An Infinite Number of Monkeys*

David Beaver
dib@illc.uva.nl
Centre for Cognitive Science, University of Edinburgh
Department of Philosophy, University of Amsterdam

June 1995

Abstract

Many theories of presupposition fail to account for the pragmatic nature of presuppositional justification. An alternative account will be presented in which accommodation is essentially non-structural: the material to be accommodated cannot be predicted purely by consideration of the form of the utterance, but instead makes essential use of world-knowledge and common-sense reasoning. The account will essentially consist of the extension of (a reformulated version of) the Context Change Potential (CCP) model due to Heim, Karttunen and Stalnaker, with a pragmatic module.

1 Introduction

Life is short. There is not enough time to explain everything. As speakers, or writers, we are forced to make assumptions. It is common to be advised to fix in ones mind a picture of the audience, that is, to make an advance decision as to what the audience can be expected to know. Often, especially given limitations of time for speaking or space for writing, one is forced to take much for granted. As a result, cases of presupposition failure, the situation occurring when the speaker or writer takes for granted something of which the hearer or reader is not previously aware, are surely common. Somehow, hearers and readers cope, and usually without complaining.

The author, in most *genres*, assumes that the text will be read linearly, and further assumes, optimistically, that readers will gather information throughout

*A version of this paper is due to appear in a special issue of *Acta Linguistica Hungarica* devoted to proceedings of the 5th Hungarian Conference on Logic and Language, which took place in Noszvaj, August 1994. It is a shorter version of a manuscript currently in preparation "A Plea for Common Sense."

the reading process. So what the author has is not a fixed picture of the common ground with the intended readership, but a rather rough cut and idealised movie of how this common ground should develop. Each frame in the movie approximates what is common between relevant aspects of (1) the author's beliefs at time of writing, and (2) the readers' beliefs as they reach some point in the text.

At risk of straining the cinematic metaphor somewhat, it could be said that the text itself is analogous to a script, but with detailed screenplay and directorial instructions omitted. Barring a major scientific breakthrough, the corresponding film will never be put on general release, so precisely how the writer intends the information state of idealised readers to evolve as they read is never made public in all its technicolour glory.

Shortly (in §3), I will discuss how readers' information states do evolve, as they construct their own movies on the basis of the script. In directing their own films readers second-guess the intentions of the original writer-director. But to understand how readers work out the writer-director's intentions, it is firstly necessary to know more of the craft of the writer-director. In the coming section I will elaborate on how, working from an assumption as to initial conditions, the author envisages the evolution of the common ground. It will be helpful to adopt some of the formalism of recent dynamic semantics. I will build particularly on ideas of Stalnaker [Sta74], Karttunen [Kar74] and Heim [Hei83], and use formal techniques related to those discussed by Groenendijk and Stokhof, e.g. [GS91], and Veltman [Vel91].

2 The Writer's View of the Common Ground

I will now present what will here be called Presupposition Logic, a simple propositional system with a dynamic semantics and dynamic notion of semantic entailment. Further discussion and motivation can be found in [Bea92, Bea93b, Bea94, Bea95].

Presupposition Logic provides a model of how a speaker or author envisages the common ground evolving. This evolution is iterative, since the common ground at any instant provides the context in which a given chunk of text is interpreted, and it is the effect of this interpretation which determines what the common ground will be going into interpretation of the next chunk. It is no longer controversial to assert that the interpretation process relies on such iteration, but there remains some question as to the course-grainedness of the iteration. For instance in the work of Gazdar [Gaz79] (and also related proposals such as Mercer's [Mer92]) it is whole sentences which produce a change in the context of interpretation. However in Karttunen-Heim style treatments of presupposition such as that introduced in this paper, as well as in treatments of anaphora due to Heim [Hei82], Kamp [Kam81] and Groenendijk and Stokhof [GS91], a finer grained iteration is involved, with sub-sentential constituents producing their own effects on local contexts of interpretation.

We begin by assuming some set of atomic proposition symbols. A model is a pair $\langle W, I\rangle$, where W is a set of worlds and I is an interpretation function mapping each atomic proposition symbol to a subset of W. The Context Change Potential (to borrow Heim's terminology) of a formula ϕ written $[\![\phi]\!]$ is a set of pairs of input and output contexts, where a context is the writer's view of the common ground. Following Stalnaker, a context is thought of as a set of possible worlds, the set containing all and only those worlds compatible with the information supposed to be common. I will write $\sigma[\![\phi]\!]\tau$ to indicate that the pair $\langle\sigma,\tau\rangle$ is in the $[\![\phi]\!]$ relation, i.e. an input context σ updated with the information in ϕ can produce an output context τ.

Defining contexts as subsets of W introduces a natural lattice structure with union and intersection as meet and join, and this lattice provides to an intuitive model-theoretic characterisation of the amount of information in a given context. A minimal context (with respect to some model) may be defined as the set of all worlds (in that model): this is the state of blissful ignorance in which no information about the world is available. Similarly a maximally informative non-contradictory context would be a singleton set: the available information rules out all except one world. However, it is possible to add even more information to such a context, and in case this information contradicts previous information we will arrive at a context containing no worlds, a truly maximal but contradictory context.

Definition D1, below, gives an update semantics for the language of propositional logic. The first clause says that the result of updating a context with an atomic proposition is an output containing only those worlds in the input which are in the extension of the proposition. A context can be updated with a conjunction of two formulae (the second clause) just in case it can firstly be updated with the left conjunct to produce an intermediary context (v in the definition), and this context can be updated with the right-hand conjunct to produce the final output (τ). The third clause says that a context can be updated with the negation of a formula just in case there is some state that can be obtained by updating the context with the negated formula itself, in which case the result of updating with the whole formula is the set of worlds in the input which are not present in the update with the negated formula. In other words, the effect of updating with the negation of a formula is to remove all information compatible with the formula. The final two clauses define implication and disjunction by (carefully selected) standard equivalences.

Definition D1 **(Update Semantics for Propositional Logic)** *For all models* $\mathcal{M} = \langle W, I\rangle$ *and information states* σ, τ, *the relation* $[\![.]\!]^{\mathcal{M}}$ *(superscript omitted where unambiguous) is given recursively by:*

$$
\begin{array}{lrcl}
(1) & \sigma[\![p_{\text{atomic}}]\!]\tau & \textit{iff} & \tau = \{w \in \sigma \mid w \in I(p)\} \\
(2) & \sigma[\![\phi \wedge \psi]\!]\tau & \textit{iff} & \exists v \;\; \sigma[\![\phi]\!]v[\![\psi]\!]\tau \\
(3) & \sigma[\![\neg\phi]\!]\tau & \textit{iff} & \exists v \;\; \sigma[\![\phi]\!]v \;\wedge\; \tau = \sigma \backslash v \\
(4) & \sigma[\![\phi \rightarrow \psi]\!]\tau & \textit{iff} & \sigma[\![\neg(\phi \wedge (\neg\psi))]\!]\tau \\
(5) & \sigma[\![\phi \vee \psi]\!]\tau & \textit{iff} & \sigma[\![\neg(\neg\phi \wedge \neg\psi)]\!]\tau
\end{array}
$$

A context σ satisfies a formula ϕ ($\sigma \models \phi$) if updating adds no new information, producing an output identical to the input. One formula entails another ($\phi \models \psi$) if any update with the first produces a context in which the second is satisfied:

Definition D2 **(Satisfaction and Dynamic Entailment)**

$$
\begin{array}{lcl}
\sigma \models \phi & \textit{iff} & \sigma[\![\phi]\!]\sigma \\
\phi \models \psi & \textit{iff} & \forall \sigma, \tau \;\; (\sigma[\![\phi]\!]\tau \Rightarrow \tau \models \psi)
\end{array}
$$

Over the standard propositional language this notion of entailment is classical. However, we will now extend the language with a presupposition operator ∂: the resulting logic will be non-classical. For example, commuting conjunctions will no longer uniformly preserve validity. The intuition behind the following definition is that a formula $\partial\phi$ ("the presupposition that ϕ") places a constraint on the input context, only allowing update to continue if the presupposed proposition is already satisfied.

Definition D3 **(Presupposition Logic)** *Presupposition Logic is defined over the language of Propositional Logic with an extra unary operator* ∂. *It has the dynamic notion of semantic entailment above, and semantics consisting of the update semantics for atomic propositions and standard connectives combined with the following interpretation for* ∂*-formulae:*

$$
\sigma[\![\partial\phi]\!]\tau \quad \textit{iff} \quad \sigma \models \phi
$$

As an example the sentence "Mary realises that John is sleepy" might be said to correspond to a formula in Presupposition Logic of the form $\partial p \wedge q$, where p is atomic proposition that John is sleepy, and q is an atomic proposition that Mary has come to believe that John is sleepy. In this paper I will not be concerned with the details of how such logical forms may be derived from natural language, or with the question of whether it is reasonable to use a representation in which presuppositions are divided explicitly from assertions using the ∂-operator. However, both these issues are dealt with elsewhere: see

[Bea93a, Bea93b, Bea95] where two sorted versions of classical type theory are used to provide a Presupposition Logic style semantics for a fragment of natural language.

Suppose a formula contains a presuppositional sub-formula. What will the presuppositions of the whole formula be? This is the presupposition projection problem of Langendoen and Savin [LS71], except applied to Presupposition Logic rather than natural language. A formula $\partial\phi$, "the presupposition that ϕ", defines an update if and only if ϕ is satisfied, so it is natural to say that in general a formula presupposes all those formulae that must be satisfied by the input context in order for there to be an update. We say that a context σ admits a formula ϕ (written $\sigma \rhd \phi$) if and only if it is possible to update σ with ϕ, this being a formalisation of Karttunen's notion of admittance in [Kar74]. In that case one formula ϕ presupposes another ψ (written $\phi \gg \psi$) just in case every context that admits the first satisfies the second. Note that admittance provides a counterpart to so-called *presupposition failure*, what happens when updating cannot continue because presuppositions are not satisfied.

Definition D4 **(Admits ($\rhd$) and Presupposes ($\gg$))**

$$\begin{array}{lll} \sigma \rhd \psi & \textit{iff} & \exists\tau\sigma[\![\phi]\!]\tau \\ \phi \gg \psi & \textit{iff} & \forall\sigma\ \sigma \rhd \phi \Rightarrow \sigma \models \psi \end{array}$$

It is clear that the definition's for the semantics of the ∂-operator and the meta-logical $\gg$ relation are closely related. For example we have that for any ϕ, $\partial\phi \gg \phi$. Indeed, $\gg$ could have been equivalently defined in terms of ∂, defining $\phi \gg \psi$ *iff* for some χ, $[\![\phi]\!] = [\![\partial\psi \wedge \chi]\!]$. We can now study projection in Presupposition Logic. As detailed in F1, the system behaves just as anyone familiar with Karttunen's 1974 system would expect. In particular, note that a formula may fail to carry a presupposition of one of its component subformulae, but instead carry a logically weaker conditionalised variant.

Fact F1
If $\phi \gg \psi$ then:

$$\begin{array}{rcl} \neg\phi & \gg & \psi \\ \phi \wedge \chi & \gg & \psi \\ \phi \rightarrow \chi & \gg & \psi \\ \phi \vee \chi & \gg & \psi \\ \chi \wedge \phi & \gg & \chi \rightarrow \psi \\ \chi \rightarrow \phi & \gg & \chi \rightarrow \psi \\ \chi \vee \phi & \gg & (\neg\chi) \rightarrow \psi \end{array}$$

Before going on, I should perhaps point out a shortcoming of Presupposition Logic *qua* model of natural language. Whilst Presupposition Logic is like natural language in allowing both assertion and presupposition of information, it is very unlike natural language in that presuppositions in Presupposition Logic seem so pointless. Granted, they have some interesting logical properties, as F1 shows, but further they seem somewhat gratuitous. In natural languages, the phenomenon of presupposition is not an arbitrarily occurring feature, but an essential part of the way language works. Presuppositions are not simply poor cousins that hang around whilst assertions do the real work of communication. Rather, presuppositions can serve an essential function in determining what actually is asserted. For instance, a definite description "the king of France", would ordinarily (say in a history lesson) be used to identify an object so that something can be predicated of it, and the presuppositions associated with the description, of existence and maybe uniqueness of the object, are not incidental, but essential to the functioning of the description. In Presupposition Logic, presuppositions do not play such important roles. However, we will see in the next section that presuppositions like those of Presupposition Logic could still play a part in the communication process. Although such presuppositions do not directly help to establish what is asserted, as is the case in natural language, they can at least provide a murky reflection of what has been taken for granted by the writer, that is, they can give readers an insight into what the writer took to be the initial common ground.

3 The Naive Reader

Readers who updated their own information state according to principles like those behind the semantics of Presupposition Logic would be stymied whenever some information was presupposed which they did not have. In this situation the reader would lack any further means of updating. So Presupposition Logic, as it stands, does not provide a good model of the evolution of the information state of a hearer or reader. But suppose that you only had a Presupposition Logic-like semantics to help you understand a text. How would you use that semantics to glean information?

If an infinite number of monkeys with typewriters were given time, some of them might produce this text. What if an infinite group of monkeys schooled in the writer-director approach to writing described above, were given some rhetorical goal (for instance, to produce my contribution to the Noszvaj symposium) but no description of the intended audience? Not knowing the audience, they might choose the initial common ground randomly, although after that each monkey's view of the common ground at a particular point in the text would be fully determined by what they typed. Now although as a reader you do not know what initial common ground has been assumed when you read a text, you can reason for any particular choice of initial conditions how the common ground would evolve. And this leads me to a suggestion for how a dynamic semantics like that given for Presupposition Logic could be used to understand a text without presupposition failure being problematic.

Begin the reading process by imagining an infinite number of monkeys (or as many as you can manage) with an infinite number of assumed initial common grounds, your goal being to find out which monkey wrote the text. As you read, separately update each of these contexts. At various stages presuppositional constructions will be encountered, and these are what sort out the wheat from the chaff. For whenever something is presupposed which has not been explicitly introduced earlier in the text, a number of monkeys drop out of contention, and there remain only those monkeys for which the assumed common ground corresponding to that point in the text satisfies the presupposition. In general, this process may not tell you exactly which monkey was responsible, but it will at least limit the options, and it will simultaneously tell you quite a lot about what information you were intended to have after reading the text.

Let us consider a simple example concerning a language with only two proposition letters p and q. Assume a model with four possible worlds, w^{pq}, w^{p}, $\mathrm{w}^{q}$and w: in the first both p and q hold, in the second only p, in the third only q, and in the fourth, neither. Monkeys can be identified with the set of worlds which they associate with the common ground, so each monkey's assumed context corresponds to a subset of $\{\mathrm{w}^{pq}, \mathrm{w}^{p}, \mathrm{w}^{q}, \mathrm{w}\}$. Thus there are 16 relevantly different monkeys, each of which we can represent using four digit binary numbers. Let the units, twoes, fours and eights digits correspond to worlds w, w^{q}, $\mathrm{w}^{p}$and w^{pq}respectively, such that a given digit is 1 if and only if the corresponding world is in the assumed common ground. For instance, monkey 1010 assumes that the common ground contains worlds $\mathrm{w}^{pq}$and w^{q}.

You read a text $\partial p \wedge q$, and wish to know which monkey wrote it. It is easily verified, from the semantics presented in the previous section, that the only pairs in $[\![\partial p \wedge q]\!]$ are $\langle 0000, 0000\rangle$, $\langle 0100, 0000\rangle$, $\langle 1000, 1000\rangle$ and $\langle 1100, 1000\rangle$. In words: an inconsistent context (0000) remains inconsistent after update with the formula; a context containing only world w^{p}can be updated, but this world is a non-q world, and is eliminated by the update with q, producing an inconsistent context; an update with an input context containing only world w^{pq}returns the input as output; a context containing worlds $\mathrm{w}^{pq}$and w^{p}, thus a context where p is established but q is still uncertain, can be updated with the effect of removing world w^{p}, the non-q world, to produce a context containing only world w^{pq}. No other contexts can be updated. For instance it is not possible to update the context 1011, which satisfies neither p or q, because the presence of non-p worlds causes presupposition failure. So the semantics of presupposition logic tells us that one of the monkeys 0000, 0100, 1000 or 1100 must have written the text.[1]

[1]One could use simple Gricean principles to reason further: the maxim of informativeness tells us that some information should have been transmitted. Thus neither 0000 (a state from which no further update is possible) nor 1000 would have been written by pragmatically sophisticated monkeys, since they both produce identity updates. If the maxim of *quality* is taken to rule out updates with contradictory information, then 0000 is ruled out once again (because it contradicts all information), and 0100 is also ruled out, because update then leads

The many-monkeys strategy can easily be formalised in terms of the semantics of Presupposition Logic. A reader's information state is identified with a set of contexts — I will use the term *information set* — and is thus a subset of the powerset of worlds.[2] A state can be updated with a formula by updating each of the member contexts separately, so producing the following definition of the update of a state I with a formula ϕ:

Definition D5 **(Updating Information Sets)**

$$I + \phi = \{\tau \mid \exists \sigma \in I \ \sigma[\![\phi]\!]\tau\}$$

By definition, let us say that an information set satisfies a formula only if its member contexts satisfy the formula:

Definition D6 **(Satisfaction by an Information Set)**

$$I \models \phi \quad \textit{iff} \quad \forall \sigma \in I \ \sigma \models \phi$$

The earlier notion of entailment could easily be defined in terms of the new notion of information, as the following fact demonstrates. (Here $\mathcal{P}(\mathcal{W})$ is the powerset of the set of worlds W.)

to the inconsistent state. Thus the only remaining monkey would be 1100, who assumed p and used the formula to communicate that q. In spite of the appeal of this sort of approach to formalisation of Grice's work (an approach which is perhaps reminiscent of that adopted by van der Sandt [vdS92]) I will pursue it no further here.

[2] Observe that two assumptions have been implicitly made, firstly that the author and the reader have agreed upon a common set of atomic proposition letters, and secondly that they use identical models (up to isomorphism). Without this latter assumption, there would be no question of the reader speculating as to which worlds were in the author's assumed common ground, for there would be no sense in which the reader could ever know anything of the author's worlds or *vice versa*. One might ask whether these assumptions, in effect that author and reader use the same vocabulary and lexical semantics, are significant. Adding uncertainty about the vocabulary would complicate technically, but it seems that not much would be gained. For there is no technical reason why the chosen set of atomic proposition symbols could not be much larger than the vocabulary that could conceivably be used, say the set of all finite sequences of Roman characters and Arabic numbers. The assumption of a common model is more metaphysically puzzling, but also, I will argue, need not be problematic.

There is a natural ordering amongst the models of Presupposition Logic, some being more constrained than others. An unconstrained model would be such that for every world w and proposition symbol p, there is another world w' which is in the extension of exactly the same proposition symbols as w, except for being in the extension of p only if w is not, and *vice versa*. A term model constructed by identifying each subset of the proposition symbols with a world, and identifying the extension of each proposition symbol with the set of worlds containing that symbol, would in this sense be completely unconstrained. A model would be more constrained if there were certain sets of proposition symbols such that no world was in the extension of all and only those symbols. Meaning postulates produce constrained models. For current purposes we may suppose that the model is unconstrained. This means only that we have not stated any special conventions about how proposition symbols should be interpreted, and we are, in a sense, in a state of mutual ignorance. In choosing an unconstrained model, we assume only that there are no special assumptions. To me, at least, this does not seem worrying.

Fact F2

$$\begin{array}{lll} \phi \models \psi & \textit{iff} & (\mathcal{P}(\mathcal{W}) + \phi) \models \psi \\ & \textit{iff} & \forall I\ (I + \phi) \models \psi \end{array}$$

To return to the earlier example, we have the following update: $\{0000, \ldots, 1111\}\ +\ \partial p \wedge q\ =\ \{0000, 1000\}$. The reader who accepts the veridicality of the text can then learn both p and q, since both these propositions are satisfied in $\{0000, 1000\}$. However, this is not to say that there is no longer a difference between presupposed and asserted formulae. Consider update with the negation of the previous example formula. We have that $\{0000, \ldots, 1111\}\ +\ \neg(\partial p \wedge q) = \{0000, 0100\}$, and whilst the result state $\{0000, 0100\}$ still satisfies p, it no longer satisfies q: in fact it satisfies $\neg q$. Thus the presupposition (that p) is effectively projected out of the negation, whilst the assertion, as would be expected, is not.

4 The Sophisticated Reader

The naive reader might imagine an infinite number of monkeys, and use only information from the text to help find out which monkey is the author. But other information is available, if not of an absolute character. We cannot initially say of any given proposal as to the assumed common ground that it is impossible, and to this extent it is necessary to consider all possibilities. But we can say that some proposals are relatively more plausible than others. The sophisticated reader considers what assumption the author is *likely* to have made as to the initial common ground.

I cannot say how we decide what the author is likely to have assumed. One strategy would be to guess that the world about which the author was trying to communicate obeys similar, or identical, general principles as your own world does, and that the author had a similar knowledge of those principles. That is, you could guess that the author's common sense knowledge is similar to your own. In fact the model I will now propose will be of sufficient generality that the particular method of guessing at the author's assumptions will not be important. The assumptions of the author, whatever they are, determine a Presupposition Logic context, a set of worlds. A reader's knowledge of which assumptions are most plausible determines an ordering over these contexts, what I will call a *plausibility ordering.* A plausibility ordering relative to some model is a reflexive, transitive binary relation over a subset of the powerset of the set of worlds.[3]

[3] Laszlo Kalman, in a talk presented at the University of Amsterdam, also proposed a treatment of presupposition involving default knowledge encoded using orderings. One difference is that he envisaged a model in which possible worlds were ordered rather than a model in which sets of worlds are ordered. However, I think we are in agreement on the basic issues, in particular on the fact that some encoding of default knowledge must be used in the treatment of presupposition.

For an ordering π, $\sigma \geq_\pi \tau$ is written for $\langle\sigma,\tau\rangle \in \pi$, and $\sigma >_\pi \tau$ is taken to mean that both $\sigma \geq_\pi \tau$ and $\tau \not\geq_\pi \sigma$. An ordering π can be updated with a new formula by considering every pair in the ordering, and updating each element of the pair separately according to the principles of Presupposition Logic. The following definition is obtained:

Definition D7 **(Updating Plausibility Orderings)**

$$\pi + \phi \quad = \quad \{\langle\sigma,\tau\rangle \mid \exists\langle\sigma',\tau'\rangle \in \pi \; \sigma'[\![\phi]\!]\sigma \;\wedge\; \tau'[\![\phi]\!]\tau\}$$

Under this definition, certain contexts may drop out of contention in the update process, just as with the naive updating process considered earlier. An example may clarify. Suppose that $[\![\phi]\!] = \{\langle\sigma,\sigma'\rangle, \langle\upsilon,\upsilon'\rangle\}$, and that we wished to update the ordering $\pi = \{\langle\sigma,\sigma\rangle, \langle\tau,\tau\rangle, \langle\upsilon,\upsilon\rangle, \langle\sigma,\tau\rangle, \langle\sigma,\upsilon\rangle, \langle\tau,\upsilon\rangle\}$ with ϕ. Returning to the earlier metaphor, the reader is considering three different *movies* that the writer-director might have intended, and at the current point in the text the candidates for the correct frame are σ, τ and υ, with a plausibility ordering $\sigma \geq_\pi \tau \geq_\pi \upsilon$. The reader should now verify that $\pi + \phi = \{\langle\sigma',\sigma'\rangle, \langle\upsilon',\upsilon'\rangle, \langle\sigma,\upsilon\rangle\}$. Observe that since τ cannot be updated with ϕ, there is no next frame from the film containing τ in the new ordering, and we are left with only two candidate films, with current frames σ' and υ' ordered $\sigma' \geq_{\pi+\phi} \upsilon'$. So the fact that frames σ and υ were in a certain ordering relation means that the next frames in those films are in the corresponding ordering relation after update. A more sophisticated model would perhaps allow juggling of orderings in the update process, to allow for what Grice-style conversational analysis might tell us about the author's knowledge and intentions.

Before considering how we might make use of plausibility orderings, let us see how the earlier notion of entailment could be defined in terms of them. The *domain* of an ordering π, written $\star\pi$, can be defined as the set of contexts which are at least as plausible as themselves in the ordering, and this allows retrieval from a plausibility ordering of a corresponding information set. This in turn permits the definition of a notion of satisfaction of a formula by a plausibility ordering in terms of the earlier notion of satisfaction by an information set:

Definition D8 **(Domain of an ordering and 'Ordinary' Satisfaction w.r.t. an ordering)**

$$\begin{array}{rcl} \star\pi & = & \{\sigma \mid \sigma \geq_\pi \sigma\} \\ \pi \models \phi & \textit{iff} & \star\pi \models \phi \end{array}$$

Given such a notion of satisfaction, it should be clear that it would be straightforward to define a notion of entailment equivalent to that given earlier. However, it is also possible to define alternative notions of entailment relative to any given plausibility ordering. Let us say that the set of *preferred contexts* in an ordering π, written $\uparrow\pi$, is the set of all contexts which are at least as plausible as any context in the ordering. Then we can say that an ordering π *preferentially satisfies* a formula ϕ, written $\pi \rhd \phi$, if the set of preferred contexts in π satisfies ϕ. Preferential satisfaction is a weaker notion than satisfaction, in that an ordering may preferentially satisfy more formulae than it satisfies. We may now say that a formula ϕ preferentially entails a formula ψ relative to an ordering π, written $\phi \rhd_\pi \psi$, if updating π with ϕ produces an ordering which preferentially satisfies ψ. Here are the formal definitions:

Definition D9 **(Preferential Satisfaction and Entailment)**

$$
\begin{array}{rcl}
\uparrow \pi & = & \{\sigma \mid \forall \tau \in \star\pi \ \sigma \geq \tau\} \\
\pi \rhd \phi & \textit{iff} & \uparrow \pi \models \phi \\
\phi \rhd_\pi \psi & \textit{iff} & \pi + \phi \rhd \psi
\end{array}
$$

In the case of a trivial ordering consisting of the cross-product of the powerset of worlds $\pi_0 = \mathcal{P} \times \mathcal{P}$, for which every set of worlds is at least as plausible as every other set of worlds, this notion collapses into the earlier entailment:

Fact F3

$$\phi \models \psi \quad \textit{iff} \quad \phi \rhd_{\pi_0} \psi$$

Consider the following pair of sentences:

1 If Jane takes a bath, Bill will be annoyed that there is no more hot water.

2 If Jane wants a bath, Bill will be annoyed that there is no more hot water.

An utterance of (1) does not suggest to me that there actually is no more hot water, but only that if Jane takes a bath, there will be no more hot water. On the other hand, (2) suggests strongly that there is no more hot water. Put another way, (1) is compatible with the standard CCP prediction of a conditional reading, but (2) is not.

The current theory will predict the contrast provided the following plausibility assumptions hold:

- At least one alternative in which it is established that there is no hot water is more plausible than all alternatives in which it is not known whether there is hot water, but in which it is known that if Jane wants a bath then there will be no hot water.
- A alternative in which it is not known whether or not there is hot water but in which it is established that if Jane has a bath then there will be no more hot water must be at least as plausible as all alternatives where it is definitely established that there is no hot water.

The general question is, why would it be reasonable to expect plausibility orderings to have such properties? My answer to this question is on the one hand both simple and obvious, and on the other hand both awkward to implement and incompatible with many contemporary theories of presupposition. Many linguists will surely find it unpalatable. The answer is: common sense.

Let me expand on this. The contrast between (1) and (2) results from our ability to find a common-sensical explanation of the lack of hot water in terms of somebody having taken a bath, but in our inability to fully explain a lack of hot water in terms of somebody simply wanting a bath. The simple assumption that there is a finite amount of relevant hot water — about a bathful — is sufficient to allow justification of there being no more hot water in situations where Jane has just taken a bath. However, the same simple assumption would not suffice in the case of (2), and a number of other assumptions would be needed, such as the assumption that if Jane wants a bath then she will definitely take one. Thus it is the relative plausibility of assumptions not explicitly mentioned in the text of the example sentences that determines what is implicated.

Let us see how some of this analysis of (1) and (2) may be crudely formalised. For expositional purposes, I will ignore many obviously relevant issues, such as temporal connections between antecedent and consequent clauses in the conditionals. Let us represent "Jane takes a bath" as "JTB", "Jane wants a bath" as JWB", "there is no hot water" as NHW, and "Bill will be annoyed that there is no more hot water" as ∂NHW $\wedge$ BA. Now suppose that our common sense knowledge of the relative plausibility of different assumptions is encoded in a plausibility ordering π. The two conditions required of π may be formalised as follows:

- $\exists\sigma \in \pi\ \sigma \models \text{NHW}$ and
  $\forall\tau \in \pi$ if $(\tau \not\models \text{NHW}$ and $\tau \models\ \text{JWB} \rightarrow \text{NHW})$ then $\sigma >_\pi \tau$
- $\exists\sigma \in \pi\ \sigma \not\models \text{NHW}$ and $\sigma \models\ \text{JTB} \rightarrow \text{NHW}$ and
  $\forall\tau \in \pi$ if $\tau \models \text{NHW}$ then $\sigma \geq_\pi \tau$

If π conforms to these requirements, then we have the following preferential entailments:

$$\begin{array}{lcl} \text{JWB} \rightarrow (\partial\text{NHW} \ \wedge\ \text{BA}) & \rhd_\pi & \text{NHW} \\ \text{JTB} \rightarrow (\partial\text{NHW} \ \wedge\ \text{BA}) & \not\rhd_\pi & \text{NHW} \\ \text{JTB} \rightarrow (\partial\text{NHW} \ \wedge\ \text{BA}) & \rhd_\pi & \text{JTB} \rightarrow \text{NHW} \end{array}$$

In other words, with respect to π, (1) preferentially entails that there is no more hot water, and (2) preferentially entails not there is no more hot water, but that if Jane takes a bath then there is no more hot water.

5 Discussion

The theory developed in this paper can be thought of as an attempt at a formal characterisation of what Lewis [Lew79] called *accommodation*, most authors since having followed his terminology, with the exception of Seuren [Seu85], who refers to an apparently similar process which he terms *backward suppletion*. But in its current form the model differs markedly from existing proposals, in that most writers have taken accommodation to be some sort of repair strategy, something that happens when the interpretation process goes wrong. Lewis seems to picture accommodation as a covert adjustment of what he calls the *conversational score*, a sort of creative accounting needed to make conversational ends meet. Van der Sandt's accommodation, to take a more recent example, is a sophisticated cut-and-paste operation on discourse representation structures. Thus accommodation has been viewed as an essentially non-monotonic operation, overwriting our previous record of what had happened in a discourse to fit with new demands. The view espoused in this paper has been quite the contrary. Accommodation is seen as a *monotonic* operation, in the sense that it does not replace or destructively revise our information about a speaker or author, but further instantiates our knowledge, reducing the range of possibilities for what the speaker was assuming.

Yet this monotonicity is not crucial, it is merely symptomatic of the differences between the approach I have espoused and others. The main claim I wish to make is that when we accommodate, we look not only at the record of what has been said, but also behind what has been said, and consider explicitly what the author might have intended and what the author might have expected. Once we explicitly include in our model the reader's uncertainty about the authors assumptions, and if we place no restrictions on the reasoning power of the reader, as a side effect our model will tend to predict that the reader's information state develops monotonically.

So we see two extremes. Firstly there is the Lewis-ian model where the reader keeps a score of what has been said, but occasionally retrospectively alters that score to fit the facts. Secondly there is the model presented here, where the reader's energy is devoted to second-guessing the author's assumptions, and the reader's own score is thus a higher order entity recording all the possibilities for what the author takes the score to be. The reality must lie between these extremes: reasoning about the author is expensive from a processing point of view, so readers are forced to take short cuts. This means that readers cannot simultaneously consider all the alternative contexts the author might be assuming, and will occasionally be forced to back-track and search for alternative contexts that fit the facts better than those under consideration. Yet it remains the case that some sort of reasoning about the writer's assumptions is involved, and that a model of accommodation is unlikely to be successful unless it takes into account such reasoning.

I am very much in agreement with Richmond Thomason, who's notion of accommodation is yet more sophisticated than the one I have proposed here, since it takes into account not only the beliefs, but also the communicative goals of other agents: "Concentrating on accommodation means shifting to reconstructed reasoning that underlies utterances. And it suggests that certain reasoning processes, such as intention recognition and cooperation are central. Successful accommodation requires that we first recognize someone's intention to achieve a goal, and then establish goals of our own that will assist in achieving this goal."[Tho90] For the sake of concreteness, I have attempted to escape the need for this sophistication by restricting my attention to a very simple presuppositional language, and by concentrating on the understanding of texts, where the scope for development of cooperative plans is naturally limited.

The model I am proposing is certainly not the first theory of presupposition to take into account the beliefs of other agents. Indeed, in a certain sense Gazdar's theory [Gaz79] and some of its descendants (eg. [vdS88], [Mer92], [Blo93]) go further in this direction than I have. What I particularly like about these theories is their utilisation of epistemic logics allowing for explicit reasoning about agents beliefs. However, in all of thes theories the type of reasoning about what has been assumed by a speaker or writer is somewhat limited. Given that a presupposition trigger has been used, the only question is whether the triggered presupposition is or is not believed by the speaker. On the other hand in the current model a much richer connection between the triggered presuppositions and the assumptions of the speaker or writer is allowed.

I will now briefly discuss one point of confusion in the account I have offered. For (1), there is the possibility either of accommodating that *if Jane takes a bath there is no hot water* or accommodation that *there is no hot water* simpliciter. The weaker conditional appears to be preferred. On the other hand, the treatment of (2) seems to involve the reverse situation, with a logically stronger (ie. non-conditionalised) proposition being accommodated. How can it be that a sometimes a logically stronger proposition is considered more plausible than a logically weaker one, and sometimes *vice versa*[4]? This apparent enigma arises from a misidentification of the accommodation process with the addition of single propositions, and a misidentification of the ordering over epistemic alternatives as an ordering over propositions. The plausibility ordering should not, however, be thought of as an ordering over propositions, but over closed sets of propositions, ie. *theories*. Thus I am not committed to any claim about the conditional *if Jane takes a bath there is no hot water* being more plausible than the proposition *there is no hot water*, and I am not committed to *there is no hot water* being more plausible than *if Jane wants a bath there is no hot water*. Rather, I would claim that there must be a relevant closed theory containing the first conditional which is more plausibly taken to be the common ground than every relevant closed theory containing the simple proposition. Similarly, there

[4]This objection, that plausibility appears oddly disconnected from entailment, was first made to me by Henk Zeevat. The same objection can be found in [GeuMS].

must be a theory containing the simple proposition which is more plausible than every theory which contains the second conditional but not the simple proposition. There is no reason to assume that any relation of inclusion (the *theory* level counterpart of propositional entailment) holds between these various theories.

So contexts could profitably be thought of as logically closed theories, and what is accommodated does not hang on the relative plausibility of alternative propositions, but on the relative plausibility of alternative theories. This particular point would become much clearer if the mechanisms described in this paper were presented more proof theoretically. In Karttunen's earliest presentation of a context change model, that in [Kar74], contexts are defined as as sets of propositions. It would be natural to recast the current model in this way. A similarly constructive alternative would be to use DRS's as contexts, in which case one might end up with a picture close to that of van der Sandt. And rather than supposing that a reader simultaneously keeps in mind all the different possible contexts, we could model a readers information states in terms of a single context σ together with some representation of common-sense knowledge about the world. Then the update operation would become essentially abductive. Given that the speaker has uttered ϕ, the reader's task would be to establish the minimal most plausible set of assumptions the speaker could be making which contains σ and allows update with $\phi$to produce the new context. Such developments will have to await future research. But it can be seen that the formal theory presented in in this paper is a more model-theoretic counterpart of the sort of abductive picture of the interpretation process for which Jerry Hobbs (see e.g.[HSAM93]) has long been arguing.[5]

References

[Bea92] David Beaver, *The Kinematics of Presupposition* Proceedings of the Eighth Amsterdam Colloquium, ILLC, University of Amsterdam

[Bea93a] 'Kinematic Montague Grammar', in Kamp, H. (ed.), DYANA-2 deliverable R2.2a, University of Amsterdam

[Bea93b] Beaver, D., *What Comes First in Dynamic Semantics*, ILLC report LP-93-15, University of Amsterdam.

[Bea94] Beaver, D., "When Variables Don't Vary Enough", in Harvey, M. & L.Santelmann (eds.), *SALT 4*, Cornell

[5] One could also, I think, draw a profitable analogy between the treatment of presupposition in this paper and the common-sense driven analyses of discourse and temporal relations in the work of Asher, Lascarides and Oberlander (see e.g. [LA93]). One might take various temporal expressions (e.g. tensed verbs) as presupposing some temporal reference with accommodation then being necessary to build the most plausible link between previously introduced temporal discourse referents and the new reference point.

[Bea95] Beaver, D., *Presupposition and Assertion in Dynamic Semantics*, PhD Dissertation, University of Edinburgh.

[BeaMS] Beaver, D., "Presupposition", to appear in van Benthem, J. and A.ter Meulen (eds.), *The Handbook of Logic and Linguistics*, Elsevier

[Blo93] Blok, P., *The Interpretation of Focus: an epistemic approach tp pragmatics*, Doctoral Dissertation, university of Groningen

[Gaz79] Gazdar, G., *Pragmatics: Implicature, Presupposition and Logical Form*, Academic Press, New York

[GS91] Groenendijk, J. and M. Stokhof, *Dynamic Predicate Logic*, in Linguistics and Philosophy 14(1), 1991, 39–100.

[GeuMS] Geurts, B., *Presupposing*, Manuscript, University of Osnabrück

[Hei82] Heim, I.,*On the semantics of Definite and Indefinite Noun Phrases*, PhD dissertation, University of Amherst.

[Hei83] Heim, I., *On the Projection Problem for Presuppositions*, in *WCCFL '83*

[Hei90] Heim, I., 'Presupposition Projection', in *Reader for the Nijmegen Workshop on Presupposition, Lexical Meaning, and Discourse Processes*, University of Nijmegen

[HSAM93] Hobbs, J., M. Stickel, D. Appelt and P.Martin*Interpretation as Abduction*, Artificial Intelligence 63, 69–142

[Kam81] Kamp, H., 'A Theory of Truth and semantic Representation', in Groenendijk, Janssen & Stokhof (eds.) *Formal Methods in the Study of Language*

[Kar74] Karttunen, L.,'Presuppositions and Linguistic Context', Theoretical Linguistics 1

[LS71] Langendoen, D. and H. Savin, 'The Projection Problem for Presuppositions', in Fillmore & Langendoen (eds.), *Studies in Linguistic Semantics* Holt, Reinhardt & Winston, New York

[LA93] Lascarides, A., and N. Asher, *Temporal Interpretation, Discourse Relations and Commonsense Entailment*, Lingusistics and Philosophy 16:5, pp 437–495

[Lew79] Lewis, D., 'Scorekeeping in a Language Game' in *Bäuerle et al.(eds.) Semantics from Different Points of View*, Berlin

[Mer92] Mercer, R.,*Default Logic: Towards a Common Logical Semantics for Presupposition and Entailment*, Journal of Semantics 9:3

[vdS88] van der Sandt, R., *Context and Presupposition*, Croom Helm, London

[vdS92] van der Sandt, R., *Presupposition Projection as Anaphora Resolution*, Journal of Semantics 9:4

[Seu85] Seuren, P., *Discourse Semantics*, Basil Blackwell, Oxford

[Soa82] Soames, S., *How Presuppositions Are Inherited: a Solution to the Projection Problem*, Linguistic Inquiry 13:483–545

[Sta74] Stalnaker, R., 'Pragmatic Presuppositions', in Munitz & Unger, (eds.) *Semantics and Philosophy*, NYP

[Tho90] Thomason, R., "Accommodation, Meaning and Implicature: Interdisciplinary Foundations for Pragmatics", in Cohen, P., J. Morgan and E. Pollack (eds.) *Intensions in Communication*, Bradford Books, MIT Press, Cambridge, Mass, pp. 325–363.

[Vel91] Veltman, F., *Defaults in update semantics*, in DYANA deliverable R2.5.C, (also: to appear Journal of Philosophical Logic)

[Zee92] Zeevat, H., *Presupposition in Update Semantics*, Journal of Semantics 9:4