Overview of My Philosophical Research

Rich Thomason

Version of: November 15, 2006

This will be posted at

http://www.eecs.umich.edu/~rthomaso/documents/general/research-overview.pdf

Outline

1. **A problem with research in philosophy (and one strategy for overcoming it).**
2. **How a research program in philosophical logic can lead into Linguistics and AI.**
3. **Deontic logic, practical reasoning, reasoning about action and change, agent architectures.**
4. **Interlude: What is nonmonotonic logic?**
5. **A logicist program for practical reasoning. (Hasty, rough indication).**
6. **Reasoning about the attitudes of other agents, achieving mutuality, and the modularity of the attitudes.**
7. **An architecture for reasoning in pragmatics. (I.e., for the interpretation and generation of discourse.)**
8. **Other topics (shown, not discussed)**

A Problem with Research in Philosophy (And One Strategy for Overcoming It)

Why Work in Other Fields?

- **Like Linguistics?**
- **And Artificial Intelligence?**

- **One reason is that this is where a lot of the action is in philosophical logic,**
- **And you want to follow the action.**
- **I guess this is a tactical reason.**

But However

- **I think there are more important, strategic reasons for philosophy.**
- **Philosophy is hard.**
- **It is *very, very* hard to do anything that is really new, and substantive**
- **There is a lot of redefining success.**
- **That is, many people forget the history and the literature on a problem.**
- **And you can often see the wheels spinning,**
- **E.g., generating points in position space,**
- **Or far-fetched examples.**

A Program

- **Maybe a good way to avoid trying to say something that was said better over 2000 years ago is to use techniques and marshall evidence that were unknown then.**
- **Also, there are a number of problems in philosophy that are too complicated to think about without the aid of models and formal techniques to structure the topic.**
- **Linguisitcs provides new techniques and evidence that can help with philosophical problems.**
- **Artificial Intelligence provides new techniques and evidence that can help with philosophical problems.**
- **But when a genuine philosophical problem arises that can be dealt with in these areas, you're unlikely to find it solved.**
- **So to use the techniques, you have to learn to be a linguist, or to be a computer scientist.**

How a Research Program in Philosophical Logic Can Lead into Linguistics and AI.

- **A large part of philosophical logic is concerned with using techniques from symbolic logic to**
  - **deal with nonmathematical domains**
  - **and maybe to account for nonmathematical types of reasoning**

- **In the course of doing this, you generate alternative theories,**
  - **For instance, of presupposition,**
  - **Or of conditionals,**
- **And you would like to know which is right.**
- **This leads you into linguistics.**

- **In many cases, you have as alternatives a semantic and a pragmatic account**
- **This happens with presupposition and conditionals**
- **And the semantic account is less plausible than the pragmatic one,**
- **But the pragmatic account is more or less hopelessly underdeveloped as a theory, and untestable.**
- **This is frustrating.**

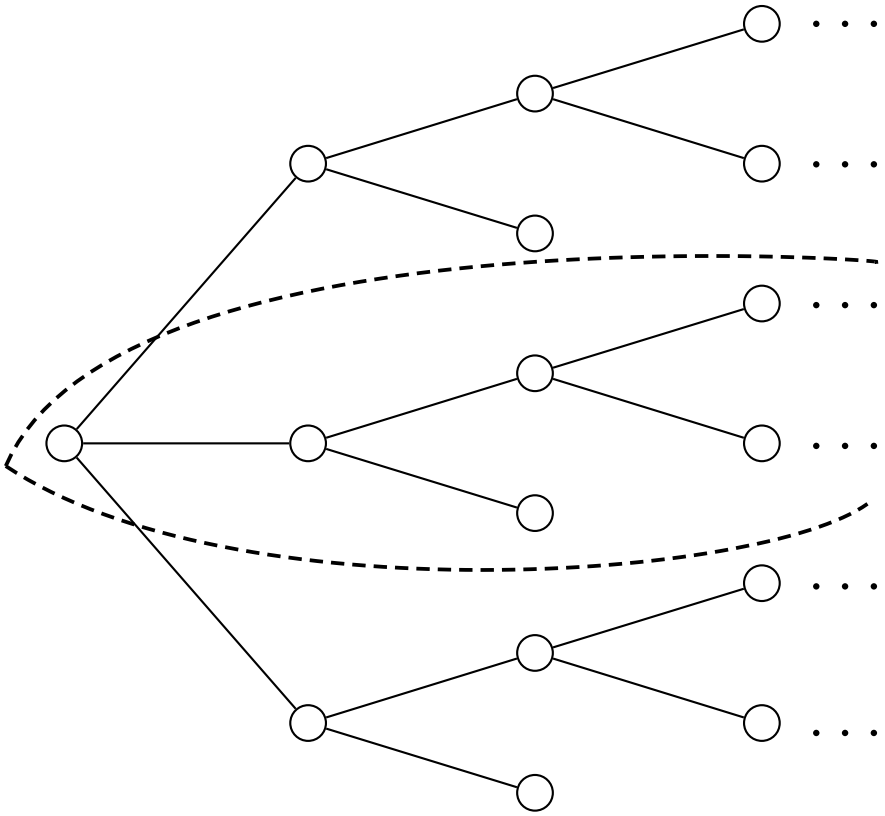
- **Pragmatic accounts like Grice's are underdeveloped because they would get too complicated to manage if you tried flesh them out properly.**
- **But Artificial Intelligence has methods for dealing with problems like this.**
  - **Use computer programs to store and test large sets of rules.**
  - **Simplify the general problem by working with a restricted domain in which the knowledge required to do the resoning is limited**
  - **Piggyback.**
  - **Try to gradually scale up.**

Topic 1: Deontic logic, practical reasoning, reasoning about action and change, agent architectures.

Deontic Logic

$M, w \models \mathbf{O}\phi$ **iff** $M, w' \models \phi$ **for all** w'**,** $w \, R \, w'$

Deontic Logic and Branching Time



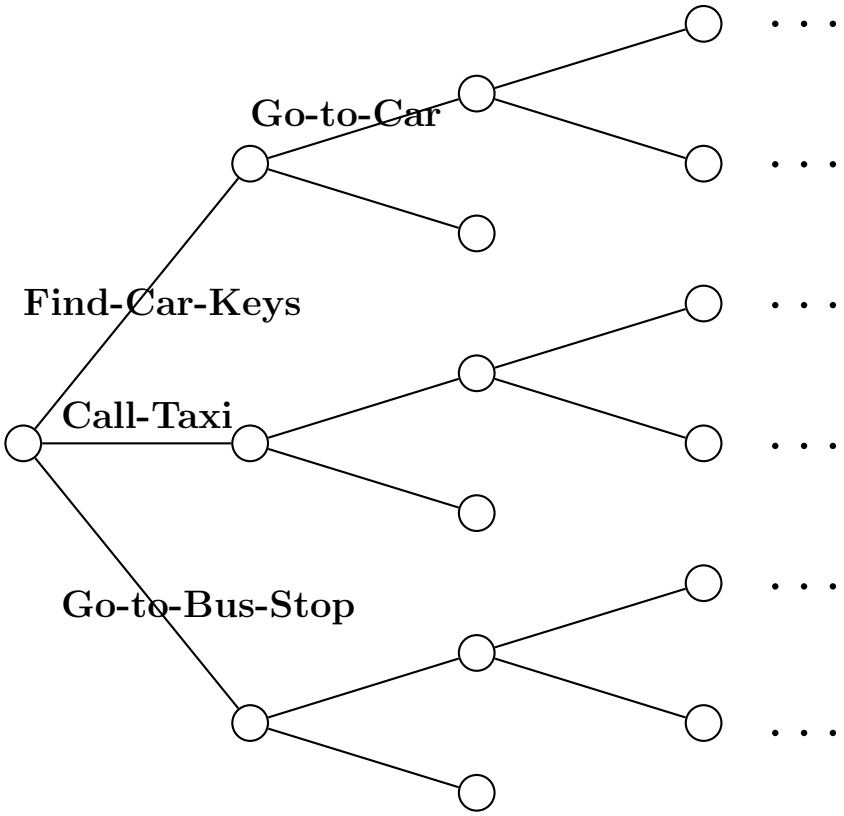
Upshot

- **Linguistic evidence (which was discussed at the recent workshop) suggests that all this is constrained by a contextually fixed presupposed set of alternatives—what you have to do.**
- **And that there is a preference relation underlying the O operator, which looked absolute in the earlier work.**
- **There are some suggestive points here, but the relationship between these models and practical reasoning has always struck me as pretty tenuous.**

Reasoning about Action and Change

- **Idea: make the transitions in a temporal model *action-driven*.**
- **An agent has a repertoire of actions—these change the state of a world.**
- **Use domains like the blocks-world to generate examples.**

Getting to the Airport



Actions as Operators on States

- **States are like possible worlds—they make propositions (or "fluents") true or false.**
- **Actions change states.**
- **They have effects (direct changes that will be enforced if the action is performed). These changes can have "ramifications" or causal side-effects.**
- **They have preconditions that must obtain when the action is performed for the action to have its effects.**
- **All this is put in the form of a logical theory.**
- **In the earliest, simplest theories there is only one agent, there are no exogenous changes, and there are no sources of uncertainty.**

Predicting the Future

- **Even in the simplest cases, the problem of figuring out what the world will be like if you perform an action is nontrivial.**
- **Especially when you try to explain the locality of actions by invoking the *law of commonsense inertia***

  > **Changes can only occur if the performance of the action in the initial state provides a reason for them to change.**

- **The best solutions to this problem make an interesting contribution to the problem of causality.**

Planning as Regression from a Goal
Plan Verification as Theorem Proving

- **An agent in an initial situation s has a repertoire of actions and a propositional goal G, which is false in s_0.**
- **A *plan to achieve* G is a series of actions, which will take you from s_0 to a state in which G is true.**
- **For the plan to be correct, the preconditions of the $n+1$st action have to be true in the nth state.**
- **A natural way to search for a plan is to create *subgoals*.**
- **A planning agent needs *beliefs* and *goals* (aka desires) and uses planning to generate *plans* (aka intentions).**
- **Hence, BDI agents.**

Interlude: Nonmonotonic Logic

Monotonicity Defined

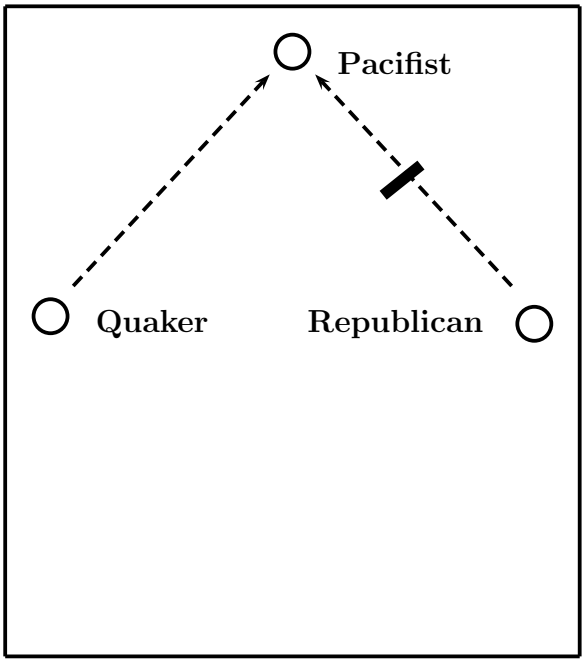
- **In classical logic (and in mathematics), reasoning yields persistent conclusions.**
- **Any logic delivers a consequence relation $\vdash$ between premises and conclusion.**
- **Monotonicity is a property of $\vdash$.**
- **Monotoniticy: If $T \vdash B$ then $T, A \vdash B$.**
- **Here, $\vdash$ is the relation of logical consequence, T is a set of premises, B is a formula, and T, A is the result of augmenting T with a formula A.**
- **Commonsense reasoning is not like this: I believe my printer is in my office, because I left it there. But if someone tells me that the door to my office has been forced, I may retract this belief.**
- **A nonmonotonic logic delivers a nonmonotonic consequence relation.**

Rule-Based NM Logics

- **There are many approaches to NM Logic.**
- **The rule-based approach (Ray Reiter) adds *default rules***

$$\mathbf{A_1}, \ldots \mathbf{A_n} \hookrightarrow \mathbf{B}$$

- **Default rules can conflict: the Nixon Diamond.**



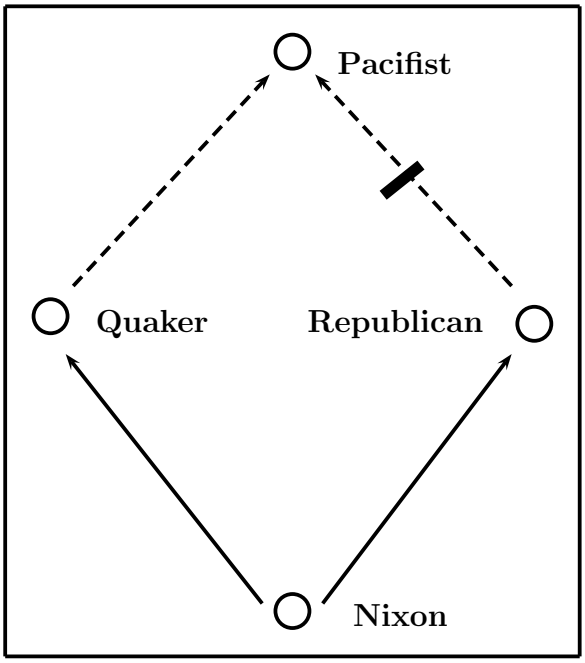
$$\mathbf{Quaker}(x) \hookrightarrow \mathbf{Pacifist}(x)$$
$$\mathbf{Republican}(x) \hookrightarrow \neg\mathbf{Pacifist}(x)$$

Rule-Based NM Logics

- **There are many approaches to NM Logic.**
- **The rule-based approach (Ray Reiter) adds *default rules***

$$\mathbf{A_1}, \ldots \mathbf{A_n} \hookrightarrow \mathbf{B}$$

- **Default rules can conflict: the Nixon Diamond.**



$$\mathbf{Quaker}(x) \hookrightarrow \mathbf{Pacifist}(x)$$
$$\mathbf{Republican}(x) \hookrightarrow \neg\mathbf{Pacifist}(x)$$
$$\mathbf{Nixon}(x) \rightarrow \mathbf{Quaker}(x)$$
$$\mathbf{Nixon}(x) \rightarrow \mathbf{Republican}(x)$$

Multiple Consequence Sets

- **How to deal with conflicts?**
- **Treat consequence as a relation between premises T and multiple consequence sets $E_1, E_2, \ldots$.**
- **Reasoners can disagree about what consequences to draw from rule-sets.**
- **Or maybe other forms of reasoning can be invoked to resolve conflicts.**

Return to Practical Reasoning

The Reasoning Target

- **I have to teach summer school in Chicago.**
- **I'd like to have a car, but it's too far to drive, and it would be too expensive to rent a car.**
- **So I'll fly to Chicago.**
- **I'll need some clothing and a big box of books, and a laptop computer.**
- **It would be a hassle to get all that stuff to the airport and take it on the plane.**
- **So I'll ship the books.**
- **So I might as well ship all the stuff I'll need.**
- **But the laptop is expensive—I'll carry that.**

Some Limitations to the Classical Planning Formalisms

1) The logicist planning community doesn't recognize the importance of plan evaluation.

2) There is no good way to deal with plan monitoring, plan modification.

3) There is no way to deal with uncertainty and risk, at least within the limits of this formalism.

4) Goals are simply given, and on some accounts are adhered to until achieved or shown infeasible. There is no reasoning about desires.

Prima Facie and All-Things-Considered Beliefs

Example 1. Beliefs about the porch light.

(i) I have a reason to believe the porch light is off, because I asked my daughter to turn it off.

(ii) I have a reason to believe the porch light is on, because the last time I saw it, it was on.

(iii) All things considered, I believe the porch light is off, because my daughter is pretty reliable.

Wishes/Wants

Part of commonsense practical reasoning consists in the practicalization of desires. Immediate desires needn't be feasible, and typically will conflict with other immediate desires. We do not expect all of these wishes to survive as practical goals. The ones that do survive I will call *wants*.

This distinction seems to correspond to one important difference beteen the way 'wish' and 'would like' on the one hand and 'want' on the other are typically used. In particular:

Interaction with Beliefs

- **Wishes can conflict with beliefs.**

  Example 2.
  I'd like to take a long vacation.
  I'd need to get time off from work to take a long vacation.
  But: **I can't get time off from work.**

- **Wishes can conflict with each other, in lignt of background beliefs.**

  Example 3.
  I'd like to take a long vacation.
  But: **I'd like to save more money this year.**
  And: **I can't save more money this year and take a long vacation.**

- **Wishes can conflict with intentions, or more generally with adopted plans. This point is made by Michael Bratman, David Israel, and Martha Pollack. See Bratman, 1987.**

- **For present purposes, it is not important to distinguish between wants and intentions.**

Formalize with Two Sorts of Defaults

- **Wishes are like *prima facie* beliefs.**
- **So, use defaults for both. But use a notation that lets us keep track of which is which.**

  $$A \overset{\mathbf{B}}{\hookrightarrow} C$$

  versus

  $$A \overset{\mathbf{D}}{\hookrightarrow} C.$$

  (Note: we are limiting ourselves to normal discourse.)
- **Wants are like all-things-considered beliefs.**
- **So treat both as conclusions in a selected extension generated by the defaults. Don't distinguish the two types of conclusions notationally. But we can account for the difference in terms of the reasons that explain why a conclusion belongs to the extension.**
- **Competing preferences need to be resolved in choosing extensions. This leads room for (local) quantitative reasoning.**

A Reasoning Example

Example 4. Part I: **Commonsense reasoning.**
(Imagine a restaurant scenario.)

1. **I'd like to have some coffee.**
2. **For me to have coffee, coffee will have to be available.**
3. **I'd like to have decaf if I have coffee.**
4. **Defaf must be available if coffee is available.**
5. **Coffee is available.**
6. **For me to have decaf coffee, I'll need to order decaf coffee.**
7. ***So:*** **I'll order decaf coffee.**

The Formalization

$\top \overset{\mathbf{D}}{\hookrightarrow}$ **Coffee**

Coffee $\overset{\mathbf{B}}{\hookrightarrow}$ **Available**

Coffee $\overset{\mathbf{D}}{\hookrightarrow}$ **Decaf**

Available $\overset{\mathbf{B}}{\hookrightarrow}$ **Decaf-Available**

Available

Decaf $\overset{\mathbf{B}}{\hookrightarrow}$ **Order-Decaf**

Logical Consequences

There is one extension, which is generated by the following choices:

$\{\textbf{Coffee}, \textbf{Available}, \textbf{Decaf}, \textbf{Decaf-Available}, \textbf{Order-Decaf}, \}$

Note: **The use of the premise Decaf $\overset{\text{B}}{\hookrightarrow}$ Order-Decaf is a makeshift. The selection of an action to achieve an end should be carried out by means of a planning process. I intend to explain in a later paper how to integrate the formalism with planning.**

Reasoning about Other Agents' Attitudes, Achieving Mutuality, And the Modularity of the Attitudes.

The Reasoning is Pervasive

- **Everyday examples of the following sort show that our beliefs about other people's attitudes are detailed and extensive.**

  Case 1. Given: **that a person a is sitting next to me on an airplane and is reading an American newspaper.**
  I believe: **that she believes that Donald Rumsfeld recently resigned as Secretary of Defense.**

Case 2. Given: **everything in Case 1, and that a is an academic.**

I believe: **that she doesn't approve of Bush's policies.**

Case 3. Given: **everything in Case 2, and that a is a philosopher.**

I'm not sure: **whether she believes that the frame problem is a problem having to do with reasoning about actions.**

Case 4. Given: **everything in Case 3.**

I'd guess: **that a doesn't know what the qualification problem is, so I'd guess that a doesn't believe that the qualification problem has to do with reasoning about actions.**

Characteristics of the Reasoning

- **The level of detail is extremely rich.**
- **Any account of the reasoning that isn't equally detailed can't be at all plausible.**
- **However, our intuitions about the reasoning are relatively shallow.**
- **Intuitively, such beliefs seem almost to be immediate;**
- **At least, they come to mind more or less effortlessly,**
- **And reflecting on them doesn't reveal a breakdown into steps.**

Theoretical Importance of Mutuality

- **More or less independently, researchers in many different areas, including:**
  - **Philosophy**
  - **Microeconomics**
  - **Distributed Systems**
  - **Psycholinguistics**

  have concluded that robust mutual knowledge is essential in the theoretical models they have constructed of social knowledge and reasoning.

Formalizing Mutual Belief

– **Proposition p is mutually believed by group $\{\mathrm{a}, \mathrm{b}\}$:**

$[\mathrm{a}]p$, $[\mathrm{b}]p$ **are true.**

$[\mathrm{a}][\mathrm{b}]p$, $[\mathrm{b}][\mathrm{a}]p$ **are true.**

$[\mathrm{a}][\mathrm{b}][\mathrm{a}]p$, $[\mathrm{b}][\mathrm{a}][\mathrm{b}]p$ **are true.**

$\vdots$

How Do We Obtain Mutual Belief?

- **(Lewis, Schiffer, accepted by many others.)**
- **There are circumstances which when present necessitate mutual belief.**
- **So when these circumstances are recognized by both participants, mutuality is guaranteed.**
- **The problem is that you can find exceptions to any plausible example.**

An Idea

- **Herbert H. Clark and Michael Schober, "Understanding by Addressees and Overhearers," Cognitive Psychology, 24:259–294, 1989.**

  The common ground between two people—here, Alan and Barbara—can be divided conceptually into two parts. Their *communal common ground* represents all the knowledge, beliefs, and assumptions they take to be universally held in the communities to which they mutually believe they both belong. Their *personal common ground* represents all the mutual knowledge, beliefs, and assumptions they have inferred from personal experience with each other.

  Alan and Barbara belong to many of the same cultural communities . . .

  1. ***Language*: American English, Dutch, Japanese**
  2. ***Nationality*: American, German, Australian**
  3. ***Education*: University, high school, grade school**
  4. ***Place of Residence*: San Francisco, Edinburgh, Amsterdam . . .**

  . . . People must keep track of communal and personal common ground in different ways. For communal common ground, they need encyclopedias for each of the communities they belong to. Once Alan and Barbara establish the mutual belief that they are both physicians, they can immediately add their physician encyclopedias to their common ground.

Modularity: Subagent Modalities

- **Associate a set $\mathcal{I}_a$ of *subagents* with each agent a.**
- **Each subagent $i \in cI_a$ induces a Kripke relation $R_{a,i}$ over possible worlds.**
- **There is an ordering $\preceq_a$ over $\mathcal{I}_a$. If $i \preceq j$, then i can access information from j.**
- **This means that if $i \preceq_a j$ and $wR_{a,i}w'$, then $wR_{a,j}w'$.**
- **We want to think of each subagent as being associated with a set of features that classify propositions, like *what any American English-speaking university-educated person could be expected to believe.***

Achieving (Belief in) Mutuality by Default

- **We can prove a theorem along the following lines.**
- **Let T be a theory that contains no formulas involving b's beliefs or abnormality predicates, and that also contains $\textsc{Said}(p)$ where p is a propositional atom.**
- **Then T circumscriptively implies $[a][\textsc{mut}]p$.**

A Moral for Belief

- **Belief is not a monolithic attitude.**
- **Think of a large familiy of attitudes that we may be more or less willing to use for practiacal purposes,**
- **Each managed by a specialist,**
- **Which may be permanant or more or less *ad hoc*.**
- **Communication between specialists may be limited in some ways.**

An architecture for pragmatic reasoning (I.e., for the interpretation and generation of discourse.)

The Reasoning that Produces Inferred Meanings is Robust

"A Very Simple Story" (by Wendy Lehnert)

When the balloon touched the light bulb, it broke. This made the baby cry. Mary gave John a dirty look and picked up the baby. He shrugged and picked up the balloon.

- **The balloon was originally inflated.**
- **The balloon broke (not the light bulb).**
- **The light bulb was on.**
- **The light bulb was hot.**
- **The heat caused the balloon to break.**
- **The balloon exploded.**
- **The explosion made a loud noise.**
- **The baby was scared.**
- **The loud noise scared the baby.**
- **The baby cried because it was scared.**
- **Mary was mad at John.**
- **Mary was mad at John for making the baby cry.**
- **Mary communicated this to John by the way she looked at him.**
- **Mary picked up the baby to comfort it.**
- **John (not the baby) shrugged and picked up the balloon.**
- **John was not overly concerned.**
- **John will throw the balloon away.**

The Problem

It takes a combination of:

- **Linguistic knowledge**
- **World knowledge**
- **Intelligent specialization of these context**

to relate meanings to the linguistic forms that express them.

The Problem Shows up All Over the Place

- **Ambiguity resolution**
- **Implicature**
- **Resolution of discourse relations**
- **Anaphora resolution**
- **Interpretation of metaphor**
- **Resolution of noun compounds**

Context Can be Important

1. Mary was in the hospital.

A man took her flowers.

2. Mary was jogging in Central Park

A man took her money.

Project with Matthew Stone

- **Global Objectives:**
  - **Integrate language processing with nonlinguistic reasoning and knowledge.**
  - **Do this for interpretation and generation, without duplicating the knowledge sources.**
  - **Try to account for the entire range of pragmatic reasoning.**
  - **And do this in a way that provides a mechanism for the operation of context.**
- **Of course, this has to be done in incremental steps.**

Abduction 1

Premisses

1. **The sidewalk is wet.**
2. **If it has rained recently, the sidewalk will be wet.**
3. **If the sprinkler has been on recently, the sidewalk will be wet.**

Conclusion

It must have rained recently

Abduction 2

Premisses

1. $\texttt{wet}(s_2)$
2. $\texttt{rain}_{.1} \rightarrow \texttt{wet}(s_2)$
3. $\texttt{sprinkler}_{.9} \rightarrow \texttt{wet}(s_2)$

Conclusion

rain

Abduction 3—Stickel's Algorithm

Backward-chain from goal through Horn-clause rules searching for a proof, as in Prolog.

But, instead of failing when a proof is not available, incorporate this with a best-first search for a set of low-cost literals that will enable a proof when added to the knowledge base.

No added costs for multiple uses of same assumption.

There are no consistency-checks, and the search is not exhaustive.

An Example from the Navy Casualty Reports Domain

KB		LF
$lube\text{-}oil(o_3)$	$\rightarrow$	$lube\text{-}oil(o)$
$alarm(r_5)$	$\rightarrow$	$alarm(a)$
$for(r_5, o_3)$, $for(X, Y) \supset nn(Y, X)$	$\rightarrow$	$nn(o, a)$
		$\boxed{sound'(e, a)}$

– This is an example of Jerry Hobbs'
– Arrows indicate inferences from KB
– From facts about a particular kind of lube oil o_3 and a particular alarm r_5
– To LF conjuncts
– The box indicates a literal that is assumed rather than derived.
– Matthew and I would say: the rererents that are inferred (the sample of lube oil and the alarm) need to be made salient by the context for this to work.
– for is too ambiguous to be a good logical target for NN resolution

Abductive Discourse Planning

Abductive planning infers actions from goals.

In planning stretches of discourse,

- **The goals are communication goals;**
- **The actions are speech acts;**
- **Utterances are methods of performing the speech acts.**

Sketch of an Axiom

The axiom for proposing says that

(1) **If the presuppositions of a proposal to to put c_7 into room r_1 are in the common ground**

(2) **and an utterance realizing the proposal occurs**

(3) **and other default (low-cost) assumptions are created,**

then **the corresponding common goal will be created.**

- **The axioms can be used for both interpretation and generation.**
- **The only difference in these uses is that some assumption costs are different, depending on whether the axiom is used**
  - **– the speaker for generation**
  - **– or by the hearer for interpretation.**

For instance, the axiom that provides methods of performing introductions look like this.

- $\texttt{introduce}'(e_1, \textbf{\textit{spkr}}, e_3)\langle L, L\rangle$
- $\wedge\, \texttt{have}'(e_4, \textbf{\textit{spkr}}, f)\langle L, L\rangle$
  $\wedge\, \textit{in-cg}'(e_3, e_4)\langle L, L\rangle$
  $\wedge\, \texttt{utter}'(e_2, \textbf{\textit{spkr}}, \texttt{i-have1-couch}, c, p, e_3)\langle L, L\rangle$
  $\wedge\, \texttt{couch}'(f)\langle H, L\rangle$
  $\wedge\, \texttt{color}'(f, c)\langle H, L\rangle$
  $\wedge\, \texttt{price}'(f, p)\langle H, L\rangle$
  $\supset\, \textit{embodies}\,(e_2, e_1)$

If f has a couch f of color c and price p, then uttering 'I have a t couch for $\$p$' will embody the speech act of introducing f.

When an interpreter uses this axiom, the utterance is known and the speech act is being inferred.

The color is given in the utterance, the speaker wants the hearer to assume the item has this color. So the assumption cost for the hearer is low.

When the generator uses this axiom, the goal is known and the utterance is being inferred.

The color that is given in the utterance has to be found in the speaker's knowledge base; it can't be assumed in the course of the derivation.

So the assumption cost for the hearer is high.

The notation $\texttt{color}\,(f, c)\langle H, L\rangle$
means that the cost for the speaker is high, the cost for the hearer is low.

Preferences in Interpretation

– Informational Preferences

Prefer discourse that builds on mutual information. (Hobbs, Clark, presupposition theorists, Asher & Lascarides)

– Attentional Preferences

– Prefer reference to salient objects (Grosz & Sidner, Grosz, Joshi & Weinstein, assorted psycholinguists)

– We need to model both kinds of preferences, and how they interact.

Our Idea about Interpretation

An interpretation of an utterance is
an explanation of how the utterance
creates a new discourse context
in which its content is

true and
prominent.

Our Idea about Generation

Generating an utterance is planning to produce
an explanation by the hearer of how the utterance
creates a new discourse context
in which its content is

true and
prominent.

Our Idea about Context

– **Context supplies both a body of information and a ranking of salience.**

Both rankings change dynamically as discourse is updated.

Contexts and Context Change

– **A context is a structure $\langle i, a \rangle$ with an informational component i and an attentional component a.**

– **Utterances correspond to operators on context.**

– **For instance, "Susan met Mary" updates the informational component with the claim that that Susan met Mary. But it also changes the attentional component by making Susan prominent.**

– **Representing the first update:**

$$i : \texttt{met}'(\texttt{Susan}, \texttt{Mary}).$$

– **Representing the second update:**

$$\begin{aligned} & a_2 : \textit{in-focus}^*(\texttt{Susan}) \\ \wedge\ & a_2 : \textit{in-focus}(\texttt{Mary}) \\ \wedge\ & a_1[\texttt{Susan} \leq \texttt{Mary}]a_2 \end{aligned}$$

(1) `Mary` **is the central focus in the resulting attentional context.**

(2) `Susan` **is in focus in the resulting attentional context.**

(3) a_2 **is the result of modifying** a_1 **so that** `Mary` **is preferred to** `Susan`**.**

An Example that Puts it All Together

Susan met Mary.

She asked her a question.

She answered no.

Some Things I Didn't Talk About

- **Lexical semantics and the meaning of *-able*.**
- **The logic of ability.**
- **Vagueness.**
- **Counterpart theory.**
- **The logic of context and contextual reasoning.**