Plausibility Revision in Higher-Order Logic With an Application in Two-Dimensional Semantics

Erich H. Rast
New University of Lisbon

1 Introduction

In this article, a qualitative notion of subjective plausibility and its revision based on a preorder relation are implemented in higher-order logic. This notion of plausibility is used for modeling pragmatic aspects of communication on top of traditional two-dimensional semantic representations. First, some prerequisites will be laid out in section 2. Higher-order logic and applicative categorial grammar are introduced and one way of implementing two-dimensional semantics in this setting is laid out. In section 3 a notion of subjective plausibility and its revision is implemented. It is then shown in section 4 how this apparatus may be used for modeling interpretative assumptions, which provide the basis for the modeling of communication in a sender-receiver framework in which the semantic content of messages is semantically underdetermined.

2 Prerequisites

2.1 Higher-Order Logic

Typed higher-order logic (HOL) with generalized Henkin models is used in what follows. Since we have a particular application of higher-order logic to natural language semantics in mind not all technical details are discussed here and we refer to seminal work such as [BBSS08, Car97] for more information. The notation used here differs slightly from the one in [BBSS08] and will be explained below.

Syntax. Base types are e for objects, t for truth-values, and c as the intensional base type for contexts and circumstances of evaluation also known as modal indices. If α, β are types then $(\alpha\beta)$ is a type. Parentheses in type specifications may be left out and in this case right-associativity is assumed; so for example eet is a shortcut for $(e(et))$. Constant base terms are represented by alphanumeric sequences of letters and special symbols, whereas variables are represented by single letters such as x, y, z, P, Q, and R. Constants may also be represented by a single letter, in case of which this is mentioned explicitly.

Compound terms are built as follows. If A is a term of type $\beta\alpha$ and B is a term of type β, then (AB) is a term of type α. If A is a term of type α and x is a variable of type β then $(\lambda x A)$ is a term of type $\beta\alpha$. The identity symbol $=$ is of type $\alpha\alpha t$ for any type α. Types of terms are indicated as superscripts. Parentheses around terms may be left out and in case of doubt one might refer to the type of the respective terms in order to determine the correct bracketing. Infix notation and the usual logical symbols are used for common relations like identity and the quantifiers, i.e. we write $a = b$ instead of $((= a)b)$ and $\forall x A$ instead of $(\forall(\lambda x A))$. Multiple λ-abstractions are abbreviated and dot-notation is used, i.e. we write $\lambda xy.A$ instead of $(\lambda x(\lambda y A))$. Parentheses are often left out but we stick to the notation (AB) of λ-calculus instead of operator-argument notation $A(B)$. Infix notation is also used for the usual truth-functions like $\wedge$, $\rightarrow$, and $\vee$. Notice that in [Chu08] and texts following this tradition types are read in the opposite direction than here: $(et)t$ in the our notation corresponds to $o(o\iota)$ in Church's notation.

The following variables including their indexed variants will be used: s, t, u, v, w are of type c; x, y, z of type e; i, j of type cce; C, C' of type $eccct$; P, Q of type cct; and p, q of type ct. Parentheses and types are left out when they can be inferred from the context.

Semantics. Following the seminal [Hen08] it is stipulated that every term has an interpretation but terms of compound type $(\alpha\beta)$ are interpreted only over a subset of the set of all functions from D_α to D_β. Thus, a generalized Henkin model for higher-order logic with identity consists of a collection of domains D_α for terms of primitive type α, where $D_t = \{T, F\}$, domains $D_{(\alpha\beta)} \subseteq D_\beta^{D_\alpha}$ for terms of compound type $(\alpha\beta)$, and an evaluation function $[\![.]\!]^g$ from terms of type γ to their denotation in D_γ under an assignment g. If A is a variable of type γ then $[\![A]\!]^g = g(A)$, where $g(A) \in D_\gamma$. If A is a constant term of primitive type γ then $[\![A]\!]^g \in D_\gamma$. If A is of the form $(A^{\beta\alpha} B^\beta)$ then $[\![(AB)]\!]^g = [\![A]\!]^g\,([\![B]\!]^g)$, i.e. it is interpreted as functional application. The term $(\lambda x^\beta A^\alpha)$, where A is a term of type α, is interpreted as the function $f \in D_\alpha^{D_\beta}$ such that $[\![A^\alpha]\!]^{g[x/a]} = f(a)$ for any $a \in D_\beta$, where $g[x/a]$ is the same variable assignment as g except that $g(x) = a$. As one might expect $=$ is interpreted as identity, i.e. $[\![((=^{\alpha\alpha t} A^\alpha) B^\alpha)]\!]^g = 1$ if $[\![A^\alpha]\!]^g = [\![B^\alpha]\!]^g$ and 0 otherwise. $[\![(\wedge A^t) B^t]\!]^g = 1$ if $[\![A]\!]^g = 1$ and $[\![B]\!]^g = 1$ (0 otherwise), and $[\![\neg^{tt} A^t]\!]^g = 1$ if $[\![A]\!]^g = 0$ and 0 otherwise. The universal quantifier $\forall(\lambda x A)$ may be defined as $(\lambda x A) = (\lambda x \top)$, where $\top$ is defined as $((=^{(ttt)(ttt)t} =^{ttt})\ =^{ttt})$ like in [And08, 85]. The other logical connectives and the existential quantifier are defined as usual.

We will also make use of a description operator ι^α for any type α except t. Let us define, as usual, $\exists! x^\alpha B^{\alpha t}$ by the scheme $\lambda A^{\alpha t} \exists y^\alpha [Ay \wedge \forall z^\alpha (Az \rightarrow y = z)](\lambda x.B)$. Based on this definition the axiom of descriptions $\exists! x^\alpha [P^{\alpha t} x \rightarrow P(\iota P)]$ (for any P) is often taken for dealing with the iota operator in an axiomatic setting. However, we also wish to give the iota operator an explicit denotational semantics without having to deal with partial functions or using a 3-valued logic. For this purpose a common trick shall be used like in [Kap89].

Let † be an alien element $\dagger \in D_\alpha$ for any type α except t. As a model constraint, let further $[\![(A^{\alpha t})]\!]^g (\dagger) = F$. Given that, let $[\![\iota^\alpha A^{\alpha t}]\!]^g$ denote the unique $a \in D^\alpha$ such that $[\![A]\!]^g (a) = T$ if there is one, and † otherwise. As an alternative we could have used a partial type theory based on a 3-valued Kleene system or a 4-valued Belnap-style logic as in [Mus95]. However, the resulting multiplicity of connectives and the question how to interpret them would have lead astray from the main topic of this paper of how to integrate a simple revision operation with a two-dimensional semantics. Moreover, even a simple preorder revision operation might not be directly transferable from classical to a many-valued logic, as it is for obvious reasons very likely that additional case distinctions will be required in the many-valued setting in comparison to the classical one.

λ-Calculus. The following, well-known rules of λ-calculus will often be used in examples:

$$(\lambda x A) \overset{\alpha}{\Leftrightarrow} (\lambda y A[x/y]) \qquad \alpha\text{-conversion} \qquad (1)$$

$$((\lambda x A)B) \overset{\beta}{\Leftrightarrow} A[x/B] \qquad \beta\text{-conversion} \qquad (2)$$

$$(\lambda x A) \overset{\eta}{\Leftrightarrow} A \qquad \eta\text{-conversion} \qquad (3)$$

where x may not be free in A for η-conversion, x and B must be of the same type, and $A[x/y]$ is the same term as A except that all free occurrences of x in A have been substituted by y. Applying one of the rules as a rewrite rule from left to right is called a reduction. Particularly β-reduction will often be used in the two-dimensional semantics laid out further below.

2.2 Applicative Categorial Grammar

For giving examples a categorial grammar in the tradition of [Ajd35, Mon74, BH64, Lew70] is briefly introduced in this section. In addition to the semantic types e, t, and c syntactic categories such as n, np, and s will be used. While having syntactic categories is not strictly speaking necessary and instead additional semantic types like $e//t$, $e///t$, and so forth, could be used as in [Mon74], syntactic categories make examples more readable. If σ and τ are syntactic categories, (σ/τ) and $(\sigma\backslash\tau)$ are also syntactic categories. We write $\sigma : A^\alpha$ for a term A of semantic type α and syntactic category σ. Syntactic construction works in parallel to semantic composition and is specified by the following rules:

$$(\sigma/\tau) : A^{(\beta\alpha)}\ \tau : B^\beta \overset{f}{\Rightarrow} \sigma : (A\, B) \qquad \text{forward concatenation} \qquad (4)$$

$$\tau : B^\beta\ (\tau\backslash\sigma) : A^{(\beta\alpha)} \overset{b}{\Rightarrow} \sigma : (A\, B) \qquad \text{backward concatenation} \qquad (5)$$

Notice that unlike in other formulations of categorial grammar according to the above rules semantic representations are composed in parallel to syntactic construction but these representations are not interpreted. There is no need to consider examples up to the level of the interpretation of nonlogical constants in a model. In the present setting, semantic representations are normalized by the rules of λ-calculus and constants with mnemonic names will serve as translations of corresponding natural language expressions without actually interpreting the higher-order formulas. A small, purely extensional example shall clarify this approach.

Example 1. *The lexicon entries for 'John' and 'Mary' are* $np : j^e$ *and* $np : m^e$ *respectively and the lexicon entry for 'loves' is* $(np\backslash s)/np : \lambda yx.love^{eet}\, x\, y$*. then:*

$$\text{'John loves Mary'} \tag{6}$$
$$= np : j\ (np\backslash s)/np : \lambda yx.love\, x\, y\ np : m \tag{7}$$
$$\overset{f}{\Rightarrow} np : j\ np\backslash s : ((\lambda yx.love\, x\, y)m) \tag{8}$$
$$\overset{\beta}{\Rightarrow} np : j\ np\backslash s : \lambda x.love\, x\, m \tag{9}$$
$$\overset{b}{\Rightarrow} s : ((\lambda x.love\, x\, m)j) \tag{10}$$
$$\overset{\beta}{\Rightarrow} s : love\, j\, m \tag{11}$$

In subsequent examples many of the intermediate steps will be left out for brevity, as the reader will without doubt already be familiar with such uses of applicative categorial grammar and λ-calculus. As is well known rules (4) and (5) only specify the applicative part of categorial grammar as opposed to the full Lambek calculus for string concatenation that has been investigated by [Lam58]. In contrast to the applicative fragment defined above the proof-theoretic approach of the full Lambek calculus also allows for hypothetical reasoning. Using only the above rules it is for example not possible to assign the category s/np to the non-constituent phrase *John likes* given that *likes* is a transitive verb of the category $(np\backslash s)/np$, whereas *likes Maria* would correctly be assigned the category $np\backslash s$ (see [Car97, ch. 5]). For present purposes the applicative fragment of categorial grammar will suffice, since larger fragments or complex syntactic phenomena are not analyzed in this paper. Comprehensive surveys of full categorial grammar and its extensions can be found in [Moo97, Car97, Mor94, Mor10].

2.3 Two-dimensional Semantics

As the contributions to [GCM06] illustrate, two-dimensional semantics isn't really a uniform semantic framework; it comes in a variety of strands. In what follows, two-dimensional semantics is understood as an implementation of sentence level semantics that loosely follows [Kap89] and has been laid out in [Ras07, ch. 5-6]. According to this position, overt indexical expressions and tenses are regarded as expressions that semantically depend on features of a context, which are encoded by corresponding context parameters. In contrast to this, the circumstances of evaluation are indices over which traditional modal operators quantify implicitly. Consequently, meanings are represented as functions from contexts to functions from circumstances of evaluation (CEs) to extensions. In the present higher-order setting this means that meanings are generally of the form $\lambda us.a$, where u is a variable for a context of type c, s is another variable of type c used for traditional modal operators, and a is a term. Indexicals will then semantically depend on u, whereas ordinary nonindexical expressions will generally depend on s. The following example illustrates this approach:

Example 2. *Let the lexicon entries for 'I', 'me' be* $np : \lambda us.speaker\, u$*, for 'Mary' be* $np : \lambda us.m$*, and for 'loves' be* $(s\backslash np)/np : \lambda ijus.love\, s\, (jus)\, (ius)$*, where speaker is of type ce and love is of type ceet. then:*

$$① \text{ 'Mary loves me'} \tag{12}$$
$$= np : \lambda us.m\ (np\backslash s)/np : \lambda ijus.love\, s\, (jus)\, (ius)\ np : \lambda us.speaker\, u \tag{13}$$
$$\Rightarrow np : \lambda us.m\ np\backslash s : \lambda jus.love\, s\, (jus)\, (speaker\, u) \tag{14}$$
$$\Rightarrow s : \lambda us.love\, s\, m\, (speaker\, u) \tag{15}$$

There are a number of crucial differences between this implementation and the one by [Kap89]. First, using the same sort of entities of type c for both circumstances of evaluation and contexts is not uncommon – see for example [vSZ05] – but not in the spirit of Kaplan's original approach, as he emphasizes the conceptual distinction between contexts and circumstances of evaluation. Contexts are more fine grained than circumstances of evaluation, as long as the latter only represent what ordinary modal operators quantify over. Concededly this distinction has been watered down somehow by relativists like [Mac05, Mac07], who have proposed that circumstances of evaluations should encode more than just time intervals and possible worlds. Second, in the above example all expressions have the type $cc\alpha$ for some base type α even if they don't depend on u or s. It would be possible to simplify these examples by only using intensional types for terms that semantically depend on a context or CE respectively. This requires introducing some additional extensional and intensional evaluation rules as in [vSZ05] or making the dependences on u and s implicit. For simplicity we have chosen the more general setting with fully intensional types for all terms, as these only result in a few additional β-reductions for nonindexical expressions. The entry for the proper name 'Mary' in the above example illustrates these harmless additional reductions.

Here is a slightly more complex example of how two-dimensional semantics works in the present view.

Example 3. *Suppose that dayOf is a function of type ce that yields the day of a context/CE and assume that basic arithmetic operations are available. Let further placeOf be a function of type ccc that yields the same index as its second argument except that the place of that index is changed to that of the first argument and let* $s \prec u$ *be true iff.* u *temporally precedes* s *according to a temporal ordering relation on* d_c*. let the lexicon entry for 'was' be* $(np\backslash s)/(s\backslash s) : \lambda A^{(cct)(cct)}ius.(s \prec u \wedge a\ (locatedAt\ s\, (ius))$*, 'here' stand for* $s\backslash s : \lambda P.Pu\, (placeOf\ us)$*, and 'yesterday' stand for* $s\backslash s : \lambda P^{cct}us.(dayOf\ s) =$

$(dayOf\ u) - 1 \wedge Pus$. *Then:*

$$\text{'I was here yesterday'} \tag{16}$$

$$= np : \lambda us.speaker\, u\ (np\backslash s)/(s\backslash s) : \lambda A^{(cct)(cct)} ius.(s \prec u \tag{17}$$
$$\wedge\ A(locatedAt\, s\, (ius))\ s\backslash s : \lambda P.Pu\, (placeOf\ us)$$
$$s\backslash s : \lambda P^{cct} us.dayOf(s) = dayOf(u) - 1 \wedge Pus$$

$$\overset{f}{\Rightarrow} np : \lambda us.speaker\, u\ np\backslash s : \lambda ius.s \prec u \tag{18}$$
$$\wedge\ (locatedAt\ (placeOf\ us)(iu(placeOf\ \ us)))$$
$$s\backslash s : \lambda P^{cct} us.dayOf(s) = dayOf(u) - 1 \wedge Pus$$

$$\overset{b}{\Rightarrow} s : \lambda us.s \prec u \wedge (locatedAt\ (placeOf\ us)\ (speaker\ u)) \tag{19}$$
$$s\backslash s : \lambda P^{cct} us.(dayOf\ s) = (dayOf\ u) - 1 \wedge Pus$$

$$\overset{b}{\Rightarrow} s : \lambda us.(dayOf\ s) = (dayOf\ u) - 1 \tag{20}$$
$$\wedge\ s \prec u \wedge (locatedAt\ (placeOf\ us)(speaker\ u))$$

This example illustrates that two-dimensional semantics is mainly concerned with the semantically-encoded relations between u and s, where the former can be understood as a variable for a semantic entity that stands for an utterance situation and the latter as a variable for the situation described by a sentence. Notice that situations in the strict sense have to be non-maximal truth-makers and for this a partial logic like the one in [Mus95] is needed, so things are simplified a bit in the present account.

The above example also illustrates a problem mentioned in [Ras09] that – according to the author's opinion – as of the time of this writing has not yet been solved in the assorted literature on context dependence in a descriptively adequate way. Two-dimensional semantics cannot reasonably be required to put restrictions on the boundaries of time intervals or places that go beyond what is expressed by generic constraints between u and s, because these restrictions are not sufficiently constrained by lexical meaning. In other words, as [Bac05] has noted as one of the first, indexicals are like many other expressions in natural languages often semantically underdetermined. This is one of the reasons why a notion of interpretation is needed for a realistic modeling of an agent's understanding of sentences containing indexicals even when very rich and detailed lexicon entries for indexicals and seemingly similar expressions are assumed.

3 Adding Structure

Rational agents, including humans insofar as they are rational, are generally capable of comparing different scenarios according to their plausibility. In this section semantic entities are preordered in order to describe the plausibility that an agent associates with a context or CE. Plausibility is hereby understood in a similar sense as what [LvdT08] call normality.

3.1 Conditions for Binary Relations

The following conditions on binary relations of type *cct* will be used:

$$\text{Trans} := \lambda P.\forall stu[(Pst \wedge Ptu) \rightarrow Psu] \tag{21}$$

$$\text{Eucl} := \lambda P.\forall stu[(Pst \wedge Psu) \rightarrow Ptu] \tag{22}$$

$$\text{Ser} := \lambda P.\forall s \exists t[Pst] \tag{23}$$

$$\text{Refl} := \lambda P.\forall s[Pss] \tag{24}$$

3.2 Adding a KD45 Modality

Let R be a relation of type *ecct*, whose first argument is the agent. To characterize this relation model constraints are added that are familiar as frame conditions from modal logic. Let models be restricted to those in which $\text{Trans}(Rx)$, $\text{Eucl}(Rx)$ and $\text{Ser}(Rx)$ holds for any x. This makes R a KD45 modality. However, in the present approach states accessible via this relation do not represent belief tout court but only what an agent considers possible or relevant (short: considers) at a given moment. This may include states that the agent considers relatively unlikely. While one might argue that what an agent considers need not necessarily be represented as a KD45 modality, it seems that roughly the same arguments as those in favor of representing rational belief by a KD45 modality also speak for taking what an agent considers as a KD45 modality. What an agent considers must be understood as a weak form of belief for which positive and negative introspection ought to hold. States that are not accessible via R for an agent from a given base state are not taken into account at all by the agent. The idea is here that although the preorder that will be introduced in the next section is total an agent's beliefs at a given state only concern what he considers possible or relevant in that state. By 'state' any entity of type c is meant here and in what follows, no matter whether it stands for or is accessible from a context or some circumstances of evaluation. With the introduction of modalities states may play four distinct roles: (i) encoding relevant features of utterance situations, (ii) representing logically possible alternatives (e.g. when alethic modalities are introduced, which hasn't been done here), (iii) representing time intervals that are temporally ordered , and (iv) representing doxastic alternatives. Although it is not uncommon to conflate these aspects of semantic representation it deserves mentioning that this is not uncontroversial. For example, epistemic two-dimensionalists like [Cha04] reject equating (ii) with (iii) and it has already been mentioned that [Kap89] does not endorse using the very same entities for (i) and (ii). The ordering by plausibility to be introduced in the next section can be taken as a refinement of their role (iv) as doxastic alternatives.

3.3 Plausibility

Belief is implemented on top of R by means of a preorder relation that represents plausibility, or as [LvdT08] call it, normality. What an agent considers plausible might change from one state to another and so in our framework this relation depends on an agent and an additional state, i.e. it is of type *eccct*. A preorder is reflexive and transitive. Models therefore need to be restricted to those in which Trans($\geq xs$) and Refl($\geq xs$) holds for any agent x and base state s. We write $s \geq_{a,u} t$ for $(\geq aust)$, use $>$ for the strict part $\lambda xust.s \geq_{x,u} t \wedge \neg(t \geq_{x,u} s)$ and $\sim$ for the equivalence relation $\lambda xust.s \geq_{x,u} t \wedge t \geq_{x,u} s$. In order for the ordering to be a useful representation of plausibility the existence of a maximum must be ensured. This is achieved by the following condition:

$$\forall xup[\exists v(pv) \rightarrow \exists s(ps \wedge \neg\exists t[pt \wedge t >_{x,u} s])] \tag{25}$$

The purpose of this condition is not hard to see. Any non-empty p of type *ct* must contain some maximum elements with respect to the strict ordering $>$ generated from $\geq$, where 'non-empty', of course, in this case means that p is true at some state. When there are several $>$-maximal elements with respect to p then they are equally plausible to the given agent in the given base state. A function Max then determines the maximum q of a plausibility ordering C with respect to an agent, a base state and an intension p of type *ct*.

$$\text{Max} := \lambda xuCp.\iota q\forall s[(ps \wedge \neg\exists t[pt \wedge Cxuts \wedge \neg Cxust]) \equiv qs] \tag{26}$$

Because of condition (25) this function does not fail to denote a unique q with respect to $\geq$, any non-empty p and some x, u. What has to be shown in the next section is that it also determines a unique maximum for a revised plausibility relation.

3.4 Preorder Revision

Generally speaking, to revise a preorder by a proposition P all the relevant P-states need to be shifted above the relevant not-P states such that afterwards P-states are strictly preferred over not-P states. In the present two-dimensional setting this operation is a bit more complicated and its implementation depends on the intended interpretation of the u and s variables. first, we only want the preorder to be changed with respect to what an agent considers at a given state. Thus, only the states that are reachable by the agent's KD45 modality are taken into consideration, whereas the preorder is not changed with respect to any other states. Secondly, the agent's plausibility ordering is not changed with respect to the utterance situation, because in this case this state is taken as the one in which the change of preferences takes place.

As an auxiliary notion, let 'when A^t then B_1^t otherwise B_2^t' abbreviate $(A \rightarrow B_1) \wedge (\neg A \rightarrow B_2)$. The revision C' of an ordering relation C conditional on P for some agent x at u_0 is then characterized the following term.

$$\begin{aligned}\text{REV} := \lambda xu_0PC.\iota C'\forall yust[&\text{when}\, u_0 = u \wedge x = y \wedge Pus \wedge \neg Put \\ &\wedge Rxus \wedge Rxut \text{ then } (C'xust \wedge \neg C'xuts) \text{ otherwise } (C'yust \equiv Cyust) \\ &\wedge (C'yuts \equiv Cyuts)]\end{aligned} \tag{27}$$

Revision preserves reflexivity and transitivity of C, i.e. if Refl(Cxu) then Refl(REV $xuPC$) and if Trans(Cxu) then Trans(REV $xuPC$) for any x, u, P, and C. Proof (cf. [vBL05, 10]): According to (27) $C'xu$ only differs from Cxu for given x, u if the antecedent condition of (27) is true. (a) Reflexivity: The antecedent condition in (27) cannot be true if $s = t$ since $Pus \wedge \neg Pus$ is not satisfiable. Thus, according to the 'otherwise' clause $C'xuss \equiv Cxuss$ for any x, u, s. (b) Transitivity: If $Cxust$ and $Cxutv$ then $Cxusv$ by the transitivity condition for C. Assume that $C'xust$ and $C'xutv$ but not $C'xusv$ for some arbitrary constants s, t, and v. According to (27) and given that $Cxusv$, $\neg C'xusv$ can only be the case if $Puv \wedge \neg Pus$. Case 1: Suppose Put holds. Then according to the antecedent condition of (27) $C'xuts \wedge \neg C'xust$ would hold, contradicting the assumption $C'xust$. Case 2: Suppose $\neg Put$ holds. Then according to the antecedent condition in (27) $C'xuvt \wedge \neg C'xutv$ would hold, contradicting the assumption $C'xutv$. QED. Revision also preserves the existence of a maximum, i.e. a property analogous to (25) also holds for revised plausibility. Proof: Suppose $\geq$ is revised by P for x, u. Assume Pus and $\neg Put$ for some particular states s, t to the effect that according to (27) s is strictly preferred over t by some x in some u after revision. Assume there is a q^{ct} such that t is in its maximum with respect to $\geq, x$, and u. Case 1: qs holds; then s is in the new maximum of q after revision, since it is now strictly preferred over t and nothing else has changed. Hence, (25) is preserved. Case 2: qs doesn't hold. Then the succedent of (25) is not violated by s because s is not in q and t therefore remains in the maximum of q for x, u. QED.

To give an example of the representation in general, let us write 111 for a state s for which some P_1us, P_2us, and P_3us are true, 010 for an s for which P_1us is false, P_2us is true, and P_3us is false, 110 for an s for which P_1us is false, P_2us is true, and P_3us is false, and so forth. Suppose the agent initially considers $011 \sim 010 > 001 \sim 000 > 111 \sim 110 > 101 \sim 100$ in some context u, i.e. he prefers $\lambda s.\neg P_1us$ states over P_1u states, P_2u over $\lambda s.\neg P_2us$ states and is indifferent about P_3u states. After revision by P_1u he considers $111 \sim 110 > 101 \sim 100 > 011 \sim 010 > 001 \sim 000$. So after revision 111 and 110 are equally plausible to the agent and both of them are more plausible than any other state he takes into consideration. Figure 1 illustrates this simple re-ordering of states. Except for the two-dimensional setting this is a standard representation of qualitative preferences and their soft upgrade (see [vBL05, 9], [Liu08, 23]).

To express a real change it would, of course, also be possible to formulate a temporal revision operation as a relation between two context variables u_0 and u_1 such that $u_0 \prec u_1$ and the agent has revised preferences at u_1. For the current purpose an atemporal conditionalization operation will suffice, though.

4 Applications

The addition of a simple preorder allows for distinctions that in a more traditional setting are not available. Great care must be taken not to mix up two different conceptual issues, though. On one hand higher-order logic is custom-

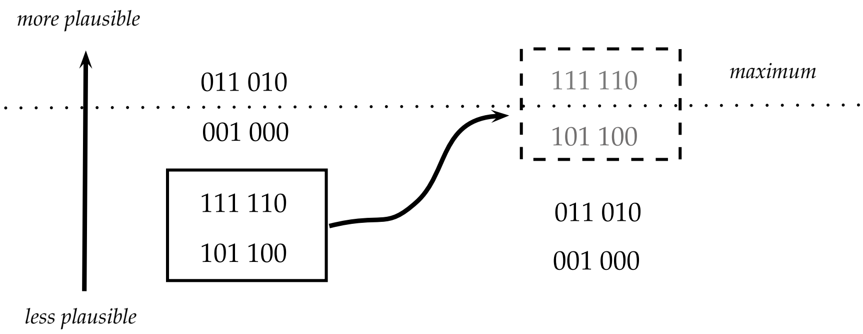

Figure 1: Plausibility revision by pushing p-worlds on top.

arily used for describing the meaning of natural language sentences. On the other hand it is also possible to express aspects of ideally rational interpretation within such a framework which do not represent the meaning of a natural language expression in a straightforward and compositional way. This distinction is addressed in the following paragraphs, before a closer look is taken at applications of preorder revision.

4.1 Interpretative vs. Linguistic Belief

By a *linguistic notion* we understand one that encodes the truth-conditions of a natural language sentence. For example, the following definition can be used to represent the ascription that x believes P in u_0, s_0, where indexicals are not interpreted.

$$\text{Bel} := \lambda x u_0 s_0 C P. \forall s_1 [(\text{Max}\, x s_0 C (R x s_0)) s_1 \to P u_0 s_1] \tag{28}$$

Here is an example of this notion in use:

Example 4. *Let 'Bob' be* $np : \lambda us.b$ *and 'believes' be* $np\backslash(s/s) : \lambda iPus.(\text{Bel}\,(ius)us \geq P)$. *Then:*

$$\text{'Bob believes Mary loves me'} \tag{29}$$

$$= np : \lambda us.b\; np\backslash(s/s) : \lambda iPus(\text{Bel}\,(ius)us \geq P) \text{ ①} \tag{30}$$

$$\overset{b}{\Rightarrow} s/s : \lambda Pus.(\text{Bel}\, bus \geq P) \text{ ①} \tag{31}$$

$$\overset{f}{\Rightarrow} s : \lambda us.\text{Bel}\, bus \geq (\lambda u's'.love\, s'\, m\, (speaker\, u')) \tag{32}$$

$$\overset{\alpha,\beta}{\Rightarrow} s : \lambda us.\forall s'[(\text{Max}\, bs \geq (Rbs))s' \to love\, s'\, m\, (speaker\, u)] \tag{33}$$

An *interpretative notion*, on the other hand, does not directly reflect part of the compositional meaning of a natural language sentence; it rather represents an aspect of the interpretation of that sentence in a given state by a given agent.

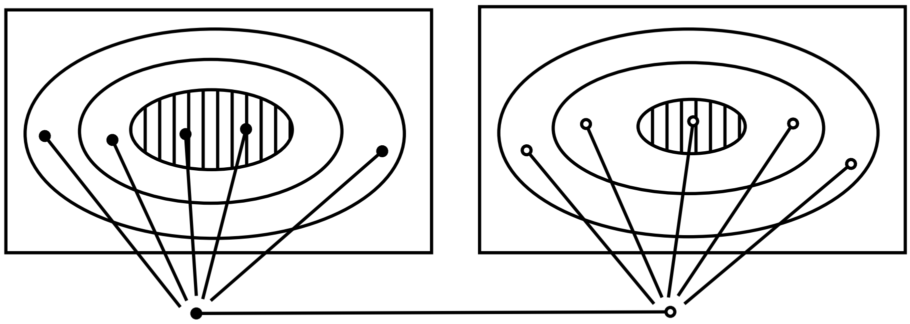

Figure 2: Example of the structure of an interpretative belief.

To the above notion of linguistic belief corresponds *interpretative belief* defined as follows.

$$\mathrm{IBL} := \lambda x u_0 s_0 C P. \forall u_1 s_1 [([\mathrm{Max}\, x u_0 C (R x u_0)] u_1 \wedge [\mathrm{Max}\, x s_0 C (R x s_0)] s_1) \rightarrow P u_1 s_1] \quad (34)$$

IBL represents x's interpretation of p based on what he considers most plausible and does not directly correspond to a natural language attitude ascription, because it also evaluates embedded indexicals according to what the agent believes (cf. [Ras07, 277-9], [Ras09, 524-5]). Figure 2 depicts the structure underlying interpretative belief. In a more conventional two-dimensional setting based on normal modal logic without a plausibility ordering a similar effect can be achieved by using a diagonalization operator in the scope of a belief modality. Assuming structural symmetry between contexts and indices as we do here, this operator can be defined syntactically in a double-index modal logic as $M, c, i \vDash \Delta\phi$ iff. $M, i, i \vDash \phi$. It 'de-rigidifies' indexicals that semantically depend on the context parameter c by evaluating them with respect to the index i. If a belief modality implicitly quantifies over this index, the diagonalized indexical is consequently evaluated with respect to this modality. Likewise, in our setting an interpretative notion takes into account those utterance contexts an agent currently considers most plausible and evaluates indexicals with respect to them. In contrast to this, context parameters are never shifted by the formal analogue to a natural language attitude ascriptions except for the modeling of so-called 'context-shifting indexicals' [Sch03].

There is a potential problem of interpretative belief when indexicals and demonstratives with a uniqueness condition such as 'I' or 'there' are interpreted but several contexts seem equally plausible to the agent in the given base state, as is for example depicted on the left side of figure 2. However, such cases are not more problematic than interpretations of nonindexical expressions with uniqueness condition (or, in fact, with any presupposition of number) such as definite descriptions. Moreover, from our pragmatic perspective violations of number presuppositions are less problematic than when they are viewn from a semantic perspective, because our representation merely reflects the fact that sometimes an agent might be in doubt about the most adequate interpretation of such an expression. By taking into account additional information and correspondingly revising her assumptions and beliefs an agent might later resolve the conflict and arrive at a unique interpretation.

4.2 Interpretative Assumptions and Their Revision

During communication discourse participants$_i$ maintain a model of what other discourse participants believe according to their$_i$ opinion. This model is to some extent specified by an agent's *interpretative assumptions*, which are defined as follows. Agent x's interpretative assumption that P in u_0, s_0 with respect to y is given by

$$\begin{aligned} \mathrm{IAS} := \lambda x y u_0 s_0 C P.\forall u_1 u_2 s_1 s_2[([\mathrm{Max}\, x u_0 C(R x u_0)]u_1 \\ \wedge [\mathrm{Max}\, y u_1 C(R y u_1)]u_2 \wedge [\mathrm{Max}\, x s_0 C(R x s_0)]s_1 \\ \wedge [\mathrm{Max}\, y s_1 C(R y s_1)]s_2) \rightarrow P u_2 s_2], \end{aligned} \tag{35}$$

or by the weaker term

$$\begin{aligned} \mathrm{IAW} := \lambda x y u_0 s_0 C P.\forall u_1 s_1 s_2[([\mathrm{Max}\, x u_0 C(R x u_0)]u_1 \\ \wedge [\mathrm{Max}\, x s_0 C(R x s_0)]s_1 \wedge [\mathrm{Max}\, y s_1 C(R y s_1)]s_2) \rightarrow P u_1 s_2]. \end{aligned} \tag{36}$$

According to the strong version the agent takes into account what he believes that the other agent believes about the utterance situation, whereas according to the weak version the agent only takes into account his own beliefs about the utterance situation when determining whether a possibly context-dependent proposition P is true according to his interpretative assumptions. There does not seem to be any clearcut criterion for deciding which of them is more adequate. Perhaps in real-world interpretation human agents consciously or unconsciously process several potential interpretations at once and both kinds of interpretative assumptions might play a role in this.

It should be mentioned at this place that it would be more adequate to model beliefs and assumptions with distinct modalities. Using one modality for both genuine beliefs and assumptions is a simplification that might need to be corrected in a more realistic approach. That being said, a first step when interpreting an utterance is to revise one's interpretative assumptions about the speaker by the meaning of that utterance as long as the speaker is considered honest and sincere. This revision can readily be expressed on the basis of the apparatus introduced so far. Revision on the basis of strong interpretative assumptions is represented by

$$\begin{aligned} \mathrm{RAS} := \lambda x y u_0 s_0 P.\iota Q \forall u_1 u_2 s_1 s_2[([\mathrm{Max}\, x u_0 \geq (R x u_0)]u_1 \\ \wedge [\mathrm{Max}\, x s_0 \geq (R x s_0)]s_1 \wedge [\mathrm{Max}\, y u_1 (\mathrm{REV}\, y u_1 P \geq)(R y u_1)]u_2 \\ \wedge [\mathrm{Max}\, x s_1 (\mathrm{REV}\, y s_1 P \geq)(R y s_1)]s_2) \equiv Q u_2 s_2], \end{aligned} \tag{37}$$

and the corresponding weak notion by

$$\begin{aligned}\mathrm{RAW} := \lambda xyu_0s_0P.\iota Q\forall u_1u_2s_1s_2[(&[\mathrm{Max}\,xu_0 \geq (Rxu_0)]u_1 \\ &\wedge [\mathrm{Max}\,xs_0 \geq (Rxs_0)]s_1 \wedge [\mathrm{Max}\,xs_1(\mathrm{REV}\,ys_1P \geq)(Rys_1)]s_2) \\ &\equiv Qu_1s_2].\end{aligned} \tag{38}$$

Each of these functions yields a revised content of type *cct* in u_0, s_0 given some P of the same type and two agents. As in the previous definitions the difference between the strong and the weak version lies in the way the context variable u_0 is handled. While the strong notion characterizes Q with respect to what the receiver believes that the speaker believes about both s_0 and u_0 conditional on P, the weaker notion only takes into account the receiver's interpretation of the context simpliciter.

Iterating the KD45 modality is not mandatory, i.e. (Ryu_1) and (Rys_1) could have been replaced by (Rxu_0) and (Rxs_0) respectively in (37) and (38), since the respective plausibility orderings for computing the maximums already depend on different agents and base situations. In our setting, if $Rxus$ holds then Rys represents what x considers in u to be considered by y just like in the case of iterated modalities in modal logic. So a partitioning like Rxu strictly speaking only represents what x considers in u without taking into account iterated considerations and not what x considers in u tout court.

4.3 A Limited Form of Interpretation

Suppose a speaker has uttered P and the receiver revises his own beliefs by the result of revising his interpretative assumptions about the speaker. This is a form of interpretation, albeit not a very sophisticated one, and might at first glance be represented as follows:

$$\mathrm{IPS} := \lambda xyusPQ.\mathrm{IBL}\,xus(\mathrm{REV}\,xu(\mathrm{RAS}\,xyusP) \geq)Q \tag{39}$$

In other words: After x has revised his (interpretative) beliefs by his interpretative assumptions regarding y, which have been revised by y's utterance P, x believes Q. IPS does for a variety of reasons not reflect interpretation in real-world communication situations, though. First, of course, people do not unconditionally accept what someone else has said even when it has been filtered through an interpretation process. Second, IAS, RAS, and IPS are too strong, because speakers interpret indexicals *as rigid expressions.* To put it in other terms, even though indexicals are interpreted by the receiver they are not generally interpreted according to the receiver's beliefs about what the sender believes. A more adequate notion of interpretation is therefore based on the weaker variants introduced above. The weaker variant IPW is defined like IPS except that RAS is exchanged with RAW:

$$\mathrm{IPW} := \lambda xyusPQ.\mathrm{IBL}\,xus(\mathrm{REV}\,xu(\mathrm{RAW}\,xyusP) \geq)Q \tag{40}$$

It expresses an agent x's interpretation of an agent y's utterance where indexicals are interpreted with respect to x's beliefs about the utterance context u_0. Notice, however, that an interpreter often needs to take both RAS and RAW into account. Suppose Bob utters the seemingly uninformative sentence 'I'm here' while being mistaken about his current location. Say, as far as Mary knows, he erroneously believes he is at the Continental, whereas he is in fact at the Grand Hotel and Mary also believes so. Then her linguistic and extralinguistic behavior is governed by both kinds of interpretative assumptions. For example, when talking to Bob on the phone, she might reply: 'No you're not there. You went to the Grand Hotel, not the Continental!' The following examples illustrates this case.

Example 5. *Let 'am' be* $(np\backslash s)/(s\backslash s) : \lambda A^{(cct)(cct)} ius.(A\ (locatedAt\ s\,(ius)))$*. Then:*

$$③\ \textit{'I am here'} \tag{41}$$

$$= np : \lambda us.speaker\, u\ (np\backslash s)/(s\backslash s) : \lambda A^{(cct)(cct)} ius.A(locatedAt\, s\,(ius)) \quad s\backslash s : \lambda P.Pu\,(placeOf\ us) \tag{42}$$

$$\xrightarrow{f} np : \lambda us.speaker\ np\backslash s : \lambda ius.locatedAt\,(placeOf\ us)(iu(placeOf\ us)) \tag{43}$$

$$\xrightarrow{b} s : \lambda us.locatedAt\,(placeOf\ us)\,(speaker\, u) \tag{44}$$

There are models for which $(\mathrm{RAS}\, mbu^* s^* \geq ③)$ *and* $(\mathrm{RAW}\, mbu^* s^* \geq ③)$ *are identical and ones for which they differ under the same intended interpretation of the constants* m, b, u^*, s^* *and other lexical entries. When they differ the receiver might point out that he suspects that the speaker has erroneous beliefs about the utterance context.*

5 Summary and Prospects

We have implemented plausibility and its revision within a two-dimensional semantics in higher-order logic and given some examples of interpretative notions that arise in such an approach. A long-term goal of this project is to 'logify' more aspects of interdependent, non-Gricean interpretation, which seems to be particularly promising for dealing with nonindexical context dependence whose resolution depends on background assumptions and underlying encyclopedic knowledge. Staying within a broadly-conceived montegovian framework as opposed to the more common modal logical approach allows one to integrate pragmatic notions with traditional semantics.

Two major issues have not been addressed: First, an agent sometimes takes new epistemic alternatives into account that were previously not taken into consideration. This boils down to revising R by P either by cutting off links to Rxu-reachable $\lambda s.\neg Pus$ states for given x, u, i.e. by a conjunctive condition on R, or by expanding R to reach the most plausible Pxu-worlds, i.e. by a disjunctive condition on R that makes use of $\geq$ and Max. While the latter form of revision in the style of [Gro88] can be implemented on the basis of a preorder, it is non-trivial to make it preserve properties of the accessibility relation such as Trans and Eucl. Second, a realistic account of interpretation needs to allow for a checking step by means of which the revised interpretative assumptions are compared with what the interpreter already believes. Given the limitations of a purely qualitative approach it is generally hard to find a checking operation that is suitable for modeling the interpretation and possible acceptance of natural language utterances and implement the corresponding non-prioritized belief revision. Both issues need to be addressed in future work.

References

[Ajd35] Kazimierz Ajdukiewicz. Die syntaktische Konnexität. *Studia Philosophica 1*, pages 1–27, 1935.

[And08] Peter B. Andrews. General models and choice in type theory. In Christoph Benzmüller, Chad E. Brown, Jörg Siekmann, and Richard Statman, editors, *Reasoning in Simple Type Theory*, pages 83–92. College Publications, London, 2008. First publ. in The Journal of Symbolic Logic Vol. 37, No. 2 (June 1972).

[Bac05] Kent Bach. Context ex machina. In Zoltán Gendler Szabó, editor, *Semantics versus Pragmatics*, pages 16–44. Oxford UP, Oxford, 2005.

[BBSS08] Christoph Benzmüller, Chad E. Brown, Jörg Siekmann, and Richard Stratman, editors. *Reasoning in Simple Type Theory*, number 17 in Studies in Logic, London, 2008. College Publications.

[BH64] Yehoshua Bar-Hillel. *Language and Information*. Addison-Wesley, Reading, MA, 1964.

[Car97] Bob Carpenter. *Type-Logical Semantics*. MIT Press, Cambridge, Massachusetts, 1997.

[Cha04] David Chalmers. Epistemic two-dimensional semantics. *Philosophical Studies*, 2004.

[Chu08] Alonzo Church. A formulation of the simple theory of types. In Christoph Benzmüller, Chad E. Brown, Jörg Siekmann, and Richard Statman, editors, *Reasoning in Simple Type Theory*, pages 35–47. College Publications, 2008.

[GCM06] Manuel García-Carpintero and Joseph Macià, editors. *Two-Dimensional Semantics*. Oxford UP, Oxford, New York, 2006.

[Gro88] Adam Grove. Two modellings for theory change. *Journal of Philosophical Logic*, 17(2):157–170, May 1988.

[Hen08] Leon Henkin. Completeness in the theory of types. In Christoph Benzmüller, Chad E. Brown, Jörg Siekmann, and Richard Statman, editors, *Reasoning in Simple Type Theory*, pages 49–59. College Publications, 2008.

[Kap89] David Kaplan. Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals. In Joseph Almog, John Perry, and Howard Wettstein, editors, *Themes from Kaplan*, pages 481–564. Oxford University Press, Oxford, New York, 1989.

[Lam58] Joachim Lambek. The Mathematics of Sentence Structure. *American Mathematical Monthly 65*, (65):154–170, 1958.

[Lew70] David Lewis. General semantics. *Synthese*, 22(1-2):18–67, 1970.

[Liu08] Fenrong Liu. *Changing for the Better*. Number DS-2008-02 in ILLC Dissertation Series. Institute for Logic, Language, and Computation, Amsterdam, 2008.

[LvdT08] Jérôme Lang and Leendert van der Torre. From belief change to preference change. In *The 18th European Conference on Artificial Intelligence ECAI 2008*, 2008.

[Mac05] John MacFarlane. Making sense of relative truth. In *Proceedings of the Aristotelian Society*, volume 105, pages 321–39, 2005.

[Mac07] John MacFarlane. Nonindexical contextualism. *Synthese*, 166(2):231–50, 2007.

[Mon74] Richard Montague. The Proper Treatment of Quantification in Ordinary English. In Richmond H. Thomason, editor, *Formal Philosophy: Selected Papers of Richard Montague*, pages 247–270. Yale University Press, London, 1974.

[Moo97] Michael Moortgat. Categorial type logics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 93–178. MIT Press, Cambridge, MA, 1997.

[Mor94] Glynn Morrill. *Type Logical Grammar: Categorial Logic of signs*. Kluwer Academic Publishers, Dordrecht, 1994.

[Mor10] Glyn V. Morrill. *Categorial Grammar: Logical Syntax, Semantics, and Processing*. Oxford UP, New York, Oxford, 2010.

[Mus95] Reinhard Muskens. *Meaning and Partiality*. CSLI Publications, Stanford, 1995.

[Ras07] Erich Herrmann Rast. *Reference and Indexicality*. Logos Verlag, Berlin, 2007.

[Ras09] Erich Rast. Context and interpretation. In Jesus M. Larrazabal and Larraitz Zubeldia, editors, *Meaning, Content, and Argument*, Proceedings of the ILCLI International Workshop on Semantics, Pragmatics, and Rhetoric, pages 515–534. University of the Basque Country Press, San Sebastian, 2009.

[Sch03] Philippe Schlenker. A Plea for Monsters. *Linguistics and Philosophy*, (26):29–120, 2003.

[vBL05] Johan van Benthem and Fenrong Liu. Dynamic logic of preference upgrade. ILLC Tech Report PP-2005-29, University of Amsterdam, Institute for Logic, Language & Computation, 2005.

[vSZ05] Arnim von Stechow and Thomas Ede Zimmermann. A problem for a compositional treatment of de re attitudes. In Gregory N. Carlson and Francis Jeffrey Pelletier, editors, *Reference and Quantification: the Partee Effect*, pages 207–228. CSLI Publications, 2005.