A Modal Contrastive Logic: The Logic of 'But'

J.J.Ch. Meyer and W. van der Hoek

UU-CS-1995-18 (revised version of UU-CS-1994-07)
May 1995

Utrecht University
Department of Computer Science
Padualaan 14, P.O. Box 80.089,
3508 TB Utrecht, The Netherlands,
Tel. : ... + 31 - 30 - 531454

A Modal Contrastive Logic: The Logic of 'But'

J.J.Ch. Meyer and W. van der Hoek

Technical Report UU-CS-1995-18 (revised version of UU-CS-1994-07)
May 1995

Department of Computer Science
Utrecht University
P.O.Box 80.089
3508 TB Utrecht
The Netherlands

ISSN: 0924–3275

A MODAL CONTRASTIVE LOGIC:
The Logic of 'But'[1]

J.-J. Ch. Meyer [2]
W. van der Hoek

Utrecht University
Department of Computer Science
P.O. Box 80.089, 3508 TB Utrecht, The Netherlands

ABSTRACT
In this paper we present a modal approach to "contrastive logic", the logic of contrasts as these appear in natural language conjunctions such as 'but'. We use a simple modal logic, which is an extension of the well-known **S5** logic, and base the contrastive operators proposed by Francez in [F] on the basic modalities that appear in this logic. We thus obtain a logic for contrastive operators that is more in accord with the tradition of intensional logic, and that, moreover—we argue—, has some more natural properties. Particularly, attention is paid to nesting contrastive operators. We show that nestings of 'but' give quite natural results, and indicate how nestings of other contrastive operators can be done adequately. Finally, we discuss the example of the Hangman's Paradox and some similarities (and differences) with default reasoning.

0. INTRODUCTION
As the pre-eminent vehicle people use for communication, natural language plays a crucial role in AI research. Artificially intelligent systems that are supposed to be really useful for non-specialists in AI and computer science, will have to be able to communicate with human beings in their own language, and therefore the construction of these systems will have to involve natural language interfaces. In order to device these, we need to have a perfect understanding of natural language as well as a formal treatment of it. However, natural language is such a rich topic that it is extremely difficult to formalize it to such a degree that it is suited for mechanical processing.

Problems start already with very basic natural language concepts. The meaning of natural language conjunctions such as 'but' and 'already' are traditionally considered to be beyond the scope of formal logic. 'But', for instance, has the flavour of the

---

[1]*But but us no buts,* as they say.
[2]Also partially supported by Nijmegen University, Toernooiveld, 6525 ED Nijmegen, The Netherlands.

conjunction 'and', but it also contains something beyond that, with a touch of contrast (or surprise), which—as some would say—would not be amenable to formal treatment. (Since these conjunctions are ubiquitous in natural language conversations, a formal / logical treatment of these would be the least we should have for a proper understanding and representation of natural language having mechanical processing in mind.)

Nevertheless, in a clear and stimulating paper ([F]), Nissim Francez argues for a logical treatment of natural language conjunctions such as 'but'. Furthermore, he claims that ordinary logic is not sufficient to express the contrastive nature of 'but' rather than the neutral conjunction 'and'. Since 'but' is like 'and', but with an extra flavour of contrast, he proposes the notion of a *bilogic*, which–roughly speaking–consists of two semantic structures: one to deal with the expected world and one to deal with the actual world. In this paper we will refer to these structures as *bi-structures*.

Although we agree entirely to both the need for a logical treatment of these conjunctions in general and the idea of a discrepancy between expected and actual aspects in them in particular, we do not believe that one has to introduce a new kind of logic (viz. bilogic) for their treatment. In fact, we shall show that a particular kind of modal logic will do. Modal logic is well-known in the literature of (philosophical) logic (cf. e.g. [Ch], [HC]), and becomes more popular rapidly among computer scientists and AI researchers for the formal treatment of various manners of knowledge representation (cf. [HM] and [MH6]).

In this paper we propose a simple extension of the well-known logic S5, called $S5AP_{(n)}$, to treat contrastive aspects in natural language conjunctions. Both the expected and actual ingredients can be handled by modal operators in an adequate manner. In this way we can also obtain a more natural treatment of nested contrastive operators than in the bilogic approach of [F]. Moreover, our approach seems fit to express more intricate notions / operators, due to the fact that we have the availability of various modalities in our approach. Although we do certainly not want to go as far as claiming that we thus obtain "*the* logic of 'but'" (as we shall see, there is still room for debate regarding the choices we have made), we *do* believe that our logic is a good first (or rather second) approximation of such a logic.

The logic $S5AP_{(n)}$ was earlier used successfully for the treatment of other notions in the representation of (common-sense) knowledge and reasoning, viz. counterfactuals ([MH3]), as well as nonmonotonic reasoning methods such as default reasoning ([MH1,2,4,5]), including the infamous Yale Shooting Problem as a case study ([Me]). If, moreover, we may view this paper as a very first step towards representing and reasoning about knowledge expressed in natural language, $S5AP_{(n)}$ thus appears to be a potential representation language for quite a substantial part of AI-related knowledge.

This paper is organised as follows: after this introduction we begin in section 1 with the (re)introduction of the logic $S5AP_{(n)}$, giving language, semantics and the logical system proper. In section 2 we shall recall the operators Francez uses in [F] and show how to treat these operators in the logic $S5AP_{(n)}$. We elaborate on some consequences, including some similarities and differences with Francez' approach. In section 3 we insert a pause in our exposition with a discussion of the well-known Hangman's Paradox, as an illustration of the use of our approach as a "logic of surprise". In section 4 we take up the main issue again with a treatment of nested contrastive operators. We show that nestings of 'but' are more natural in our approach, and indicate how the use of an $S5AP_{(n)}$-like modal logic also enables an adequate treatment of nestings of other contrastive operators. In section 5 we indicate some relations with defaults, well-known from the area of nonmonotonic reasoning ([Rei1], [Rei2]). Finally we summarize our results in a conclusion in section 6.

1. THE LOGIC $S5AP_{(n)}$

In [MH1,2,5] we have introduced an epistemic logic ([Hin]) to treat defaults. To do so we took an S5-like modal language extended with special modal operators A denoting *actual truth* and P denoting *preferred* or *practical belief.* (The logic resembles auto-epistemic logic (AEL) of [Mo], but has a much simpler interpretation.) In the present paper we shall use this logic again, but with a modified interpretation: instead of knowledge we now interpret the basic S5-modality as an objective kind of certainty / inescapability and, moreover, we here interpret preferred belief as *expectation*, which will be a crucial ingredient of contrastive logic. We now introduce the logic formally.

Assume a fixed collection **P** of primitive propositions. The language **L** we shall use is the minimal set of formulas closed under the following rules: (i) $\mathbf{L} \supseteq \mathbf{P}$, (ii) if $\varphi, \psi \in \mathbf{L}$, then $\neg\varphi$, $\varphi \wedge \psi$, $\varphi \vee \psi$, $\varphi \supset \psi$, $\varphi \equiv \psi \in \mathbf{L}$, (iii) if $\varphi \in \mathbf{L}$, then $L\varphi$, $M\varphi$, $A\varphi$, $P_i\varphi \in \mathbf{L}$ ($i = 1,...,n$). Informally, $L\varphi$ is read as "φ is necessary / certain", in the sense that φ is inescapable: it holds in all a priori (ontologically) possible (in the sense of imaginable) worlds, $M\varphi$ as "φ is (considered) ontologically possible", $A\varphi$ as "φ is actually true (true in the real world)", and $P_i\varphi$ as "φ is preferred or *expected* (within frame of reference i)". As we shall see below, a frame of reference (or mind) refers to a preferred subset of the set of *a priori* possible worlds that together constitute a body of knowledge or, rather, preferred or expected belief. Note that the formula $A\varphi$ has *no* epistemic reading: it does *not* say that it is *known* that φ holds in the actual world. It merely states that φ holds in the actual or real world. This enables us to speak about truth in the actual world, not depending on the possible world we evaluate the formula. Another way to view this is that the A-operator refers to the actual world as it is known by an omniscient external observer. This omniscient observer can always tell whether some assertion that is expected to be true by some (human or artificial) user of language, is actually true. It is also important to note that the L-operator here denotes an objective kind of certainty rather than a subjective kind of knowledge like in [MH1,2,5]: in the present paper we use L as an ontological necessity operator stating something about truth in all conceivable worlds. We will return to this later, when we discuss the interplay between the L-operator and the actuality operator A.

Formally, expressions of **L** are interpreted by means of Kripke-structures of the form $(\Sigma,\pi,\rho,\Sigma_1,\ldots,\Sigma_n,\wp,\wp_1,\ldots,\wp_n,\mathfrak{R})$, where Σ is a non-empty collection of possible worlds, π: $\mathbf{P} \rightarrow \Sigma \rightarrow \{t,f\}$ is a truth assignment to the primitive propositions per world, $\rho \in \Sigma$ is the real or actual world, $\Sigma_i \subseteq \Sigma$ are sets (clusters) of preferred worlds, $\wp = \Sigma \times \Sigma$ (and $\wp$ is hence reflexive, transitive and euclidean), $\wp_i = \Sigma \times \Sigma_i \subseteq \wp$ (i.e., $\forall s,t : \wp_i(s,t) \Leftrightarrow t \in \Sigma_i$) is transitive and euclidean, and $\mathfrak{R}$ is the constant functional relation $\Sigma \times \{\rho\}$, pointing to the real world ρ from every world in Σ. We let $\mathbf{S5AP_{(n)}}$ denote the collection of Kripke-structures of the above form (the **A** in $\mathbf{S5AP_{(n)}}$ referring to actualities, and the **P** in $\mathbf{S5AP_{(n)}}$ referring to preferences).

The formal interpretation of the language now reads: for $\mathbb{M} = (\Sigma, \pi, \rho, \Sigma_1,\ldots, \Sigma_n, \wp, \wp_1,\ldots, \wp_n, \mathfrak{R})$,

(i) $(\mathbb{M},s) \models p$ iff $\pi(p)(s)=t$
(ii) $(\mathbb{M},s) \models \neg\varphi$ iff not $(\mathbb{M},s) \models \varphi$
(iii) $(\mathbb{M},s) \models \varphi \wedge \psi$ iff $(\mathbb{M},s) \models \varphi$ and $(\mathbb{M},s) \models \psi$
(iv) $(\mathbb{M},s) \models L\varphi$ iff $(\mathbb{M},t) \models \varphi$ for all $t \in \Sigma$ (or, equivalently, for all t such that $\wp(s,t)$)
(v) $(\mathbb{M},s) \models M\varphi$ iff $(\mathbb{M},t) \models \varphi$ for some $t \in \Sigma$ (or, equivalently, for some t with $\wp(s,t)$)
(vi) $(\mathbb{M},s) \models A\varphi$ iff $(\mathbb{M},\rho) \models \varphi$ (or, equivalently, $(\mathbb{M},t) \models \varphi$ for the unique t with $\mathfrak{R}(s,t)$)
(vii) $(\mathbb{M},s) \models P_i\varphi$ iff $(\mathbb{M},t) \models \varphi$ for all $t \in \Sigma_i$ (or, equivalently, for all t with $\wp_i(s,t)$),
and $\mathbb{M} \models \varphi$ iff $(\mathbb{M},s) \models \varphi$ for all $s \in \Sigma$, and $\models \varphi$ iff $\mathbb{M} \models \varphi$ for all $\mathbb{M} \in \mathbf{S5AP_{(n)}}$, as usual.

It is possible to axiomatize (the theory of) $\mathbf{S5AP_{(n)}}$ as follows (cf. [MH2]): take the S5 system for the modality L (and dual $M \equiv \neg L \neg$) and use K45 for the P-modalities, resulting in the main body of the system $S5AP_{(n)}$:

(1) the axioms of Propositional Calculus (PC),
(2) $L(\varphi \supset \psi) \supset (L\varphi \supset L\psi)$,
(3) $L\varphi \supset \varphi$,

(4) $M\varphi \supset LM\varphi$,
(5) $P_i(\varphi \supset \psi) \supset (P_i\varphi \supset P_i\psi)$,
(6) $P_i\varphi \supset P_iP_i\varphi$,
(7) $\neg P_i\varphi \supset P_i\neg P_i\varphi$,
completed by the modus ponens and necessitation (for L) rules.

Moreover, the following relating axioms are added:
(8) $L\varphi \supset P_i\varphi$,
(9) $LP_i\varphi \equiv P_i\varphi$,
(10) $\neg P_i$ **false** $\supset (P_iP_j\varphi \equiv P_j\varphi)$,
(11) $\neg P_i$ **false** $\supset (P_iL\varphi \equiv L\varphi)$.

Furthermore, for the A-modality we employ:
(12) $A(\varphi \wedge \psi) \equiv A\varphi \wedge A\psi$,
(13) $A\neg\varphi \equiv \neg A\varphi$,
(14) $L\varphi \supset A\varphi$,
(15) $A\varphi \supset LA\varphi$,
(16) $AP_i\varphi \equiv P_i\varphi$,
(17) $\neg P_i$ **false** $\supset (P_iA\varphi \equiv A\varphi)$.

We assume that the S5 part of the logic is familiar. A few theorems concerning the new modalities A and P_i are the following:

1.1 PROPOSITION.
(i) $A\varphi \supset M\varphi$
(ii) $AL\varphi \equiv L\varphi$
(iii) $LA\varphi \equiv A\varphi$
(iv) $AA\varphi \equiv A\varphi$
(v) $A\varphi \equiv \neg A\neg\varphi$
(vi) $A(\varphi \vee \psi) \equiv (A\varphi \vee A\psi)$
(vii) $A(\varphi \supset \psi) \equiv (A\varphi \supset A\psi)$
(viii) $\neg A$ **false**
(ix) $MP_i\varphi \equiv P_i\varphi$
(x) $\neg P_i$ **false** $\supset (P_iM\varphi \equiv M\varphi)$

1.2 THEOREM. The system $S5AP_{(n)}$ is sound and complete with respect to $\mathbf{S5AP_{(n)}}$.
PROOF. Soundness is evident. Completeness is proved in the usual way by showing that any (maximally) consistent set of formulas is satisfiable in a canonical modal

structure in $\mathbf{S5AP_{(n)}}$ (cf. e.g. [HC2]) using the usual modal correspondences of S5 for L/M and K45 for P_i, and furthermore the fact that

(8) corresponds to the property $\wp_i \subseteq \wp$;

(9) corresponds to $\forall s,t,u\ \wp(s,t)\ \&\ \wp_i(t,u) \Rightarrow \wp_i(s,u)$;

(10) corresponds to: if $\wp_i$ is "locally serial" at s, i.e. if there exists a t with $\wp_i(s,t)$, then $\forall s,t,u\ \wp_i(s,t)\ \&\ \wp_j(t,u) \Rightarrow \wp_j(s,u)$, and $\forall s,t\ \wp_j(s,u) \Rightarrow \exists u\ (\wp_i(s,u)\ \&\ \wp_j(u,t))$;

(11) corresponds to if $\wp_i$ is "locally serial" at s, i.e. if there exists a t with $\wp_i(s,t)$, then $\forall s,t,u\ \wp_i(s,t)\ \&\ \wp(t,u) \Rightarrow \wp(s,u)$ (The reverse implication contained in (11) does not give additional information as it did in (10) above);

(12) - (14) concerning the A-modality (and the derived theorems) imply that the relation $\mathfrak{R}$ associated with it is an idempotent, functional relation $\subseteq \wp$;

(15) corresponds to $\forall s,t,u\ \wp(s,t)\ \&\mathfrak{R}(t,u) \Rightarrow \mathfrak{R}(s,u)$;

(16) to a property like (9) except with $\wp$ replaced by $\mathfrak{R}$; and

(17) to a property like (11) except with $\wp$ replaced by $\mathfrak{R}$.

(These correspondences have been studied systematically in a general modal setting in [H].)

Together these properties make a (maximally) consistent set satisfiable in a (canonical) model in $\mathbf{S5AP_{(n)}}$. By contraposition this implies that every valid assertion is derivable, as usual, proving completeness. □

Note that the introduction of the A-modality enables us to speak about valid assertions concerning the actual world. If we should use just the assertions φ instead of Aφ—as is done usually in modal logic—this would not be possible, since the validity of φ in an **S5**-modal setting is equivalent with the validity of Lφ. So in this setting we can introduce assertions about the actual world as axioms (and apply necessitation on them) without losing the actual character of these assertions. One should, however, note the intended interpretation of $LA\varphi \equiv A\varphi$, (Proposition 1.2 (iii)). It is especially here that it makes a huge difference interpreting the L-operator as an objective kind of certainty (or necessity) or that one interprets it as knowledge! In the latter interpretation $LA\varphi \equiv A\varphi$ would be absurd: it would state that something is known to be actually true exactly when it is actually true. This is not the intended meaning here, however. In the intended former reading of L as necessity the validity $LA\varphi \equiv A\varphi$ just states that what is actually true, is necessarily actually true, which means that the actual world is fixed as one particular world from the set of a priori possible worlds. So, in summary, L refers to the whole set Σ of a priori (ontologically) possible worlds, A to the (fixed) actual world in Σ, and P to the (set of) expected world(s) in Σ. This formal apparatus enables us to speak about truth in the actual world as well as truth in all a priori possible worlds, including all possible alternatives for the actual (or real) world. Using it we will be able to give meaning to Francez' contrastive operators in the next section.

2. CONTRASTIVE OPERATORS

In [F] the following operators are introduced: $\lhd\varphi$ (contrast), $\varphi\surd$ (conformity), $!\varphi$ (expectation), $\varphi!$ (actuality), φ **but** ψ (the formal pendant of the natural language conjunction 'but'). (*Actually*, Francez uses a slightly different set of symbols for these operators, *but* these were not available on our Mac™.) Informally, these operators have the following meaning:[3]

(1) The contrast operator, in $\lhd\varphi$, read as "surprisingly φ", asserts that φ holds in the actual word, but not in the expected world(s) (or 'standard' world, as Francez calls it),
(2) the conformity operator, in $\varphi\surd$, read as "truly (or indeed) φ", asserts that φ holds in the actual world, conform what was expected,
(3) the expectation operator, in $!\varphi$, read as "expectedly (hopefully / it is feared that) φ", asserts that φ is expected, i.e., holds in the expected world, which may be further interpreted as either wished or feared.
(4) the actuality operator, in $\varphi!$, read as "actually φ", asserts that φ holds in the actual world, and finally
(5) the but operator, in φ **but** ψ, asserts that φ holds in the actual world, that therefore $\neg\psi$ is expected (holds in the expected world), but that actually it holds that ψ (together with φ).

These operators can be given an adequate formal semantics within $\mathbf{S5AP_{(n)}}$, by considering them as abbreviations of $\mathbf{S5AP_{(n)}}$-modalities:

(1) $\lhd\varphi =_{def} P\neg\varphi \wedge A\varphi$;
(2) $\varphi\surd =_{def} P\varphi \wedge A\varphi$;
(3) $!\varphi =_{def} P\varphi$;
(4) $\varphi! =_{def} A\varphi$;
(5) φ **but** $\psi =_{def} L(\varphi \supset P\neg\psi) \wedge A(\varphi \wedge \psi)$.[4]

---

[3]Francez makes this formal by using his bilogic, where he distinguishes between an expected component and an actual one. Here we use our own terminology, and a semantics based on $\mathbf{S5AP_{(n)}}$.
[4]Actually, a more direct translation of Francez' semantics of '**but**' would read as $P(\varphi\supset\neg\psi)\wedge A(\varphi\wedge\psi)$. We have not chosen for this translation, since it cannot be used to derive results such as the (desirable) assertion that $\neg\psi$ is expected, and, moreover, it suffers from a number of the same problems as in [F], as will be discussed in the sequel. In fact, the modal framework of $\mathbf{S5AP_{(n)}}$ enables us to deviate from Francez' interpretation in the manner displayed in our definition in order to overcome the problems with this interpretation.

These translations are fairly obvious, when considering the semantics for the L, A and P-modalities. We note that the P-modality may stand for any of the P_i-modalities, associated with the frames Σ_i. This gives us the opportunity to vary these in more complex cases of knowledge representation. This flexibility may be very useful, since it is not obvious at all that there should only be *one* expected frame, as is the case in Francez' approach. In fact, we shall see presently, when we try to nest contrastive operators like '**but**', that it appears to be necessary to have several of these frames around!

We remark that, although (1) clearly is the intended translation of Francez' contrast operator $\lhd$, it is possible in our modal framework to define a weaker form of contrast by $\nabla\varphi =_{def} \neg P\varphi \wedge A\varphi$ ($\equiv \neg!\varphi \wedge \varphi!$), which says essentially that φ holds actually, although it was not expected. In our approach the operator P refers to a *set* of worlds, so that in general $\neg P\varphi$ is not equivalent with $P\neg\varphi$. So $\neg!\varphi$ and $!\neg\varphi$, and hence $\nabla\varphi$ and $\lhd\varphi$, are not equivalent either! We believe that this is intuitively right, since there appears to be a difference between not expecting something and expecting that something is not the case. In [F] this distinction cannot be made, since in his semantics $!\neg\varphi \equiv \neg!\varphi$.

Note also the appearance of the L-operator in clause (5) for '**but**': it appears in the first conjunct, which says that necessarily (L) φ raises the expectation that not ψ. This is intuitively sound—we do indeed want this expectation raised in all possible situations (=worlds). (In fact, $L(\varphi \supset \psi)$ is the natural implication to be used in modal contexts, the so-called *strict* implication.) However, as we shall see in a moment, the definition (5) can be simplified in the context of the logic **S5AP**. It serves as providing an intuition that is rather general and which is as little as dependent on the underlying logic S5AP as possible, so that in case one would like to vary this logic, one can still work with this general definition rather than its more simple S5AP-equivalent.

2.1 PROPOSITION ([S2]).

(i) $\models \varphi$ **but** $\psi \equiv A(\varphi \wedge \psi) \wedge P\neg\psi$

(ii) $\models \varphi$ **but** $\psi \equiv A\varphi \wedge \lhd\psi$

PROOF.

(i) First we note that: φ **but** $\psi \Leftrightarrow L(\varphi \supset P\neg\psi) \wedge A(\varphi \wedge \psi) \Rightarrow A(\varphi \supset P\neg\psi) \wedge A(\varphi \wedge \psi) \Rightarrow A(\varphi \supset P\neg\psi) \wedge A\varphi \wedge A\psi \wedge A(\varphi \wedge \psi) \Rightarrow (A\varphi \supset AP\neg\psi) \wedge A\varphi \wedge A(\varphi \wedge \psi) \Rightarrow AP\neg\psi \wedge A(\varphi \wedge \psi) \Rightarrow P\neg\psi \wedge A(\varphi \wedge \psi)$. For the converse: $P\neg\psi \wedge A(\varphi \wedge \psi) \Rightarrow LP\neg\psi \wedge A(\varphi \wedge \psi) \Rightarrow L(\varphi \supset P\neg\psi) \wedge A(\varphi \wedge \psi) \Leftrightarrow \varphi$ **but** ψ.

(ii) φ **but** $\psi \Leftrightarrow_{(i)} A(\varphi \wedge \psi) \wedge P\neg\psi \Leftrightarrow A\varphi \wedge A\psi \wedge P\neg\psi \Leftrightarrow A\varphi \wedge \lhd\psi$.
▯

This gives us a surprisingly simple characterization of the 'but' operator: Prop.2.1(i) expresses that "φ **but** ψ" is true iff both φ and ψ are actually true although it was expected that ψ does not hold, while Prop. 2.1(ii) states that "φ **but** ψ" is true iff φ is actually true and (we are tempted to say "but", but this would more or less be tautological) ψ is a surprise. This is very intuitive indeed. In fact, of course, we could have chosen this to be the definition of this operator. However, we have chosen not to do this for the reason that we like to see the above characterization of 'but' as a(n S5AP-) consequence of clause (5) of the definition given earlier. We think this definition has a sound semantical justification, which is somewhat less "contingent" in our opinion (and does not hinge on the peculiarities of our underlying logic), and, moreover, keeps a closer relation with Francez' approach.[5]

We discuss some similarities and differences between our approach and the one by Francez. As in [F] we have that

2.2 PROPOSITION.
(i) $\models \lhd\varphi \equiv$ **true but** φ
(ii) $\models \varphi$ **but** $\varphi \equiv \lhd\varphi$
(iii) $\models \neg($**false but** $\varphi)$

PROOF.
(i) $\lhd\varphi \Leftrightarrow P\neg\varphi \wedge A\varphi \Leftrightarrow P\neg\varphi \wedge A($**true** $\wedge\, \varphi) \Leftrightarrow$ **true but** φ.
(ii) φ **but** $\varphi \Leftrightarrow$ (Prop.2.1(i)) $A(\varphi \wedge \varphi) \wedge P\neg\varphi \Leftrightarrow P\neg\varphi \wedge A\varphi \Leftrightarrow \lhd\varphi$.
(iii) **false but** $\varphi \Rightarrow P\neg\varphi \wedge A$ **false** $\Rightarrow$ **false**. ▯

Furthermore, unlike Francez' operator for 'but', ours is *not* symmetric, since $L(\varphi \supset P\neg\psi)$ is obviously not equivalent with $L(\psi \supset P\neg\varphi)$. We see this as an advantage of our approach, witness the observation that in natural language 'but' does not seem to be symmetric: "The conference will be held in Holland, but not in Amsterdam" does not sound equivalent with "The conference will not be held in Amsterdam, but it will be held in Holland". The former statement seems to emphasize that it is not in Amsterdam, the latter that it is in Holland, in the sense that the first sentence conveys that it is expected (albeit refuted) that the conference will be held in Amsterdam, whereas in the

[5]It also provides a good link with the presupposition approach as followed in [WR], see the Postscriptum.

second sentence the attention is focused to the point that there is a (refuted) expectation that the conference will not be in Holland. In our setting, letting 'a' stand for "in Amsterdam" and 'h' for "in Holland", the first sentence is represented by $Pa \wedge A(h \wedge \neg a)$, and the second by $P\neg h \wedge A(\neg a \wedge h)$. The first entails the contrast $Pa \wedge A\neg a$, concentrated around "(not) being in Amsterdam", the second entails the contrast $P\neg h \wedge Ah$, concentrated around "(not) being in Holland", which is in accord with our intuition expressed above. We can say more about this, when we view the relation with defaults in section 5.

Also in our case we do not have that the binary contrast '**but**' is expressible in terms of only the unary contrast '$\lhd$'. ([F] has the equivalence φ **but** $\psi \equiv \lhd(\varphi \wedge \psi)$, which we do *not* have; we only have the left-to-right implication, see Prop. 2.4 below.) Another difference is that in our approach the following proposition holds, which states that φ **but** ψ implies that ψ comes as a surprise.

2.3 PROPOSITION.

$$\models \varphi \textbf{ but } \psi \supset \lhd\psi$$

PROOF. Directly from Prop. 2.1 (ii). □

Proposition 2.3 does not hold in the approach by Francez, and, indeed, one may question the desirability of the validity of φ **but** $\psi \supset \lhd\psi$, attributing absolute surprise to ψ if contrast holds. Our reply to this is that this merely says that as a consequent of φ **but** ψ, ψ is surprising. $\lhd\psi$ does not state that ψ is surprising in an absolute sense. There is a reason for this, viz. the antecedent φ **but** ψ. This antecedent entails an element of surprise w.r.t. ψ. This is similar to the validity of the implication $p \wedge q \supset p$: the consequent p does not hold absolutely, but given $p \wedge q$. In the case of φ **but** $\psi \supset \lhd\psi$: the consequent $\lhd\psi$ does not hold absolutely, but given φ **but** ψ. Moreover, in section 4, we will discuss an extension of the language, in which we can express where surprises come from, so that we can distinguish kinds of surprises. We then return to this issue.

Finally in this section we return to our remark above that although we do not have the equivalence φ **but** $\psi \equiv \lhd(\varphi \wedge \psi)$, but we *do* have the left-to-right implication. This is as it should be intuitively: we have that φ **but** ψ implies that it comes as a surprise that φ and ψ hold together. Moreover, we also *do* have equivalence of φ **but** ψ and $\lhd(\varphi \wedge \psi)$ under the additional assumption that φ is expected ([S2]).

2.4 PROPOSITION.

(i) $\models$ φ **but** $\psi \supset \varphi$ **but** $(\varphi \wedge \psi)$

(ii) $\models$ φ **but** $(\varphi \wedge \psi) \equiv \lhd(\varphi \wedge \psi)$

(iii) $\models$ φ **but** $\psi \supset \lhd(\varphi \wedge \psi)$

(iv) $\models$ $P\varphi \supset (\varphi$ **but** $\psi \equiv \lhd(\varphi \wedge \psi))$

PROOF.

(i) φ **but** $\psi \Rightarrow A(\varphi \wedge \psi) \wedge P\neg\psi \Rightarrow A(\varphi \wedge \psi) \wedge P(\neg\varphi \vee \neg\psi) \Rightarrow A(\varphi \wedge \varphi \wedge \psi) \wedge P(\neg(\varphi \wedge \psi)) \Rightarrow \varphi$ **but** $(\varphi \wedge \psi)$.

(ii) φ **but** $(\varphi \wedge \psi) \Leftrightarrow A(\varphi \wedge \varphi \wedge \psi) \wedge P(\neg(\varphi \wedge \psi)) \Leftrightarrow A(\varphi \wedge \psi) \wedge P(\neg(\varphi \wedge \psi)) \Leftrightarrow \lhd(\varphi \wedge \psi)$.

(iii) directly from (i) and (ii).

(iv) suppose $P\varphi$ is true. Then: φ **but** $\psi \Leftrightarrow A(\varphi \wedge \psi) \wedge P\neg\psi \Leftrightarrow$ (because of $P\varphi$) $A(\varphi \wedge \psi) \wedge P(\varphi \supset \neg\psi) \Leftrightarrow A(\varphi \wedge \psi) \wedge P\neg(\varphi \wedge \psi) \Leftrightarrow \lhd(\varphi \wedge \psi)$. □

3. INTERLUDE: THE HANGMAN'S PARADOX

The language of contrasts and surprises is not only suited to give a treatment of natural language (contrastive) conjunctions. By way of a diversion we show in this section how to represent the well-known Hangman's Paradox in our language. Since this paradox has certain elements of surprise and unexpectedness it is thus not surprising that we may use our language for this.

The version of the paradox that we will treat reads as follows. A prisoner is sentenced to be hanged on some unexpected (for him) day next week. The prisoner reasons now as follows: I cannot be hanged on the last day of the week, since then I would expect this when I wake up that day. But this implies that I cannot be hanged on the one to last day but one, since then I would expect this to happen. Proceeding in this way the prisoner concludes that he will not be hanged next week. But then, quite surprisingly, he is hanged on the third day of the coming week.

In $S5AP_{(1)}$ we can mimic this line of reasoning in the following manner. To start with, we express the statements that are given after the sentence has been passed. Let p_i stand for the proposition that the prisoner is hanged on the i-th day of the week (i = 1, ..., 7; we assume a 7-day hanging week, but with respect to short time working for hangmen also a 5-day hanging week or any other choice will do).

(1) $L(p_1 \vee ... \vee p_7)$: it is certain that the prisoner is hanged on some day in the week.

(2) $L(p_i \supset (\neg p_1 \wedge ... \wedge \neg p_{i-1}))$: certainly, if the prisoner is hanged on the i-th day, he has not been hanged on the days before.

(3) $L(p_i \supset \neg Pp_i)$ for all $i \in \{1, ..., 7\}$: the prisoner is ensured that if the hanging takes place on the i-th day, he will not expect it.

We first introduce moments of time. Let t_i stand for (the moment the prisoner wakes up on) the i-th day. Now we reason as follows. (Actually we should incorporate temporal logic as well, but we avoid this here by using 7 distinct moments of time in our model. At different moments there may be different possibilities (possible worlds) resulting in different certainties and expectations, and as it will turn out below, they indeed do differ from time to time. Technically, this means that we consider S5AP-structures at each of the 7 distinct moments of time: $\mathbb{M}_1, ..., \mathbb{M}_7$. Then by the common sense expression "at time t_i it holds that φ" we just mean formally that $\mathbb{M}_i \models \varphi$. So at each moment of time (viz. day) we consider a different S5AP-model. Note that this means that the modal operators, viz. the certainty operator L, its dual M, the actuality operator A and the expectation operator P, only refer to the possible worlds *at the moment of evaluation*, and *not* to all worlds at *all* times!)

At t_i it holds that $L(p_i \supset L(\neg p_1 \wedge ... \wedge \neg p_{i-1}))$: it is certain that on the morning of the i-th day of the week, if this day is *the* day of hanging, *it is certain* that it was not on any previous day. (In this case, he is still alive and able to reason.) Note that this is stronger than (2) above, since now (in case p_i is true) *at time point t_i* (by the passage of time) it has become *necessary (inevitable)* that the *a priori* possibilities $p_1, ..., p_{i-1}$ are ruled out. In particular, at t_7 it holds that: $L(p_7 \supset L(\neg p_1 \wedge ... \wedge \neg p_6))$, and hence $P(p_7 \supset P(\neg p_1 \wedge ... \wedge \neg p_6))$. Since from (1) it follows that $P(p_1 \vee ... \vee p_7)$, at t_7 it holds that $P(p_7 \supset Pp_7)$[6], i.e., the prisoner expects that if it will happen on the 7th day, he will expect it. Together with (3), this yields $P(p_7 \supset (Pp_7 \wedge \neg Pp_7))$, i.e., $P(p_7 \supset$ **false**), i.e., $P\neg p_7$: the prisoner expects at t_7 that he is not hanged on the 7th day. Although this is true at t_7, the morning of the 7th day, we may now also assume that $P\neg p_7$ holds at earlier moments, since the prisoner can do hypothetical reasoning about the 7th day, and reach this conclusion at any time earlier. (We admit that this is an extra assumption, but we believe it to be in line with the natural language version of the paradox we gave above.) Likewise, using the fact that $P\neg p_7$ at t_6 (by an analogous assumption on the hypothetical reasoning capabilities of the prisoner as before), one

[6]From $P(p_1 \vee ... \vee p_7)$ we obtain $PP(p_1 \vee ... \vee p_7)$, which together with $P(p_7 \supset P(\neg p_1 \wedge ... \wedge \neg p_6))$ yields $P(p_7 \supset (P(\neg p_1 \wedge ... \wedge \neg p_6) \wedge P(p_1 \vee ... \vee p_7))$, i.e., $P(p_7 \supset (P((\neg p_1 \wedge ... \wedge \neg p_6) \wedge (p_1 \vee ... \vee p_7)))$, implying $P(p_7 \supset Pp_7)$.

can reason that at t_6 it holds that $P((p_6 \supset Pp_6)$[7], and consequently $P\neg p_6$. Going on in this way, we obtain that, using $P\neg p_{i+1} \wedge ... \wedge P\neg p_7$ (which we inherit again by analogous reasons as above), at t_i it holds that $P((p_i \supset Pp_i)$ and consequently $P\neg p_i$ ($i = 1, ..., 7$). Now the fact that (at t_3) actually p_3, i.e. Ap_3, comes as a surprise: at t_3 we have $P\neg p_3 \wedge Ap_3$, i.e., $\lhd p_3$.

Finally, note the in the case of a 1-day hanging week, the paradox really becomes inconsistent, since in this case one may show directly from (1) that, at t_1, Lp_1 holds, and thus both Pp_1 and (by 3) $\neg Pp_1$ hold. Which was to be expected in this degenerated case.

4. NESTED CONTRASTIVE OPERATORS: DISCUSSION

In this section we discuss the issue of nesting contrastive operators such as '**but**'. In Francez' approach this presents some problems, since it must be decided which part of the bi-structure is responsible for which part of a nested expression, and Francez uses an operator to isolate the assertional part of an expression in a syntactical way, which is then used in the semantics. Apart from thus complicating the semantics, the operator flattens nestings of contrasts in a rather crude way. A nested expression such as (φ **but** ψ) **but** χ becomes equivalent to what in our language should read as $L((\varphi \wedge \psi) \supset P\neg\chi) \wedge A(\varphi \wedge \psi \wedge \chi)$, which is a rather "flat" expression in the sense that the contrast in it is not nested, which was to be expected. In fact, in Francez' approach the formula (p **but** q) **but** r, obviously containing a nested contrast, is equivalent with the formula ($p \wedge q$) **but** r, which has no nested contrast at all (where p, q and r are purely propositional). Moreover, in Francez' approach it also holds that (p **but** q) **but** r is equivalent with p **but** (q **but** r), which Francez claims to be an intuitive property.

In our opinion, this is too simplistic an approach. For instance, let us consider the sentence

(@) "Tweety is a bird, but Tweety cannot fly, but (still) Tweety is high up in the air (because it is in an airplane, for example)".[8]

This sentence has indeed an intuitive nesting of contrast. First, because Tweety is a bird we expect it to fly; its failure to do so comes as a surprise to us; however, the fact

---

[7]From $L(p_6 \supset L(\neg p_1 \wedge ... \wedge \neg p_5))$ we obtain $P(p_6 \supset P(\neg p_1 \wedge ... \wedge \neg p_5))$, which together with $PP(p_1 \vee ... \vee p_7)$ yields $P(p_6 \supset P(p_6 \vee p_7))$. Hence, since we also have $P\neg p_7$, and so $PP\neg p_7$, we get $P(p_6 \supset (P(p_6 \vee p_7) \wedge P(\neg p_7)))$, i.e., $P(p_6 \supset (P((p_6 \vee p_7) \wedge \neg p_7)))$, so $P((p_6 \supset Pp_6)$.
[8]This example is due to Roland Bol. Note, by the way, the similarity of this example with examples in the literature of nonmonotonicity, where nesting is also an issue of concern.

that we are told that nevertheless Tweety is high up in the air is another surprise, since objects that cannot fly, usually are not high up in the air. So, given the statement that "Tweety is a bird, but Tweety cannot fly", after having assimilated the initial surprise that Tweety cannot fly, although it is a bird, we resign ourselves in this fact, and are then surprised again to hear that Tweety is high up in the air. Clearly, the sentence "Tweety is a bird and Tweety cannot fly, but Tweety is high up in the air" is not intuitively equivalent with the sentence (@), since here we experience only one surprise instead of two in a row.

Nesting of contrasts can be treated adequately in our set-up, as follows. Consider the example above again. Let 'b' stand for "Tweety is a bird", 'f' for "Tweety can fly", and 'h' for "Tweety is high up in the air". Then the sentence (@) can be represented in our setting as (b **but** ¬f) **but** h. When we use the definition we see that this is equivalent with

$L((b\ \mathbf{but}\ \neg f) \supset P\neg h) \wedge A((b\ \mathbf{but}\ \neg f) \wedge h)$,

which is in its turn equivalent with

$L((L(b \supset Pf) \wedge A(b \wedge \neg f)) \supset P\neg h) \wedge A(L(b \supset Pf) \wedge A(b \wedge \neg f) \wedge h)$.

This implies in $S5AP_{(n)}$:

$((L(b \supset Pf) \wedge A(b \wedge \neg f)) \supset P\neg h) \wedge L(b \supset Pf) \wedge A(b \wedge \neg f) \wedge Ah$,

which in its turn implies

$P\neg h \wedge Ah \wedge Ab \wedge A\neg f \wedge A(b \supset Pf)$,

yielding, since $A(b \supset Pf)$ implies $Ab \supset Pf$,

$A(b \wedge \neg f \wedge h) \wedge P(f \wedge \neg h)$,

capturing the intuition of a subsequent, double surprise, viz. regarding "non-flying" and "being high up in the air". This becomes even more apparent, if we use different P-modalities in nested expressions. In fact, the expression we obtained above is perhaps still not entirely adequate, since in it the surprises are put on the same level, so that we cannot distinguish any more which one is the last and most dominant view. The need to distinguish is even greater in situations where the surprises conflict each other. This is also the case in the above example, when we impose additionally that "flying" necessarily implies "being high up in the air", viz. $L(f \supset h)$. This then yields $P(h \wedge \neg h)$, so both h and ¬h are expected. This calls for an ordering in expectation.

Before we discuss a possible way to do this, we pause at the allegedly desirable equivalence of (p **but** q) **but** r and p **but** (q **but** r). It is easily checked that this equivalence does not hold in our setting, and we believe that this is as it should be, witness the following example:

(1) Tweety is not a bird, but (it is the case that): [Tweety is not an insect either, but it can fly].

Although a bit awkward, one can imagine that this sentence makes sense in a situation where one mainly considers birds and insects as flying creatures. Now note the difference with

(2) [Tweety is not a bird, but Tweety is not an insect either], but it can fly.

Sentence (2) is rather strange, since the first "but" does not seem to be really contrastive: even in a situation where the only flying creatures are birds and insects, one may consider many alternatives when one does not limit oneself to flying creatures, and the first "but"-sentence does not say anything about the ability to fly. In any case the meaning of (2) seems to be rather different from that of (1). At least, this example shows that it is not evident that (p **but** q) **but** r and p **but** (q **but** r) should have the same meaning. In fact, the meaning of sentences of the form p **but** (q **but** r) generally do not seem to be very clear[9], so that we even tend to exclude sentences of this form as representations of natural language sentences. It seems to us, that nested "**buts**" should generally be read as sentences of the form ((...((p **but** q) **but** r) **but** ...) **but** s). In the next few paragraphs, where we return to the issue of an ordering in expectation, we assume this to be the case.

The use of multiple P-modalities is supported by the logic and may come in useful when we want to distinguish between contrasts of expectation on the various levels. In the case of (@) one obtains

$$L((L(b \supset P_1 f) \wedge A(b \wedge \neg f)) \supset P_2 \neg h) \wedge A(L(b \supset P_1 f) \wedge A(b \wedge \neg f) \wedge h),$$

which yields

$$A(b \wedge \neg f \wedge h) \wedge P_1 f \wedge P_2 \neg h,$$

indicating that the contrast regarding "non-flying" takes place at another level as that regarding "being high up in the air". If one uses a systematic way to associate indexes with the nestings, it becomes possible to establish the precise level of nesting of a contrast. We may then order the modalities, such that those corresponding to connectives on the highest syntactic level have the greatest semantical impact. So, *in*

[9]This becomes even more pregnant due to an observation by Pierre-Yves Schobbens ([S2]), that one can show in our logic that p **but** (q **but** r) $\equiv$ Ap $\wedge$ Aq $\wedge$ Ar $\wedge$ P **false**. This expresses that something funny is going on with respect to the expectations in this sentence: they are inconsistent. Of course, this can either mean that these sentences are truly strange in a linguistic sense, or that our logic is not capable with these sentences adequately. As yet we do not know this, see also the remarks in the Postscriptum in connection with [WR].

concreto, one associates with an occurrence of '**but**' nested at a level i the index N+1−i, where N stands for the deepest nesting of '**but**' occurring in the formula concerned (where a non-nested occurrence has nesting level 0). In the example above this algorithm has been applied. The resulting $P_1 f \wedge P_2 \neg h$ now expresses that the fact that Tweety is not high up in the air ($\neg h$) is a higher level expectation than that Tweety can fly (f). Note that this only seems to work correctly under the assumption that we only consider "**but**"-nestings of the form ((...((p **but** q) **but** r) **but** ...) **but** s). For nestings of the form p **but** (q **but** (r **but** (...))) this way of assigning expectation levels seems wrong, but on the other hand, this appears to be the precise reason why these sentences are difficult to assign a meaning to, as discussed above.

This feature of ordered expectations is reminding of a theory of ordered preferences that we have developed in the context of defaults in [MH2,4], but the difference is that in the latter situation the outcomes of less preferred default beliefs are overridden by conflicting more preferred ones, whereas in the present setting of contrasts the results of 'lower' expectations still retain their informative value and are (and should) not be discarded. We return to the relationship with defaults in some more detail in Section 5.

Once we have use multiple P_i-operators, we may also offer a better explanation of Proposition 2.3 in section 2. There we saw that φ **but** $\psi \supset \triangleleft\psi$ is a validity in our setting, which may viewed as odd if $\triangleleft\psi$ is interpreted as an absolute surprise (although we already gave some arguments in its defense). Now we have the ability to mark surprises, we may even define indexed operators:

$\triangleleft_i\varphi =_{def} P_i\neg\varphi \wedge A\varphi$;
$\varphi\surd_i =_{def} P_i\varphi \wedge A\varphi$;
$!_i\varphi =_{def} P_i\varphi$;
φ **but**$_i$ $\psi =_{def} L(\varphi \supset P_i\neg\psi) \wedge A(\varphi \wedge \psi)$,

so that we may be more precise about the origin of the expectations and the element of surprise associated with it. So now we have an indexed version of Proposition 2.3:

(*) $\models$ φ **but**$_i$ $\psi \supset \triangleleft_i\psi$,

which explains better where the surprise $\triangleleft_i\psi$ stems from, since now it is directly related to the contrastive sentence φ **but**$_i$ ψ involving an (i-)indexed expectation $P_i\neg\psi$. In any case, reading $\triangleleft_i\psi$ as an absolute form of surprise is far less obvious now, so that the validity of (*) is perhaps less questionable than its unindexed form, Proposition 2.3.

We now continue our exposition, using unindexed contrastive operators. Other results concerning nested uses of our operators include the following:

4.1 PROPOSITION.

(1) $\vdash\ !!\varphi \equiv\ !\varphi$

(2) $\vdash (\varphi\surd)\surd \equiv \varphi\surd$

(3) $\vdash \neg\mathrm{P}\ \mathbf{false} \supset (\lhd\varphi \equiv\ !\lhd\varphi\ (\equiv \mathrm{P}\lhd\varphi))$

(4) $\vdash \neg\mathrm{P}\ \mathbf{false} \supset \neg\lhd\lhd\varphi$

(5) $\vdash \mathrm{P}\ \mathbf{false} \supset (\lhd\lhd\varphi \equiv \lhd\varphi)$

PROOF

(1) $!!\varphi \equiv \mathrm{P}!\varphi \equiv \mathrm{PP}\varphi \equiv \mathrm{P}\varphi \equiv\ !\varphi$.

(2) we first prove the assertion under the assumption that $\neg\mathrm{P}$ **false**:
$(\varphi\surd)\surd \equiv \mathrm{P}(\varphi\surd) \wedge \mathrm{A}(\varphi\surd) \equiv \mathrm{P}(\mathrm{P}\varphi \wedge \mathrm{A}\varphi) \wedge \mathrm{A}(\mathrm{P}\varphi \wedge \mathrm{A}\varphi) \equiv$
$\mathrm{PP}\varphi \wedge \mathrm{PA}\varphi \wedge \mathrm{AP}\varphi \wedge \mathrm{AA}\varphi \equiv \mathrm{P}\varphi \wedge \mathrm{A}\varphi \wedge \mathrm{P}\varphi \wedge \mathrm{A}\varphi \equiv \mathrm{P}\varphi \wedge \mathrm{A}\varphi \equiv \varphi\surd$.
If P **false** holds: $(\varphi\surd)\surd \equiv \mathrm{PP}\varphi \wedge \mathrm{PA}\varphi \wedge \mathrm{AP}\varphi \wedge \mathrm{AA}\varphi \equiv \mathrm{A}\varphi \equiv \mathrm{P}\varphi \wedge \mathrm{A}\varphi \equiv \varphi\surd$.

(3) under the given assumption that $\neg\mathrm{P}$ **false** we obtain:
$\lhd\varphi \equiv \mathrm{P}\neg\varphi \wedge \mathrm{A}\varphi \equiv \mathrm{PP}\neg\varphi \wedge \mathrm{PA}\varphi \equiv \mathrm{P}(\mathrm{P}\neg\varphi \wedge \mathrm{A}\varphi) \equiv \mathrm{P}\lhd\varphi \equiv\ !\lhd\varphi$.

(4) assume that $\neg\mathrm{P}$ **false**:
since $\mathrm{A}\lhd\varphi \equiv \mathrm{A}(\mathrm{P}\neg\varphi \wedge \mathrm{A}\varphi) \equiv \mathrm{AP}\neg\varphi \wedge \mathrm{AA}\varphi \equiv \mathrm{P}\neg\varphi \wedge \mathrm{A}\varphi \equiv \lhd\varphi$, we get:
$\lhd\lhd\varphi \equiv \mathrm{P}\neg\lhd\varphi \wedge \mathrm{A}\lhd\varphi \equiv \mathrm{P}\neg\lhd\varphi \wedge \lhd\varphi \equiv$ (by (3)) $\mathrm{P}\neg\lhd\varphi \wedge \mathrm{P}\lhd\varphi \equiv \mathrm{P}(\neg\lhd\varphi \wedge \lhd\varphi)$
$\equiv$ P **false** $\equiv$ **false**. Hence we have that $\neg\lhd\lhd\varphi$.

(5) assume P **false**: now we have that $\lhd\lhd\varphi \equiv \mathrm{P}\neg\lhd\varphi \wedge \mathrm{A}\lhd\varphi \equiv \mathrm{A}\lhd\varphi \equiv \lhd\varphi$. □

The first two of these are as in [F]; (3) deviates from [F], where it holds that $!\varphi \equiv\ !\lhd\varphi$, but is not very much better: '"expectedly surprisingly" is equivalent with "surprisingly"' is perhaps somewhat more natural than '"expectedly surprisingly" is equivalent with "expectedly"', but it still sounds a bit odd. We shall return to this below; we now continue with a discussion of (4) and (5). Instead of these, Francez has (always, i.e., without any condition) $\lhd\lhd\varphi \equiv \lhd\varphi$, which is rather strange: "surprisingly, surprisingly φ" contains something like a double surprise, a kind of meta-surprise (a surprise concerning a surprise), which is not reflected in the equivalence with a single surprise. Instead, in our approach this double surprise gives rise to inconsistent expectations (4, using contraposition). As can be read from the proof of (4), this has to do with the fact that "surprisingly, surprisingly φ" implies two inconsistent expectations, viz. both the expectations that φ is surprising and that it is not. Moreover, we have that the double surprise is only equivalent with the single surprise in case we have inconsistent expectations (5).

Although we believe that our approach is a definite improvement of Francez' one, we admit that perhaps the logic is still not quite the best one can get. E.g., as we saw already, Proposition 4.1 (3) is a bit counter-intuitive, and also (1) is questionable: expectation of an expectation of φ collapses to just the expectation of φ. So, this means that the logic is not yet really fit to treat meta-expectations non-trivially. In this paper we will accept this disability to express non-trivial meta-expectations, and merely give some remarks about a possible solution, leaving a full treatment to future work. Nevertheless, we believe that this might still be done within the realm of modal logic, at the price of allowing more complicated models than just the simple S5-like ones we have introduced, as follows.

If one would like to distinguish expectations and (meta-)expectations of expectations, one may do this in our logic by dropping the appropriate axiom(s). In the case of denying $!!\varphi \equiv\, !\varphi$ it is sufficient to delete the axiom (10). If we want to deny $\lhd\varphi \equiv\, !\lhd\varphi$, it is the most reasonable thing to delete axiom (17), collapsing $PA\varphi$ and $A\varphi$ under the assumption that $\neg P$ **false**. In fact, axiom (17) states that a reference to the real world is transparent. (In other words, as we stated before, the A-operator is not pertaining to an epistemic but to an *ontological* notion.) If we want to consider meaningful nestings with respect to the P-operator(s), we must make this reference opaque, which is achieved by deleting (17). But then it is also necessary to delete axiom (15), $A\varphi \supset LA\varphi$, since otherwise we still have that $A\varphi \supset PA\varphi$.[10] Semantically, this amounts to abandoning the idea of having one fixed actual or real world, as we argued for in Section 1. Put in a different way, it means that we make the notion captured by the A-operator an *epistemic* one. However, when we do this we also need to adjust our definition (5) of the '**but**' operator, since in this case the modalities A and LA differ:

(5') φ **but** $\psi =_{def} L(\varphi \supset P\neg\psi) \wedge LA(\varphi \wedge \psi)$, or equivalently,
(5") φ **but** $\psi =_{def} L((\varphi \supset P\neg\psi) \wedge A(\varphi \wedge \psi))$,

---

[10]We may even consider alternatives such as replacing (17) by $P_iA\varphi \equiv P_i\varphi$, equating by correspondence the frame Σ_i with a singleton set containing a "real" world ρ_i, as locally perceived from the frame Σ_i. It is an advantage of using modal logic that correspondence theory is available to specify the models associated with these axioms, and we can change them as we please. However, by doing this one loses the crisp S5-like semantics of the logic. This is our main reason for not doing so in the first place. But this would also affect our simple reading of the operator L, cf. the discussion at the end of Section 1.

expressing that it is now in any situation (world) the case that if φ holds $\neg\psi$ is expected, but (in any situation) actually—in the real world that now may depend on the situation (world) in which you are evaluating, since this is not fixed any more—it is the case that both φ and ψ hold. One may verify that, using this new definition of the '**but**' operator in this more liberal set-up, the analysis of nestings of this operator we gave earlier, essentially remains the same. (Moreover, note that in case we impose (15) again, (5') becomes equivalent again with the original definition (5), so that (5') is indeed a consistent extension of (5) in the simple setting we adopted in section 2.) In conclusion we may say that using a modal logic that is K-like rather than S5-like, more refined notions of nested contrastive operators become expressible.

5. RELATION WITH DEFAULTS

As we have noted before, contrastive conjunctions are related to default statements[11]. For instance, in our example of Tweety, the fact that Tweety is a bird raised the expectation of Tweety being able to fly. As we have examined in [MH1,2,5], in default reasoning, when we know that "birds normally fly", we also raise the expectation of Tweety's flying abilities, when we say that Tweety is a bird. In fact, this default is the underlying explanation for the expectation that is raised in the first part of the "but"-sentence "Tweety is a bird, but it cannot fly". So, rather than saying that we can translate "but"-sentences into defaults, we argue that when interpreting these sentences (must? / may?) involve defaults that create the expectations. In the example above, "Tweety is a bird", together with the common-sense default "birds normally fly" in the background, creates the expectation that Tweety will fly.

Next, in contrastive statements our expectations are refuted immediately by the part following 'but'. Whereas in (pure) default reasoning we usually have some time assuming the defeasible knowledge that Tweety can fly, and taking decisions and actions accordingly (which is the very reason why default reasoning is useful), in a contrasting sentence this knowledge is defeated directly by the information that follows, (mostly) before we have had the time to take action. In the example, it is stated directly in the second ("but") part that "Tweety cannot fly".

---

[11]In the language **L** we express defaults of the form $\varphi : \psi / \chi$ (using Reiter's notation) as $\varphi \wedge M\psi \supset P\chi$. The reading of such a formula is "if φ is true and ψ is (considered) possible, then χ is preferred". When $\psi = \chi$ (where '=' stands for syntactical equality), the default is called *normal*. In [MH1] and [Me] we have shown several examples of how to use this formalism for defaults including the non-trivial Yale Shooting Problem ([HMcD]).

Another difference is that in default reasoning it is important to check whether the assertion that is to be believed by default (the conclusion that Tweety can fly, for example) is *consistent* with what is currently known. (Only in this case the default rule may be employed, cf. [Rei1,2], [MH1,2,5].) This appears to be less important in a contrastive sentence, since this sentence will typically be uttered in a situation in which it is already the case that the receiver will consider the expected conclusion possible, so that the latter will indeed experience a kind of contrast (surprise, or even shock) if s/he hears the "but"-part of the sentence. It is even somewhat strange to utter a contrastive sentence if it contains no contrast, witness the following example:

(**) Tweety is a bird, but it is a bird.

In our formal system this is represented by b **but** b, which is equivalent with $P\neg b \wedge Ab$, or, equivalently $\lhd b$: surprisingly, b. The surprise here is originating from the expectation that Tweety is not a bird, which should come from the *context* (it cannot be sensibly be expected to come from the statement that b is true). In fact, we think indeed that sentence (**) is only uttered meaningfully in a context where one expects $\neg b$ to be the case. It is still hard to imagine this, but consider a situation in which someone has heard a lot about a certain Tweety without knowing that it is a bird (say, thinking it is a person). Then s/he learns something about Tweety which he thinks is really funny for a person, e.g. that Tweety likes bird seed. Then in a flash it dawns upon to him, and utters surprised: "Ah! Tweety is a *bird*, but it is a *bird*!". (We agree that whether this rather Monty Python-like scenario is really convincing depends on the degree of emphasis given to the word "bird" when uttering this sentence.) In this context this sentence has the same meaning as "Ah! But it is a *bird*!", which more or less corresponds to the formula "**true but** b", amounting to the same formula $P\neg b \wedge Ab$.

As having said already, a further similarity between defaults and contrastive sentences is the possibility of nesting contrast, which appears to be related to the issue of specificity in default theory. This is best illustrated by an example: consider a contrastive sentence like "Bob is an adult, but he is a student, but he is clever", uttered in the context of Bob's being employed, is strongly related to the default theory:

(1) Normally, adults are employed;
(2) Normally, students are unemployed;
(3) Normally, clever students are employed (as a teaching assistant)

If we now have a clever student, named Bob, we like to derive from this default theory that he is expected to be employed, since this is the most specific information. (In the context of nonmonotonic reasoning specificity is an important issue, cf. [MH2,4].) This is also the "highest level surprise" (in the sense of section 4) in the contrastive sentence. Lower level surprises correspond to the outcomes of applying less specific defaults. So, more precisely, in a conversation on Bob's state of employment, the utterance "Bob is an adult" creates the expectation on the basis of default (1) that he is employed, which is then refuted by the utterance "Bob is a student", in its turn, by default (2), raising the expectation that he is unemployed. In the end, by uttering "Bob is clever" this is again refuted, using default (3), which suggests again that Bob must be employed. So here again we have a correspondence between expectations and defaults in the sense that the default theory functions as a common-sense background theory for the hearer of the contrastive sentence in order to evaluate the implications (with respect to Bob's state of employment) of what is uttered by the speaker.

The relation with defaults may also help in further explaining the issue of asymmetry of "but"-sentences, that we discussed in section 2. For example, consider the two statements

(i) Tweety is a bird, but it does not fly.
(ii) Tweety does not fly, but it is a bird.

If we abbreviate (i) by b **but** $\neg$f, and (ii) by $\neg$f **but** b, in our approach (i) is equal to $L(b \supset Pf) \wedge A(b \wedge \neg f)$, and (ii) is equal to $L(\neg f \supset P\neg b) \wedge A(\neg f \wedge b)$. So, in terms of background knowledge defaults, (i) involves the default "birds normally fly" (amounting to the (refuted) expectation that Tweety flies), whereas (ii) involves the default "non-flying things normally are non-birds" (yielding the (refuted) expectation that Tweety is not a bird). So this has to do with contraposition of defaults. Interestingly, in the literature this is a controversial issue, and some approaches in the literature to default reasoning do not allow/yield contraposition of defaults, including our own and some very well-known ones (cf. [Gin] and [MH1, 2, 5]).

Finally, we mention work strongly related to ours, [S], which is motivated from a completely different point of view, viz. that of system specification, where contrastive sentences such as φ **but** ψ are interpreted in such a way that φ is a default sentences itself: normally, it holds that φ, and (but) now we have that ψ. For example, in [S] an example of the form $p \wedge q$ **but** $\neg q$ is given: "JR and Sue Ellen [usually] come at parties, but Sue Ellen is not coming tonight". This then is given meaning by means of a nonmonotonic logic, where actualities take precedence over default beliefs. Although similar in spirit to our approach—our definition of φ **but** ψ, too, expresses that φ creates an expectation / default belief, which is then refuted by ψ—there are also differences: in our view φ is not (necessarily) a default itself, but yields an expectation (or default belief) together with implicit background information. This is exemplified by sentences of the kind "Tweety is a bird, but s/he does not fly". Obviously, in this sentence "Tweety is a bird" is not intended to denote a kind of default itself, but *yields* the default belief that Tweety can fly by means of an implicit default in the background, viz. "Normally, birds can fly". In our view, the sentences that are considered in [S] are special cases of "but"-sentences φ **but** ψ, where φ is itself of the form $P\varphi'$, expressing an expectation or default. (In our approach we would represent the above example by $P(p \wedge q)$ **but** $\neg q$, which is equivalent with $L(P(p \wedge q) \supset Pq) \wedge A(P(p \wedge q) \wedge \neg q)$, i.e., **true** $\wedge P(p \wedge q) \wedge A\neg q$, which is intuitively right.) A second difference is that in our approach the results of expectations still retain their informative value and are not overridden by the actual information. We believe this to be important with respect to assessing the precise meaning of a contrastive sentence, since otherwise the meaning of the contrastive sentence $p \wedge q$ **but** $\neg q$ in the example above would exactly be the same as that of the non-contrastive sentence $p \wedge \neg q$, *quod non*, obviously.

6. CONCLUSION

In the setting of the modal logic $\mathbf{S5AP_{(n)}}$ we have shown how to represent contrastive natural language conjunctions such as 'but'. By doing so we obtain a complete logic for contrastive connectives. Our approach is more in line with the tradition of treating intensional operators (using modal logic and Kripke-style semantics) than that of [F], and, as a consequence of this, we obtain a complete axiomatization of these operators without much difficulty. Moreover, it has some advantages such as the non-symmetry of 'but', and a more straight-forward treatment of nested contrastive operators. We also claim that the operators in our approach behave somewhat more intuitively than in Francez' approach.

One might object that the meaning of 'but' is still not established conclusively in our paper, since one might still question the validity of the various properties ascribed to 'but' in our setting. Although we think ourselves that most of these properties are intuitively right and desirable, we appreciate that these might be subject to some debate. In our opinion it is inherent to natural language constructs such as a connective as 'but' that their semantics contain some elements of subjective interpretation and oversimplification, so that the establishment of their exact and "right" meaning will always be controversial to some degree. We hope to have offered a first (or rather second) step towards a logical understanding of 'but' by the use of modal logic. In particular, although we admit that also our present approach to nested contrasts might need further refinement, our claim is that the paradigm of modal logic offers sufficient flexibility to do this, since (as one of the anonymous referees of this paper has put it) it allows for finer distinctions and several variants of the same natural language construct, each capturing a different meaning.

ACKNOWLEDGEMENTS

The authors are indebted to Nissim Francez for his inspiring paper [F], which in an obvious way paved the way for the approach expounded in the present paper. Moreover we had some stimulating discussions with him on the subject, some of them "on stage" during the BISFAI'91 conference. Furthermore, we would like to thank the editor Martin Golumbic, two anonymous referees and Bernd van Linder for their very helpful criticism and comments on an earlier version of this paper. Last but not least, we are indebted to Pierre-Yves Schobbens, who provided us with Proposition 2.1 and several of its consequences, some of which were rather surprising to us.

POSTSCRIPTUM

After having completed this paper we have received a reaction, mediated by the editor, which involved follow-up work by Winter & Rimon [WR], where the subject of contrastive conjunctions was taken up from a more general linguistic perspective. Winter & Rimon provide a semantic theory of several of such conjunctions including, apart from 'but', also 'although', 'yet' and 'nevertheless'. Although "inspired by" both our and Francez' paper, they disagree with us on some central points.

First of all, they discuss a richer range of possible contrasts: contrasts regarding expectations as we treat them, constitute only one of these, but also "rhetorical contrast" can be expressed by 'but': "I wrote it in agony, but I wrote it". We agree directly that in our rather straightforward approach we do not distinguish these refined notions of contrast, but we *do* think that in our modal approach we can analyze formulas of the form "$(\varphi \wedge \psi)$ **but** φ" to the extent that *some* form of contrast is appreciated adequately. By Prop. 2.1, $(\varphi \wedge \psi)$ **but** φ amounts to $P\neg\varphi \wedge A(\varphi \wedge \psi)$, expressing that it was expected that φ does not hold although actually it was true together with ψ. In the example this comes down to the statement that apparently it was expected not to write it, although actually I did so while in agony. Furthermore, the formula above is equivalent with $\lhd\varphi \wedge A(\varphi \wedge \psi)$, implying that it was quite a surprise that I wrote it. We believe that this is an acceptable outcome in this case.

Moreover, in [WR] a distinction is made between *direct* contrast and *indirect* contrast. Thus, in our terms, introducing the additional contrastive operator '**nevertheless**', we would have φ **nevertheless** $\psi = L(\varphi \supset P\neg\psi) \wedge A(\varphi \wedge \psi)$, pertaining to *direct* contrast, while φ **but** ψ would give us something like $\exists\chi$: $L(\varphi \supset P\neg\chi) \wedge A(\psi \supset \chi) \wedge A(\varphi \wedge \psi)$, expressing some kind of *indirect* contrast. We do agree that this is perhaps more adequate, but wonder about the status of this "higher-order" quantification. Even more importantly, although in [WR] the conjunction 'but' (or rather its direct version 'nevertheless') is given a semantics similar to our definition of φ **but** $\psi = L(\varphi \supset P\neg\psi) \wedge A(\varphi \wedge \psi)$, the authors split this formula in a standard truth condition part ("$\varphi \wedge \psi$") and a presupposition part ("$\Diamond(\varphi \supset \neg\psi)$"), as is rather standard in natural language semantics. Being no linguists, we wonder whether this is really needed in our case where we consider a fairly rich modal language (see below).

Furthermore, Winter and Rimon point at some problems due to the use of material implication. It has as a consequence that the rules $\Diamond\neg\varphi / \Diamond(\varphi \supset \psi)$ and $\Diamond\psi / \Diamond(\varphi \supset \psi)$ are valid that, in their turn, cause problems with the presuppositions in the treatment of [MR]. This has led Winter & Rimon to incorporate a much more sophisticated and intricate semantics regarding conditionals, viz. Veltman's data semantics [V]. (Their modal representation of 'but', on the other hand, is less rich: no counterparts of the A- and P-modalities are used, and instead of the necessity-type L-operator a possibility-type operator '$\Diamond$' ("may") is employed.) This semantics might well be more adequate for some aspects of contrastive conjunctions. However, we are not totally convinced by the argument given, since this concerns a rather weak modal (possibility-type) operator '$\Diamond$' (much weaker than the (necessity-type) L-operator that we employ): although our L-operator satisfies the same problematic rule as '$\Diamond$', it is more harmless, since the antecedent of the rule is stronger and harder to satisfy.

Finally, although Winter & Rimon adhere to the importance of nestings of contrast operators, they believe that "the empirical status of 'genuine' nesting of contrast is not completely clear" and that "a full analysis of the formal aspects of nesting would have dictated an entirely different approach" ([WR], p.385). As will be clear from our paper, we agree entirely that this matter deserves much more attention.

REFERENCES

[Ch] B.F. Chellas, Modal Logic: An Introduction, Cambridge University Press, Cambridge / London, 1980.

[F] N. Francez, Contrastive Logic, Techn. Report No. 668, Technion, Haifa, Israel, 1991.

[Gin] M.L. Ginsberg, Introduction of "Readings in nonmonotonic reasoning", Morgan Kaufmann, Los Altos, 1987.

[HM] J.Y. Halpern & Y. Moses, A Guide to the Modal Logics of Knowledge and Belief, Proc. IJCAI'85 (1985), pp. 480-490.

[HMcD] S. Hanks & D. McDermott, Nonmonotonic Logics and Temporal Projection, Artificial Intelligence 33 (1987), pp. 379-412.

[Hin] J. Hintikka, Knowledge and Belief, Cornell University Press, Ithaca NY, 1962.

[H] W. van der Hoek, Systems for Knowledge and Belief, J. of Logic and Computation 3(2), 1993, pp. 173-195.

[HC] G.E. Hughes & M.J. Cresswell, An Introduction to Modal Logic, Methuen, London, 1968.

[HC2] G.E. Hughes & M.J. Cresswell, A Companion to Modal Logic, Methuen, London, 1984.

[Me] J.-J. Ch. Meyer, An Analysis of the Yale Shooting Problem by Means of Dynamic Epistemic Logic, in: Proc. 7th Amsterdam Colloquium, (M. Stokhof & L. Torenvliet, eds.), ITLI, University of Amsterdam, 1990, pp. 317-326.

[MH1] J.-J. Ch. Meyer & W. van der Hoek, Nonmonotonic Reasoning by Monotonic Means, in Proceedings JELIA '90 (J. van Eijck, ed.), LNCS 478, Springer, 1991, pp. 399-411.

[MH2] J.-J. Ch. Meyer & W. van der Hoek, A Modal Logic for Nonmonotonic Reasoning, in: Non-Monotonic Reasoning and Partial Semantics (W. van der Hoek, J.-J.Ch. Meyer, Y.H. Tan & C. Witteveen, eds.), Ellis Horwood, Chichester, 1992, pp. 37-77.

[MH3] J.-J. Ch. Meyer & W. van der Hoek, Counterfactual Reasoning by (Means of) Default, Annals of Mathematics and Artificial Intelligence 9 (III-IV), 1993, pp. 345-360.

[MH4] J.-J. Ch. Meyer & W. van der Hoek, An Epistemic Logic for Defeasible Reasoning Using a Meta-Level Architecture Metaphor, VU-Report IR-329, Amsterdam, 1993.

[MH5] J.-J. Ch. Meyer & W. van der Hoek, A (Cumulative) Default Logic Based on Epistemic States, VU-Report IR-288, Amsterdam, 1992; extended abstract in: M. Clarke, R. Kruse & S. Moral (eds.), Symbolic and Quantitative Approaches to Reasoning and Uncertainty (Proc. ECSQARU '93, Granada), Springer-Verlag, Berlin, 1993, pp. 265-273.

[MH6] J.-J. Ch. Meyer & W. van der Hoek, Epistemic Logic for AI and Computer Science, Cambridge University Press, forthcoming.

[Mo] R.C. Moore, Semantical Considerations on Nonmonotonic Logic, Artificial Intelligence 35 (1988), pp. 75-94.

[Rei1] R. Reiter, A Logic for Default Reasoning, Artificial Intelligence 13, 1980, pp. 81-132.

[Rei2] R. Reiter, Nonmonotonic Reasoning, Annual Reviews of Comp. Sc. 2, 1987, pp. 147-187.

[S] P.-Y. Schobbens, Exceptions for Algebraic Specifications: On the Meaning of "But", Science of Computer Programming 20, 1993, pp. 73-111.

[S2] P.-Y. Schobbens, Personal communication.

[V] F. Veltman, Data Semantics and the Pragmatics of Indicative Conditionals, in: E.C. Traugott (ed.), On Conditionals, Cambridge University Press, Cambridge, 1986.

[WR] Y. Winter & M. Rimon, Contrast and Implication in Natural Language, Journal of Semantics 11, 1994, pp. 365-406.