictly related to the possibly honest formula under inspection, and neither involves extension to a stable set that is minimal in some sense, nor minimization in a class of models, they provide a convenient tool for testing honesty. Disjunction properties are particularly useful for proving that some formula is *dishonest*, as we will illustrate by reconsidering three examples.

EXAMPLE 3.37

- Using the PDP it easily follows that $\Box p \vee \Box q$ is not even weakly honest (Cf. example 3.13): $\Box(\Box p \vee \Box q) \vdash \Box p \vee \Box q$, so $\Box(\Box p \vee \Box q) \vdash \Box\{p, q\}$, yet $\Box(\Box p \vee \Box q) \nvdash p$ and $\Box(\Box p \vee \Box q) \nvdash q$ (where non-derivability is shown by providing a counter-model, as usual). That $\Box p \vee \neg p$ is not weakly honest has a similar proof, now by taking $\Pi = \{p, \neg p\}$, thus contradicting the PDP.

- Some of the earlier proofs can also be simplified. For example, example 3.9 now has a very easy proof: $\Box(p \vee q) \vdash \Box p \vee \Box\Diamond q$, yet $\Box(p \vee q) \nvdash p$ and $\Box(p \vee q) \nvdash \Diamond q$, and thus DP shows that $p \vee q$ is not naïvely honest.

- Though less comfortable, PDDP can be used for an alternative proof of example 3.18: $\varphi = \Diamond p \vee \Diamond q$ is not strongly honest since $\Box\varphi \vdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond\{p, q\}$, $\Box\varphi \nvdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond p$, and $\Box\varphi \nvdash \Box(\overline{C_{\Box\varphi}})_0, \Diamond q$. To show the latter, consider the model M from example 3.33. M verifies $\Box\varphi$, but does not verify $\Diamond q$, nor any element of $\Box(\overline{C_{\Box\varphi}})_0$. To make this last point, suppose that $M \models \Box\alpha$ for some $\alpha \in \mathcal{L}_0$. Then in the empty balloon world ϵ: $M_\epsilon \models \alpha$, thus (by propositional persistence) $\models \alpha$, and therefore $\alpha \in (C_{\Box\varphi})_0$.

4. Conclusion

We have described a new epistemic logic with the remarkable feature that on the one hand knowledge implies truth, yet on the other hand truth does not imply epistemic possibility, thus avoiding at least one type of logical omniscience. The logic is shown to be sound and complete for so-called balloon models with partial interpretation.

This logic is then used as a vehicle to study circumscription of knowledge. We have introduced different notions of honesty, each of which can be equivalently described in a number of ways. This results in a hierarchy of honesty, since we can easily prove

φ is naïvely honest $\Rightarrow$ φ is strongly honest $\Rightarrow$ φ is weakly honest.

Definitions of stability, minimality of models with respect to knowledge and disjunction properties which are given the classical study of honesty [HM85] have been reformed in such a way that they can successfully be transferred to partial logic for characterization of the three notions of honesty which evolve from our system **L**. As a summary of these characteristics the following table depicts them once more:

type	Stable sets	Balloon models	Disjunction properties
naive	$\subseteq$-minimality	$\sqsubseteq$-minimality	ordinary DP
weak	$\subseteq_0$-minimality	$\sqsubseteq_\Box$-minimality	PDP
strong	$\subseteq_\Diamond$-min. among $\subseteq_0$-min.	$\subseteq_\Diamond$-min. among $\subseteq_\Box$-min.	PDDP

As we have illustrated on a number of examples, naïve honesty is too strong (i.e. it yields too many dishonest formulas). Weak honesty gives a similar analysis of honesty as Halpern and Moses' analysis. The notion of strong honesty is preferred within the partial approach. By minimizing the 'size' of worlds, epistemic alternatives can be taken as small as possible. Moreover, their number can also be reduced, since not only knowledge, but also possibilities are minimized.

This additional optimization enables us to give an analysis of non-monotonic validity on the basis of minimal knowledge states. The following 'preferential' consequence relation would evolve from our concept of strong minimality.

$$\varphi \mid\!\sim \psi \Leftrightarrow \varphi \text{ is strongly honest and for all strongly minimal } S \in \mathcal{ST}(\varphi) : \psi \in S.$$

This relation intuitively denotes that, if φ is only known, then ψ is also known. Due to our partial background logic, we find no entailment of irrelevant possibilities, e.g.:

$$p \not\mid\!\sim \Diamond q.$$

Notice that though many non-monotonic entailments that were valid for the classical system **S5** do not qualify for our partial system **L**, such entailment still differs from (partial) consequence and derivability: we have, for example

$$(p \vee q) \not\vdash \Diamond p \quad \& \quad (p \vee q) \mid\!\sim \Diamond p.$$

We can extend the latter example to show that '$\mid\!\sim$' is indeed a non-monotonic relation: we have for instance

$$(p \vee q) \wedge \neg p \not\models \Diamond p.$$

This analysis shows the core difference with the classical approach of Halpern and Moses. Their approach also predicts the two last observations, but with them possibility is the only means to capture ignorance, and therefore $\Diamond q$ follows from p in their definition of non-monotonic inference.

References

[Acz68] Aczel, P. 'Saturated intuitionistic theories', in : H. Schmidt, K. Schütte & H. Thielle (eds.) *Contributions to Mathematical Logic*, pp. 1–13, North-Holland, Amsterdam, 1968.

[BP83] Barwise, J. & Perry, J., *Situations and Attitudes*, MIT Press, Cambridge (USA), 1983.

[Ben90] Benthem, J. van, 'Modal logic as a theory of information', in: J. Copeland (ed.) *Proceedings of the Prior Memorial Colloquium, Christchurch 1989*, to appear with Oxford University Press, Oxford UK.

[Bla86] Blamey, S., 'Partial Logic', in: Gabbay & Günthner (eds) *Handbook of Philosophical Logic*, volume 3, Reidel, Dordrecht, 1986.

[HM85] Halpern, J. & Y. Moses, 'Towards a theory of knowledge and ignorance', in Kr. Apt (ed.) *Logics and Models of Concurrent Systems*, Springer–Verlag, Berlin, 1985.

[HC84] Hughes, G. & M. Cresswell - *A Companion to Modal Logic*, Methuen, London, 1984.

[HJT94] Hoek, W. van der, J. Jaspars & E. Thijsse, 'Honesty in Partial Logic', in J. Doyle, E. Sandewall, P. Torasso (eds.) *Proceedings of the fourth international conference on Principles of Knowledge Representation and Reasoning* (KR'94), Morgan Kaufmann, San Francisco, 1994, pp. 583-594.

[Jas91a] Jaspars, J., 'Theoretical circumscription in partial modal logic', in: J. van Eijck (ed.), *Logics in AI*, Proceedings JELIA'90, pp. 301–316, LNCS 478, Springer–Verlag, Berlin, 1991.

[Jas91b] Jaspars, J., 'A generalization of stability and its application to circumscription of positive introspective knowledge', *Proceedings of the Ninth Workshop on Computer Science Logic* (CSL'90), Springer–Verlag, Berlin, 1991.

[Jas93] Jaspars, J., 'Normal forms in partial modal logic', in: C. Rauszer (ed.), *Algebraic Methods in Logic and in Computer Science*, Banach Center Publications, volume 28, Institute of Mathematics, Polish Academy of Sciences, Warsaw, 1993.

[Jas94] Jaspars, J. , *Calculi for Constructive Communication: A Study of the Dynamics of Partial States*, doctoral dissertation, ILLC Dissertation Series 1994–4 & ITK Dissertation Series 1994–1, Amsterdam, 1994.

[JT93] Jaspars, J. & E. Thijsse, 'Fundamentals of Partial Modal Logic', in P. Doherty & D. Driankov (eds.), *Partial Semantics and Non-monotonic Reasoning for Knowledge Representation* (provisional title), based on workshop Linköping (Sweden) 1992, to appear.

[Lan88] Langholm, T., *Partiality, Truth and Persistence*, CSLI Lecture Notes No. 15, Stanford

CA, 1988.

[Moo85] Moore, R., 'Semantical considerations on non-monotonic logic', *Artificial Intelligence* 25, pp. 75–94, 1985.

[ST92] Schwarz, G. & M. Truszczyński, 'Modal logic S4F and the minimal knowledge paradigm', in Y. Moses (ed.), *Proceedings of* TARK *4*, (Monterey CA), Morgan Kaufmann, Palo Alto CA, 1992.

[Sta] Stalnaker, R., *A note on non-monotonic modal logic*, unpublished manuscript, Department of Philosophy, Cornell University.

[Sti87] Stirling, C., 'Modal logics for communication systems', *Theoretical Computer Science* 49, pp. 311–347, 1987.

[Sto77] Stoy, J., *Denotational Semantics: The Scott-Strachey Approach to Programming Language Theory*, The M.I.T. Series in Computer Science, M.I.T. Press, Cambridge MA, 1977.

[Thi90] Thijsse, E. - 'Partial propositional and modal logic: the overall theory', M. Stokhof & L. Torenvliet (eds.) *Proceedings of the* 7^{th} *Amsterdam Colloquium*, volume 2, pp. 555–579, ITLI, Amsterdam, 1990.

[Thi92] Thijsse, E., *Partial logic and knowledge representation*, doctoral dissertation, Eburon Publishers, Delft, 1992.

[Tho68] Thomason, R.H. 'On the strong semantical completeness proof of intuitionistic predicate logic', *Journal of Symbolic Logic* 33:1, pp. 1–7, 1968.