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

Erich Rast
erich@snafu.de

IFL, Universidade Nova de Lisboa

LogKCA-10, November 2010
Overview

1. Introduction

2. Preliminaries: STT, Two-Dimensional Semantics

3. Plausibility and Its Revision

4. Applications: Interpretative Assumptions, Some Form of Interpretation

5. Conclusion
Motivation: Why HOL?

- Why not? - Fun, curiosity, etc.
- Descriptive adequacy: Work with realistic representations of semantic content and (halfway) realistic examples.
- Investigate how (non-Gricean) notions of interpretation can be expressed in a two-dimensional setting.
- Integrate methods from formal epistemology with tools used commonly in general semantics (Montague Grammar, TLG, CCG).
- Long term goal: Express (aspects of) the interpretation of utterances on top of a traditional, two-dimensional semantic representation.
What Needs to be Done

- Implement two-dimensional semantics.
- Implement plausibility as a preorder to obtain a form of graded belief.
- Implement preorder revision.
- Apply in two-dimensional setting.
- Distinguish interpretative notions from linguistic notions.
- Implement a simple notion of interpretation:
  - Nonindexical and indexical expressions are interpreted.
  - The hearer evaluates the semantic content of a sentence on the basis of what he assumes that the speaker believes.
Types

Primitive types: e for entities, c for states, t for \{1, 0\}. If \( \alpha, \beta \) are primitive types, then \( \alpha \beta \) is a compound type. (Parentheses are left out – right-associativity is assumed.)

Syntax

If \( A \) is a term of type \( \beta \alpha \) and \( B \) is a term of type \( \beta \), then \( AB \) is a term of type \( \alpha \). If \( A \) is a term of type \( \alpha \) and \( x \) is a variable of type \( \beta \) then \( \lambda x A \) is a term of type \( \beta \alpha \). The identity symbol \( = \) is of type \( \alpha \alpha t \) for any type \( \alpha \).

- Standard conventions: dot notation, infix notation for common connectives, \( \forall x A \) for \( \forall (\lambda x A) \), leave out parentheses, e.g. \( A \times y \) instead of \( (A \times y) \), etc.
Simple Type Theory/HOL – Semantics

**Generalized Model**

A collection of non-empty sets $D_\alpha$ for primitive types $\alpha$, $D_{(\alpha\beta)} \subseteq D_\beta^{D_\alpha}$ for compound types $(\alpha\beta)$, and a term evaluation function $\llbracket \cdot \rrbracket^g$ under assignment $g$.

**Truth in a Model**

- $\llbracket A^\gamma \rrbracket^g = g(A)$ for variable $A$, where $g(A) \in D_\gamma$
- $\llbracket A^\gamma \rrbracket^g \in D_\gamma$ for constant $A$
- $\llbracket (A^{\beta\alpha} B^{\beta}) \rrbracket^g = \llbracket A \rrbracket^g (\llbracket B \rrbracket^g)$
- $(\lambda x^{\beta} A^{\alpha})$ is the function $f \in D_\beta^{D_\alpha}$ such that $\llbracket A^{\alpha} \rrbracket^g[x/a] = f(a)$ for any $a \in D_\beta$
- $\llbracket ((\alpha\alpha t \ A^{\alpha}) B^{\alpha}) \rrbracket^g = 1$ if $\llbracket A^{\alpha} \rrbracket^g = \llbracket B^{\alpha} \rrbracket^g$ (0 otherwise)
- $\llbracket (\wedge A^t) B^t \rrbracket^g = 1$ if $\llbracket A \rrbracket^g = 1$ and $\llbracket B \rrbracket^g = 1$ (0 otherwise)
- $\llbracket \neg tt A^t \rrbracket^g = 1$ if $\llbracket A \rrbracket^g = 0$ (0 otherwise)
Applicative Categorial Grammar

\[(\sigma/\tau) : A^{(\beta\alpha)} \rightarrow_{f} \tau : B^{\beta} \Rightarrow_{b} \sigma : (A B)\]  
 forward concatenation (1)

\[\tau : B^{\beta} (\tau \setminus \sigma) : A^{(\beta\alpha)} \Rightarrow_{b} \sigma : (A B)\]  
 backward concatenation (2)

Notes:

- Unlike in some other uses of CG, semantic representations are not evaluated.
- Shortcuts are possible, e.g. by different evaluation strategies (von Stechow/Zimmerman 2005), parameterizing truth in a model to contexts and indices, lifting modalities to the meta level, etc.
- For more realistic examples the full power of the Lambek calculus (hypothetical reasoning) plus other extensions like in TLG, CCG, Mortgat (1995) are needed.
Two-Dimensional Semantics

- Based on Kaplan (1989): In addition to indices for normal modal logical operators, add context parameters.
- Kaplan (1989) does not endorse structural symmetry of contexts and indices at all – but see e.g. others such as von Stechow/Zimmerman (2005).
- In the current setting: Make meanings of type $cc\alpha$, i.e. functions from an utterance ‘situation’ (state) to a function from a topic ‘situation’ (state) to an extension of type $\alpha$.
- Ideally, situations would be used.
  - See Muskens (1995) for a partial, relation type theory.
  - It is prima facie not clear how to implement revision in a partial setting (future research needed).
Two-Dimensionalism: Example

Let the lexicon entries for ‘I’, ‘me’ be $np : \lambda us.\text{speaker } u$, for ‘Mary’ be $np : \lambda us.m$, and for ‘loves’ be $(s \backslash np)/np : \lambda ijus.\text{love } s (jus) (ius)$, where $\text{speaker}$ is of type $ce$ and $\text{love}$ is of type $ceet$, where $i, j$ are intensional types $cce$. Then:

1. ‘Mary loves me’

   $= np : \lambda us.m (np\backslash s)/np : \lambda ijus.\text{love } s (jus) (ius)$

   $\Rightarrow np : \lambda us.\text{speaker } u$

   $\Rightarrow np : \lambda us.m np\backslash s : \lambda jus.\text{love } s (jus) (\text{speaker } u)$

   $\Rightarrow s : \lambda us.\text{love } s m (\text{speaker } u)$
Of course, everything can be formulated in the object language and familiar conditions on binary relations of type cct can be expressed as properties of relations:

\[
\begin{align*}
\text{Trans} & := \lambda P. \forall stu[(Pst \land Ptu) \to Psu] \\
\text{Eucl} & := \lambda P. \forall stu[(Pst \land Psu) \to Ptu] \\
\text{Ser} & := \lambda P. \forall s \exists t[Pst] \\
\text{Refl} & := \lambda P. \forall s[Pss]
\end{align*}
\]

KD45 Accessibility \(R^{ecct}:

\[
\forall x[\text{Trans}(Rx) \land \text{Eucl}(Rx) \land \text{Ser}(Rx)]
\]
Plausibility $\geq$ of type $eccct$ is implemented as a preorder in the last two argument places:

$$\text{Trans}(\geq xs)$$
$$\text{Refl}(\geq xs)$$

with additional condition

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

As usual,

- $s \sim_{x,u} t$ iff. $s \geq_{x,u} t$ and $t \geq_{x,u} s$,
- and $s >_{x,u} t$ iff. $s >_{x,u} t$ and not $t >_{x,u} s$. 
The maximum of a non-empty proposition $p$ of type $ct$ w. r. t. relation $C$ of type $eccct$ is computed by:

$$\text{Max} := \lambda xu \underbrace{Cp}_{\lambda q} \forall s \left[ (ps \land \neg \exists t [pt \land Cxuts \land \neg Cxust]) \equiv qs \right]$$ (15)
Based on van der van Benthem/Liu (2005), Liu (2008), Lang/vander Torre (2008): To revise by $p$, shift all $p$-states on top of the non-$p$-states.
As an auxiliary notion, let ‘when $A^t$ then $B_1^t$ otherwise $B_2^t$’ abbreviate $(A \rightarrow B_1) \land (\neg A \rightarrow B_2)$. The revision $C'$ of an ordering relation $C$ conditional on $P$ for some agent $x$ at $u_0$ is then characterized the following term.

$$\text{REV} := \lambda x u_0 P C. \forall y u s t [\text{when } u_0 = u \land x = y \land P u s \land \neg P u t \land R x u s \land R x u t \text{ then } (C' x u s \land \neg C' x u t s) \text{ otherwise } (C' y u s \equiv C y u s) \land (C' y u t s \equiv C y u t s)]$$

Read: For given base situation $u$, if $P u s$ and $\neg P u t$ for two states $s, t$, then ensure that $s >' t$ for the revised preorder $>'$, otherwise leave the preorder unchanged.
Belief vs. Interpretative Belief

Linguistic Belief

Indexicals are not evaluated:

$$\text{Bel} := \lambda xu_0 s_0 CP. \forall s_1 [((\text{Max } xs_0 C(Rxs_0))s_1 \rightarrow Pu_0 s_1] \quad (17)$$

Interpretative Belief

Indexicals are evaluated:

$$\text{IBL} := \lambda xu_0 s_0 CP. \forall u_1 s_1 [([\text{Max } xu_0 C(Rxu_0)]u_1 \land [\text{Max } xs_0 C(Rxs_0)]s_1) \rightarrow Pu_1 s_1] \quad (18)$$

Interpretative notions are analogous to using diagonalization in a double-index modal logic, using an operator such as $M, c, i \models \Delta \phi$ iff. $M, i, i \models \phi$ when contexts and indices are structurally alike.
The Structure of an Interpretative Belief
Non-iterated Interpretative Assumptions

Weak interpretative assumptions:

\[ IAW := \lambda xyu_0s_0CP. \forall u_1s_1s_2[([\text{Max } xu_0C(Rxu_0)]u_1 \land [\text{Max } xs_0C(Rxs_0)]s_1 \land [\text{Max } ys_1C(Rys_1)]s_2) \rightarrow Pu_1s_2] \quad (19) \]

Reflect

- ...what the hearer believes about the utterance situation
- ...what the hearer believes that the speaker believes about the topic situation

Strong interpretative assumptions:

Additionally reflect what the hearer believes that the speaker believes about the utterance situation.
Using Revision

\[
\text{RAW} := \lambda x y u_0 s_0 P. \nu Q \forall u_1 u_2 s_1 s_2 \left[ (\left( \text{Max } x u_0 \geq (R_x u_0) \right) u_1 \wedge \left( \text{Max } x s_0 \geq (R_x s_0) \right) s_1 \wedge \left( \text{Max } x s_1 (\text{REV } y s_1 P \geq (R y s_1)) \right) s_2 \right] \equiv Q u_1 s_2 \]

- That meaning/2d-intension Q such that Q holds according to what x believes that y believes given that y accepts P (for given base states \( u_0, s_0 \)).
- This notion is based on the revision by P of what y believes according to x’s beliefs, where the utterance situation is only taken into account according to x’s beliefs.
- Corresponding strong notion: Additionally it is taken into account what y believes about the utterance situation according to what x believes.
Some Form of Interpretation

$$\text{IPW} := \lambda xyusPQ.\text{IBL } xu(\text{REV } xu(\text{RAW } xyusP) \geq )Q \quad (21)$$

- Speaker utters a sentence whose meaning is $P$ (disregarding ambiguity for simplicity).
- Assumption: Speaker is not deceptive, sincere, etc.
- $\text{RAW } xyusP$ represents what the speaker $y$ believes given that $P$ (according to $x$'s beliefs).
- $\text{IPW } a\ b\ u_0\ s_0\ P\ Q$: $Q$ holds according to $a$'s interpretative belief generated by his believes revised by what he believes that the speaker $b$ believes $given\ that\ P$ (in given base states $u_0, s_0$).
- This form of interpretation captures the hearer's interpretation of the literal meaning of an utterance on the basis of a model of what the speaker believes. ($\neq$ Gricean speaker meaning)
Example (informal)

John: I am here. (22)
Mary: No, you’re not! You went to the Continental! (23)

- John and Mary want to meet in the lobby of the Holiday In.
- Suppose John is at the Continental and believes he’s at the Holiday In.
- Suppose Mary believes that John is at the Continental and also believes that he believes that he is at the Holiday In.
- Then strong and weak interpretation differ: This explains Mary’s reaction and can explain other reactions such as her going to the Continental despite believing that the original meeting was to take place at the Holiday In.
Conclusions

- (not surprising) Interpretation and belief as in ‘to believe’ are very distinct notions.
- There is a non-Gricean level of pragmatics that can be modeled directly on top of traditional representations of truth-conditional meanings computed from the lexicon.
- Once a hearer’s model of the speaker’s believes and its update is modeled, a number of fine-grained notions of belief, revision of iterated beliefs, and interpretation become available.
Open Issues

- How to implement soft update of the KD45 accessibility relation, i.e. when the agent takes into account things not previously considered, while maintaining the properties of the relation?
- Plausibility revision is categorical; is there a way to get more realistic graded revision?
- What’s the relation between the above plausibility revision and AGM or KM belief revision/upgrade?