Presuppositions, games, and bounded rationality

Gerhard Jäger
University of Bielefeld

1 Introduction

In van der Sandt (1992), an algorithm for the computation of presupposition resolution within the framework of DRT is given. It

(a) computes a set of possible readings of a sentence containing a presupposition,

(b) imposes a set of hard constraints for possible readings and thus possibly excludes some of the outputs of step (a), and

(c) defines a preference ordering on the remaining set.

The most preferred reading according to step (c) is the actual reading of the sentence. The constraints mentioned in (b) are **local informativity** (every sub-DRS is informative in its local context), **global informativity** (every sentence is informative), and **consistency** (no sub-DRS is inconsistent).[1] The preference ordering mentioned in (c) amounts to the claim that binding is better than accommodation, and high accommodation is better than low accommodation.

This algorithm is empirically highly successful. However, some of its aspects, especially the preference of high over low accommodation, are not independently justified. This abstract attempts to derive them from general principles of pragmatics, couched in game theoretic terms.

2 Epistemic logic and game theoretic pragmatics

Following most authors in game theoretic pragmatics, I assume that communication can be modeled as a signaling game in the sense of Lewis (1969). Inspired by Stalnaker (2005), I take it that natural language expressions have a conventionalized meaning that is common knowledge between the interlocutors. However, the actually transmitted information need not coincide with this literal meaning. Rather, pragmatic communication takes place in a Nash equilibrium of the underlying signaling game. The conventionalized meaning itself need not constitute a Nash equilibrium itself, but it forms the base for computing the pragmatic equilibrium.

I assume that interests of both players (speaker and hearer) are identical – they both want the speaker to transmit as much of his private knowledge to the hearer as possible as efficiently as possible.[2]

To be more precise, I assume that the hearer is interested in information about the world, not in opinions of the speaker about the epistemic state of the hearer. This can be formalized in terms of epistemic modal logic. Suppose the literal meaning of the message that the speakers emits is ϕ, where ϕ is an expression of multimodal epistemic logic, with at least two modalities, $\Box_S$ and $\Box_H$ (for speaker and hearer respectively). The hearer can thus provisionally assume that $\Box_S\phi$. However, the decision problem that the hearer faces

[1] Issues of variable binding, including the "trapping constraint", are ignored in the present abstract.

[2] Efficiency means that signals are not "cheap": coding complexity incurs costs, i.e. negative utility.

concerns the world as such, not the speaker's opinions about the epistemic state of the hearer himself. Technically, this means that any two worlds that only differ with respect to the worlds that are accessible from them for the hearer are considered identical for the purpose of the hearer's decision problem. Therefore the hearer will update his information state with ψ where ψ is the strongest formula **not containing modal operator $\Box_H$** such that $\Box_S\psi$ can be derived from $\Box_S\phi$. Since the speaker can anticipate this inference, what is pragmatically communicated is ψ rather than ϕ.

I will leave the question open which modal logic is appropriate to model pragmatic reasoning. As a lower bound, I assume system T, but I remain agnostic as to whether the introspection axioms do or do not hold.

3 Presuppositions

Presuppositions are modeled as statements about the knowledge of the hearer. So a sentence as (1a) is interpreted as (b), which can be paraphrased as (1c).

(1) a. The king of France is bald.
b. $\Box_H A \wedge B$ (side condition: $B \vdash A$)
c. As you know, France has a king, and this king is bald.

The communicated meaning is the strongest non-modal statement that can be derived from

$$\Box_S(\Box_H A \wedge B)$$

In this example, this would be the formula B (by applying K and T once each, plus some propositional reasoning). Depending on the common ground, this can be interpreted as an instance either of binding or of accommodation. In the sequel, I will focus on accommdation, but I will return to the issue of binding later.

If a presupposition trigger is embedded under some operator, this gives rise to an ambiguity between local and global accommodation. For instance, if (1) is negated, this can be construed either as (b) or as (c).

(2) a. It is not the case that the king of France is bald.
b. $\neg(\Box_H A \wedge B)$
c. $\Box_H A \wedge \neg B$

Consider (2b). From $\Box_S(\neg(\Box_H A \wedge B))$, there is only one $\Box_H$-free proposition that the hearer can infer, nameley the tautology. From $\Box_S(\Box_H A \wedge \neg B)$, however, he can infer B via two applications of (T), plus some propositional reasoning. The current model thus predicts that in (2), only global accommodation is possible (since it is never rational for the speaker to communicate a tautology). This result is welcome, because local accommodation is only construable as a denial here – a type of speech act that goes beyond the scope of the present paper.

Let us now consider a case where a presupposition is embedded under a modal operator. Again, the presupposition can be accommodated locally or globally. (The adverb *perhaps* is translated as an epistemic possibility operator, the dual of $\Box_S$.)

(3) a. Perhaps the king of France is bald.
b. $\Diamond_S(\Box_H A \wedge B)$

c. $\Box_H A \wedge \Diamond_S B$

The strongest $\Box_H$-free formula that can be derived from $\Box_S\Diamond_S(\Box_H A \wedge B)$ (the local accommodation reading), is $\Box_S\Diamond_S(A \wedge B)$, as the following semi-formal proof shows ("PC" abbreviates "propositional calculus"):

$$\begin{array}{cl}
\Box_S\Diamond_S(\Box_H A \wedge B) & (ass.) \\
\Box_H A \rightarrow A & (T) \\
\Box_H A \wedge B \rightarrow A \wedge B & (PC) \\
\Diamond_S(\Box_H A \wedge B) \rightarrow \Diamond_S(A \wedge B) & (K+PC) \\
\Box_S\Diamond_S(\Box_H A \wedge B) \rightarrow \Box_S\Diamond_S(A \wedge B) & (K+PC) \\
\Box_S\Diamond_S(A \wedge B) & (PC)
\end{array}$$

So the communicated meaning can be paraphrased as "Perhaps there is a king of France who is bald."

If we turn to (3b), the strongest $\Box_H$-free formula derviable from $\Box_S(\Box_H A \wedge \Diamond_S B)$ is $\Box_S(A \wedge \Diamond_S B)$, which can be paraphrased as "There is a king of France, and perhaps he is bald." The corresponding proof is as follows:

$$\begin{array}{cl}
\Box_S(\Box_H A \wedge \Diamond_S B) & (ass.) \\
\Box_H A \rightarrow A & (T) \\
\Box_H A \wedge \Diamond_S B \rightarrow A \wedge \Diamond_S B & (PC) \\
\Box_S(\Box_H A \wedge \Diamond_S B) \rightarrow \Box_S(A \wedge \Diamond_S B) & (K+PC) \\
\Box_S(A \wedge \Diamond_S B) & (PC)
\end{array}$$

Note that the proof for the latter reading requires two applications of modal axioms (T and K), while the corresponding proof for the local accommodation reading requires one additional application of K.

The argument can be continued with more complex examples, like

(4) Perhaps John believes that the King of France is bald.

Here we have three options: local, intermediate and global accommodation. The corresponding proofs contain at least one application of K for the global reading, at least two for the intermediate and at least three for the local accommodation reading.

4 Bounded rationality

As indicated in the previous section, the proof that a certain accommodation reading is a Nash equilibrium requires successively more applications of modal axioms the deeper the accommodation site is embedded under modal operators. The preference for high accommodation can thus be interpreted as a strategy to avoid proof complexity. Under a game theoretic perspective, this makes sense if we take into account that **reasoning consumes resources**. This is one aspect of the often observed fact that "real" agents are not the perfectly rational beings that classical game theory (or traditional pragmatics, for that matter) assumes them to be. Economists call this insight "bounded rationality". For reasons of space I only sketch a formalization for the present application: Ambiguity is modeled as uncertainty of the hearer about the identity of the signal that the speaker emits. It is common knowledge that for each reading, the intended interpretation constitutes a Nash equilibrium. In the simplest case, the hearer does not know which signal is intended though and considers all resolutions equally likely. Reasoning incurs costs that are infinitesimal if compared to the utility of succcessful communication, but not completely negligible. Therefore a utility maximizing listener will minimize reasoning costs and thus resolve presuppositions as high as possible.

Little is known about the actual cognitive costs of reasoning. It seems plausible though to assume that modal reasoning is massively more costly than plain propositional reasoning. Counting applications of modal axioms is thus a first, if crude, approximation of this aspect of bounded rationality.

5 Conclusion

For reasons of space, I disregarded presupposition binding. Let me point out though that in van der Sandt's theory, binding involves reasoning about known variables, while accommodation requires the introduction of new variables. In a first order system, accommodation thus incurs applications of quantifier proof rules while binding doesn't. I conjecture that quantification rules incur reasoning costs comparable to quantification rules.

The hard constraints that van der Sandt assumes to restrict resolution options can be accounted for straightforwardly in the game theoretic setting. It is part of the very notion of a signaling game that contradictory or uninformative messages lead to low utility and therefore cannot be part of a Nash equilibrium strategy. Locally uninformative sub-DRSs—that correspond to redundant parts of syntactic structure—incur complexity costs for the speaker without increasing the value of the transmitted information and thus cannot be part of a rational strategy either.

REFERENCES

Lewis, David. 1969. Convention. Cambridge: Harvard University Press.

van der Sandt, Rob. 1992. Presupposition projection as anaphora resolution. Journal of Semantics 9: 333–377.

Stalnaker, Robert. 2005. Saying and meaning, cheap talk and credibility. In: Anton Benz, Gerhard Jäger and Robert van Rooij (eds.). Game Theory and Pragmatics. Palgrave MacMillan. 83–100.