A framework for modeling the relationships between the rational and behavioral reactions of assisting conversational agents

François Bouchet Jean-Paul Sansonnet

LIMSI-CNRS
Université Paris-Sud XI

December 18th 2009

EUMAS 2009

Outline

Outline

Assisting Conversational Agents (ACA)

Issues of assistance to novice users

- "Paradox of motivation" *(Carroll & Rosson, 1987)*
- Users prefer the help provided by "a friend behind the shoulder" *(Capobianco & Carbonell, 2001)*

A solution: conversational agents for assistance

- "Persona Effect": an animated agent increases credibility *(Lester, 1997)*
- Natural language: ideal modality when facing cognitive distress *(Carbonell, 2003)*

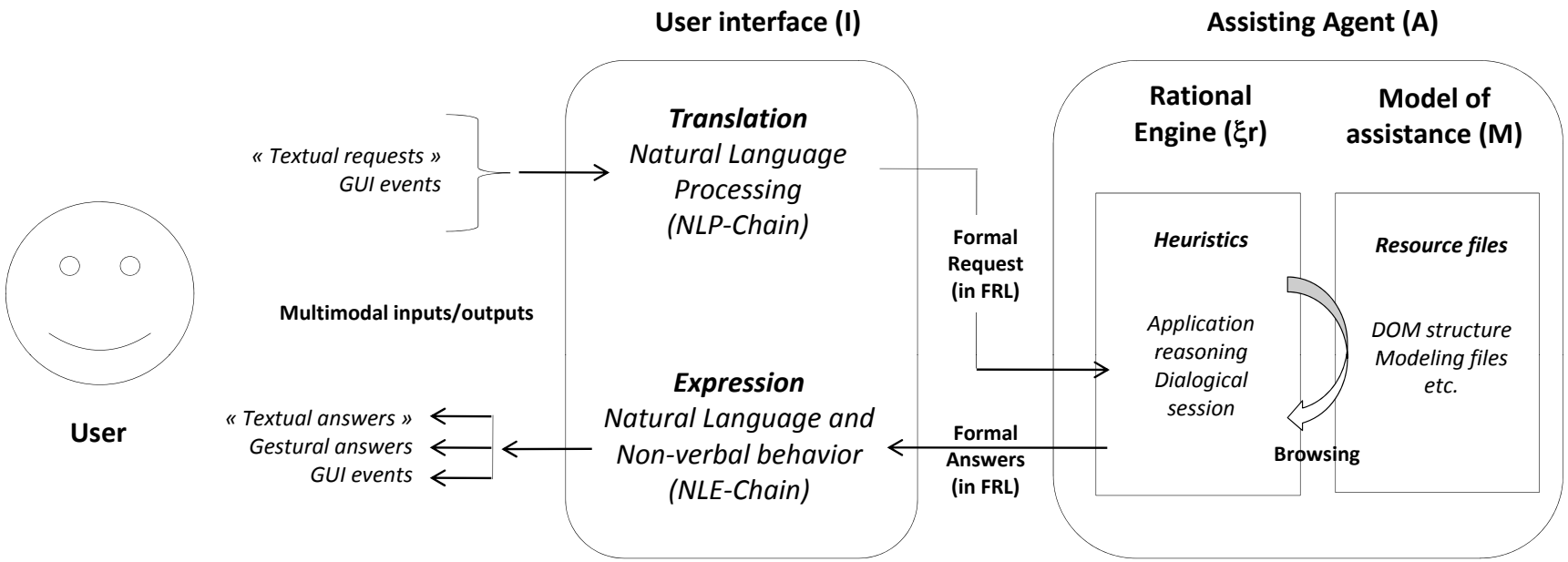
Realistic Assisting Conversational Agents

To be used, must look like "the friend behind the shoulder":

- Embodiment: movements, emotions rendering...
  $\longrightarrow$ suitable with its visual realism
  – or risks to fall into the "Uncanny valley" *(Mori, 1970)*

- Cognitively: coherent personality, credible reactions to requests...
  $\longrightarrow$ suitable with its embodiment
  – or risks to reproduce the "Clippy Effect" *(Xiao et al., 2004)*

Typical ACA architecture



DOM-Integrated Virtual Agents (DIVA) *(Xuetao et al., 2009)*

Typical ACA architecture issues

Issue: Lack of human-likeness and dialogue naturalness

1. Repetition of cooperation: the agent is always responsive
2. Repetition of answer's schemes: use of similar linguistic patterns
3. Repetition of rational reactions: independently from previously asked requests

Solution: a modified architecture

- a personality model integrated tp the model of assistance $\mathscr{M}$
- a correlated behavioral engine $\mathscr{E}_b$ working with $\mathscr{E}_r$

Typical ACA architecture issues

Issue: Lack of human-likeness and dialogue naturalness

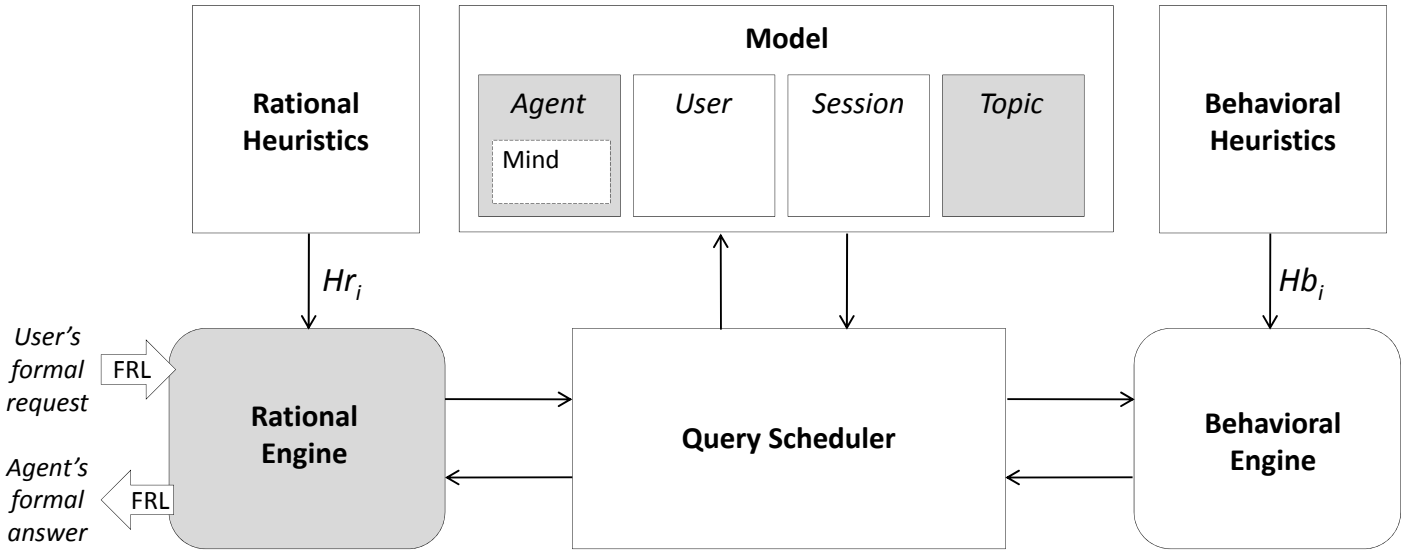
1. Repetition of cooperation: the agent is always responsive
2. Repetition of answer's schemes: use of similar linguistic patterns
3. Repetition of rational reactions: independently from previously asked requests

Solution: a modified architecture

- a personality model integrated tp the model of assistance $\mathscr{M}$
- a correlated behavioral engine $\mathscr{E}_b$ working with $\mathscr{E}_r$
  $\Rightarrow$ **Relationship between $\mathscr{E}_r$ and $\mathscr{E}_b$?**
  To define through experimentation with Rational and Behavioral (R&B) agents

Outline

General R&B agent architecture

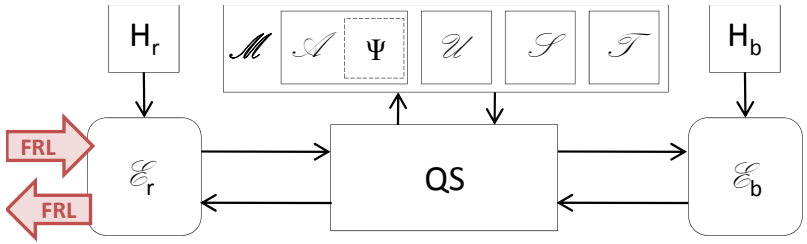


- Detailed model of assistance $\mathscr{M}$ (including agent's mind)
- Separated heuristics as symbolic representation
- Behavioral Engine $\mathscr{E}_b$
- A query scheduler

Language structure

FRL supports I/O between the user (u) and the agent (a) through the interface

Form: `PERFORMATIVE[content]`



Content

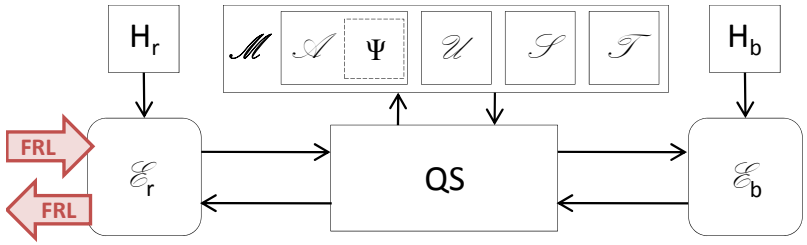
- Reference (R): element of the model $\mathscr{M}$
- Action (A): operation executable in the environment
- Proposition (P): logical proposition regarding the state of $\mathscr{M}$
- Value (V): value of an element of the model $\mathscr{M}$

Performatives

Language structure

FRL supports I/O between the user (u) and the agent (a) through the interface

Form: PERFORMATIVE[content]



Content

Performatives

- Knowledge: $\texttt{ASK}_u$[R|A|P], $\texttt{HOW}_u$[A], ...
- Actions management: $\texttt{SUGGEST}_a$[A|P], ...
- Feeling expression: $\texttt{FEEL}_u$[P], $\texttt{LIKE}_a$[R|A|P|V], ...
- Dialogue: $\texttt{AGREE}_u$[P], $\texttt{BYE}_u$[], ...

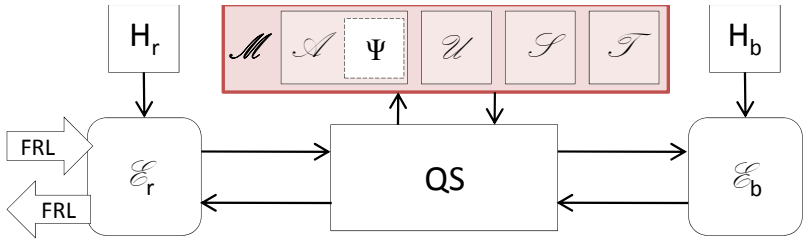
Syntax

Tree structure

- Non-terminal nodes: concepts
- Terminal nodes: symbols, numbers, booleans, strings

Skeleton of the $\mathscr{M}$ ontology

```
Model = Rootconcept[
    Concept1[
        Concept11[...],
        Concept12[...],
    ...]
    Concept2[
        Concept21[...],
        Concept22[...]
    ...],
...]
```



5 domains in the model $\mathscr{M}$:

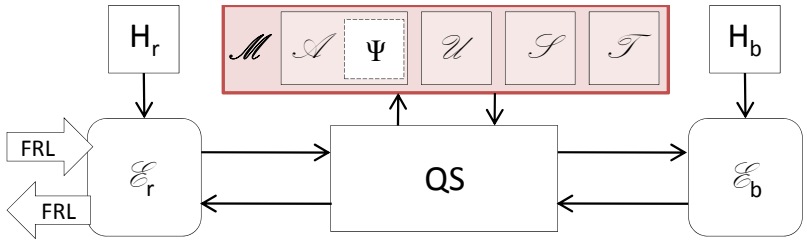
1. The agent ($\mathscr{A}$)
2. The user ($\mathscr{U}$)
3. The request ($\mathscr{R}$)
4. The session ($\mathscr{S}$)
5. The topic ($\mathscr{T}$)

$\mathscr{M} =< \mathscr{A}, \mathscr{U}, \mathscr{R}, \mathscr{S}, \mathscr{T} >$

Dynamics

$\mathscr{M}_0 =< \mathscr{A}_0, \varnothing, \varnothing, \varnothing, \mathscr{T}_0 >$

- Agent updates $\mathscr{A}$, $\mathscr{U}$, $\mathscr{R}$ and $\mathscr{S}$ according to interactions
- Application updates $\mathscr{T}$



Model Query Language (MQL)

`GET[path]`	return subtrees
`SET[path,expr]`	replaces subtree by expression
`ADD[path,expr]`	appends expression to the subtree
`DEL[path,subtree]`	deletes a subtree

...

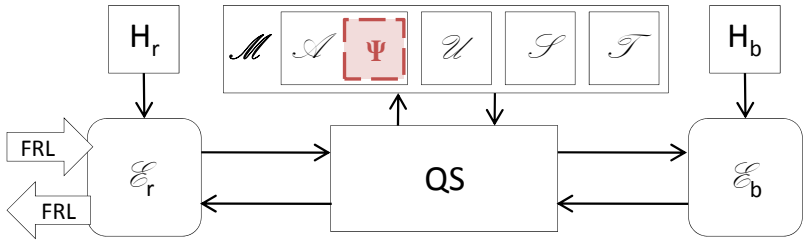
Example of path: $\mathscr{M}.\mathscr{A}$.name

A query object Q_i wraps queries

Q_i^+/Q_i^- stands for a successful/failed request

Four mental states

	Unary	Binary
Static	Trait Ψ_T	Role Ψ_R
Dynamic	Mood Ψ_m	Affect Ψ_a



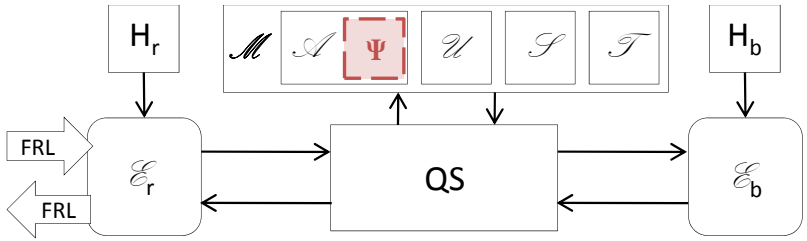
Values

In $[-1,1]$ but we distinguish 5 intervals:

$v \in [-1,-0.8]$	$<$	strongly antonymic
$v \in [-0.8,-0.2]$	-	antonymic
$v \in [-0.2,0.2]$	$=$	neutral
$v \in [0.2,0.8]$	$+$	positive
$v \in [0.8,1]$	$>$	strongly positive

Four mental states

	Unary	Binary
Static	Trait Ψ_T	Role Ψ_R
Dynamic	Mood Ψ_m	Affect Ψ_a



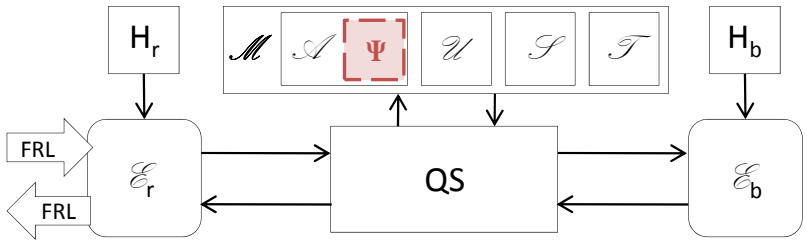
Traits Ψ_T

Classical "Big Five" *(Goldberg, 1981)* defining the personality

- *Openness*: appreciation for adventure, curiosity
- *Conscientiousness*: self-discipline and achieves goals
- *Extraversion*: strong positive emotions and sociability
- *Agreeableness*: compassion and cooperativeness
- *Neuroticism*: experience negative emotions easily

Four mental states

	Unary	Binary
Static	Trait Ψ_T	Role Ψ_R
Dynamic	Mood Ψ_m	Affect Ψ_a



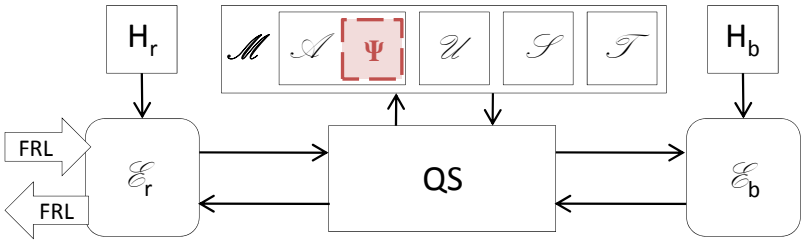
Moods Ψ_m

Personality factors changed in time by heuristics and biases

- *Energy*: physical strength
- *Happiness*: physical contentment regarding the situation
- *Confidence*: cognitive strength
- *Satisfaction*: cognitive contentment regarding the situation

Four mental states

	Unary	Binary
Static	Trait Ψ_T	Role Ψ_R
Dynamic	Mood Ψ_m	Affect Ψ_a



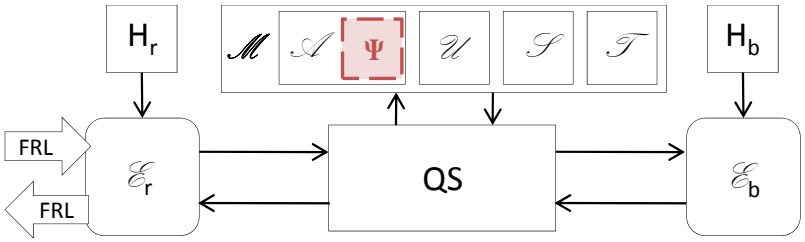
Roles Ψ_R

Static relationship between the agent and another entity of the world (*e.g.* users)

- *Authority*: right to be directive to X and reciprocally to not accept directive behaviors from X.
  Antisymmetric: `Authority(X,Y) = -Authority(Y,X)`
- *Familiarity*: right to use informal behaviors towards X.
  Symmetric: `Familiarity(X,Y) = Familiarity(Y,X)`

Four mental states

	Unary	Binary
Static	Trait Ψ_T	Role Ψ_R
Dynamic	Mood Ψ_m	Affect Ψ_a



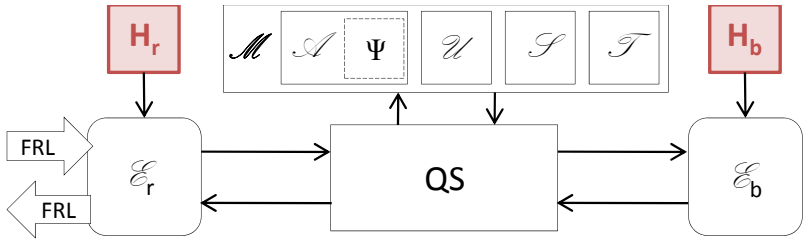
Affects Ψ_a

Dynamic relationships between the agent and another entity

- *Dominance*: power felt towards X.
  Antisymmetric: `Dominance(X,Y) = -Dominance(Y,X)`
- *Affection*: attraction and tendency to be nice to X.
  Not necessarily symmetric.
- *Trust*: feeling one can rely on X.
  Not necessarily symmetric.

Heuristic Description Language

H : identifier[FRL-pattern] :⟶
{ GuardedScript$_1$, . . . ,
GuardedScript$_n$}

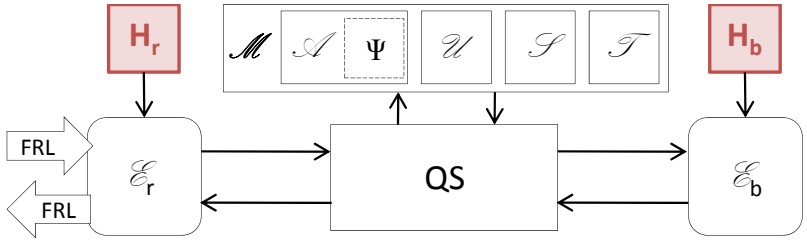


Simple rational reaction: "How old are you?"

H_r : ask-agent-attribute[$\texttt{ASK}_u$ [agent.x_]]:⟶ {
⟶ Q[i, GET[x_]],
Q_i^+ ⟶ Q[j, SET[$\mathscr{R}$.reply, $\texttt{TELL}_a$ [agent.x_, Q_i.value]]],
Q_i^- ⟶ Q[j, SET[$\mathscr{R}$.reply, $\texttt{TELL}_a$ [Q_i.failure]]]
}

Heuristic Description Language

H : identifier[FRL-pattern] :⟶
{ GuardedScript$_1$, . . . ,
GuardedScript$_n$}



Simple behavioral reaction: "I don't like you"

H_b : dislike-agent[DISLIKE$_u$[agent]]:⟶ {
⟶ {Q[i, MAP[energy, $\lambda\ x.x*0.9$]],
Q[j, MAP[confidence, $\lambda\ x.x*0.9$]]
}
$Q_i^+ \wedge Q_i$.val $< -0.5 \longrightarrow$ ADD[$\mathscr{R}$.reply,
TELL$_a$[energy, "tired"]]
$Q_j^+ \wedge Q_j$.val $< -0.5 \longrightarrow$ ADD[$\mathscr{R}$.reply,
TELL$_a$[confidence, "depressed"]]
}

Outline

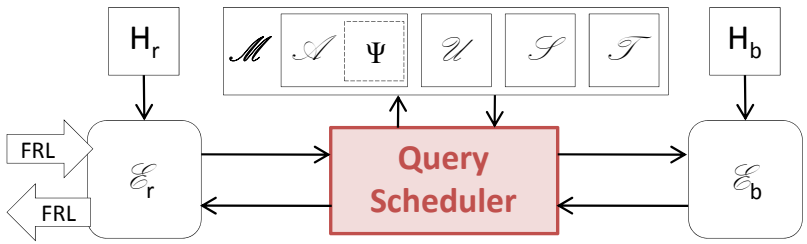
The CB scenario

Reminder: our objective is to be able to test the functioning of $\mathscr{E}_r$ and $\mathscr{E}_b$ together

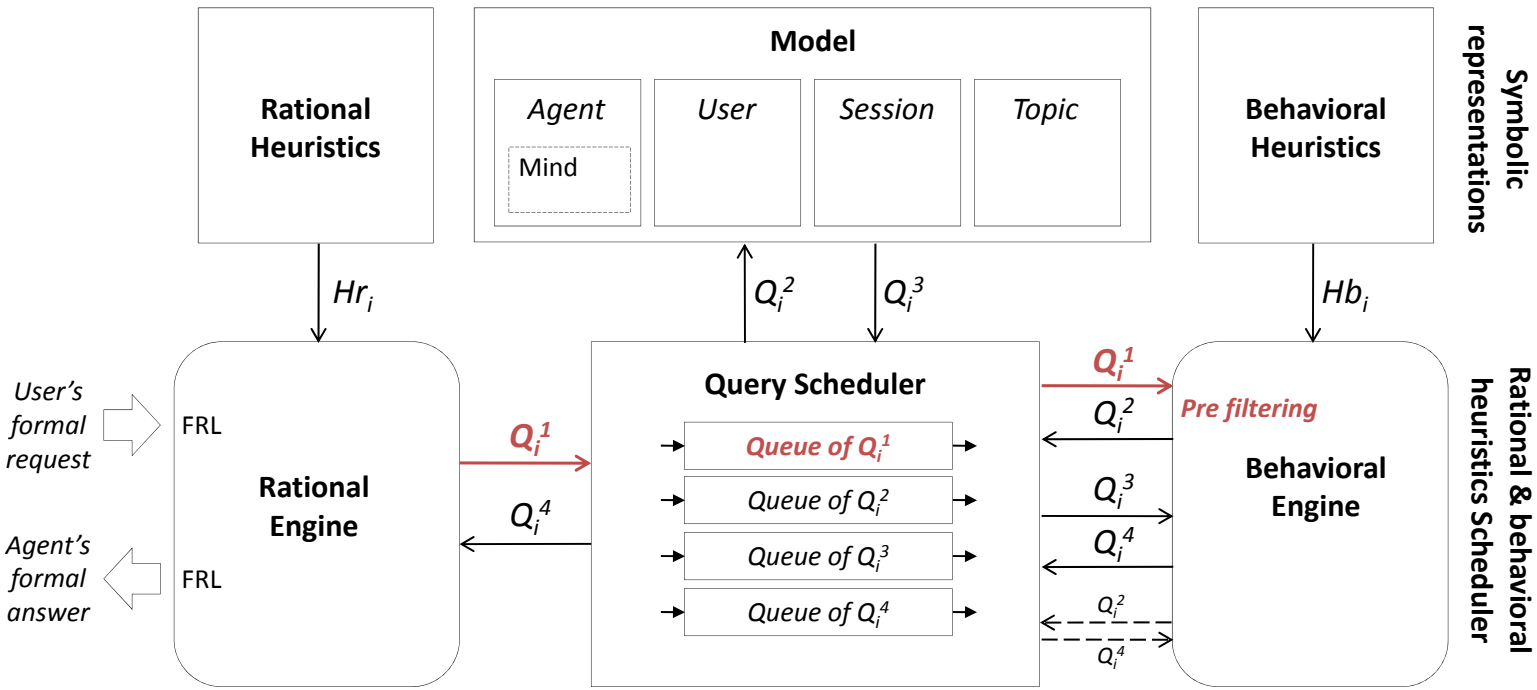
Hypotheses

1. $\mathscr{E}_r$ and $\mathscr{E}_b$ are constructed independently
2. $\mathscr{E}_r$ and $\mathscr{E}_b$ work independently

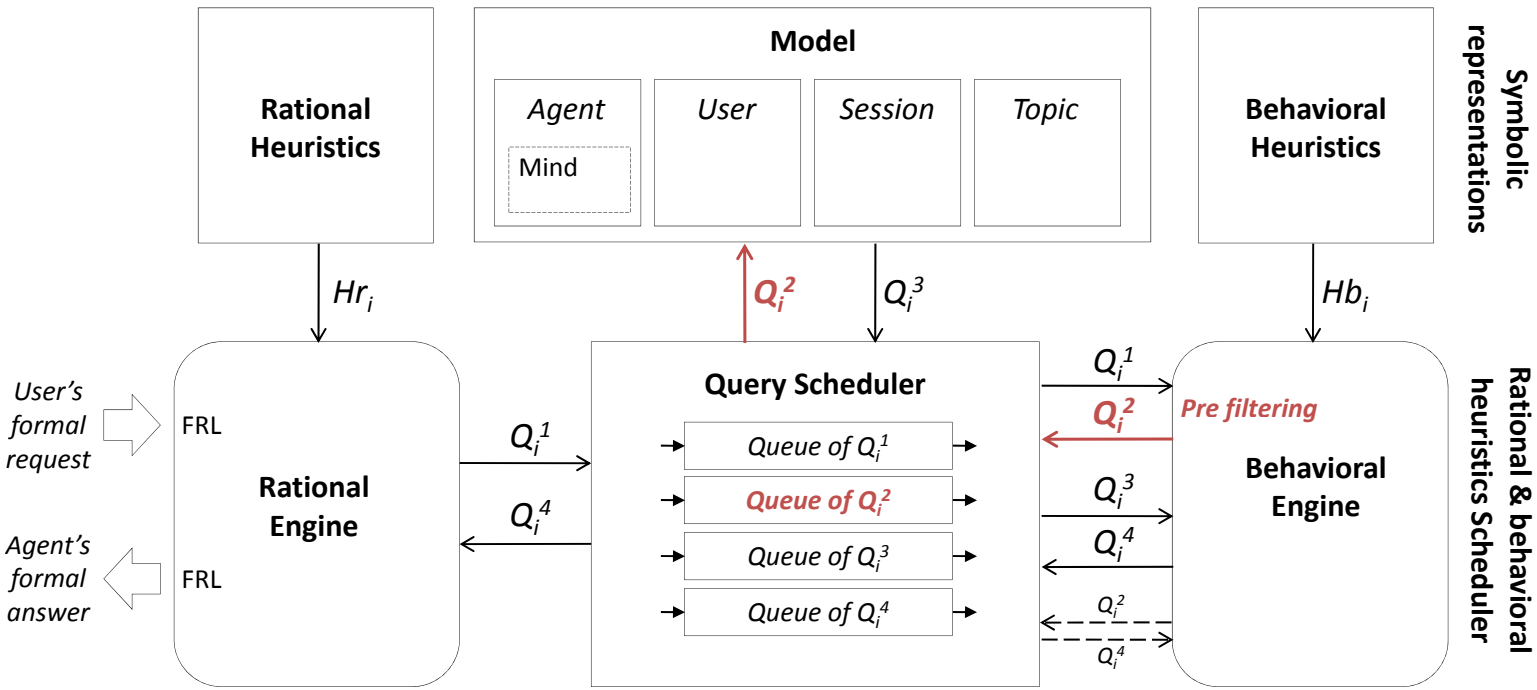
Practically, the Query Scheduler (QS) let $\mathscr{E}_b$ prefilter and postfilter requests to and from $\mathscr{E}_r$



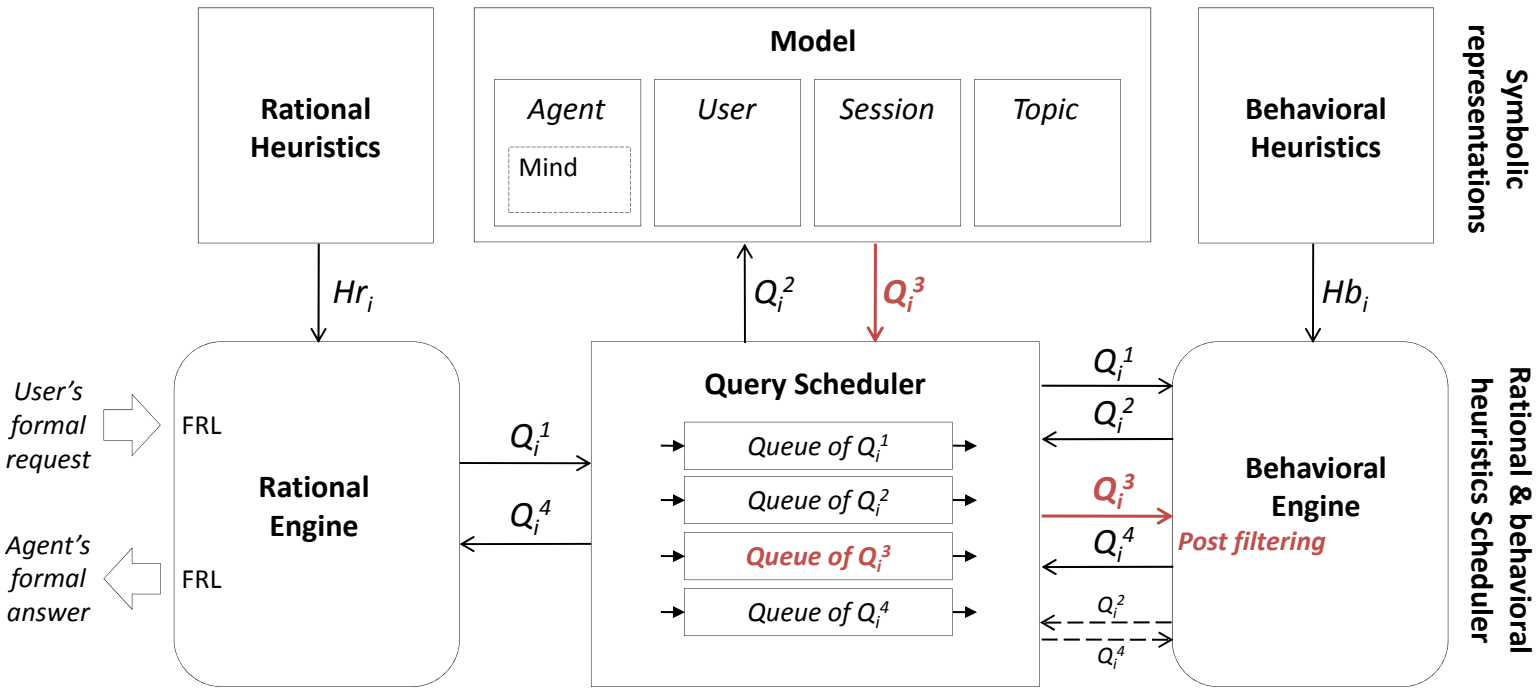
The CB dynamic functionning



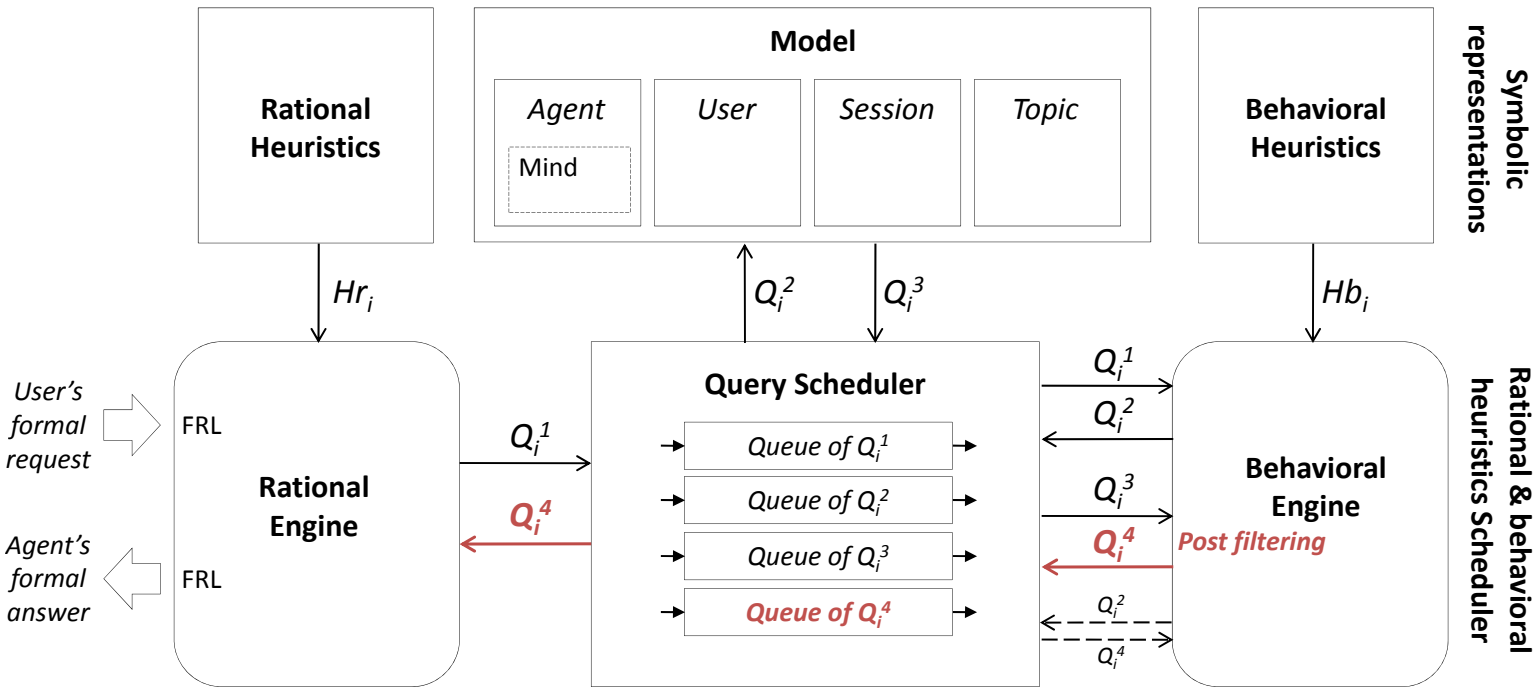
The CB dynamic functionning



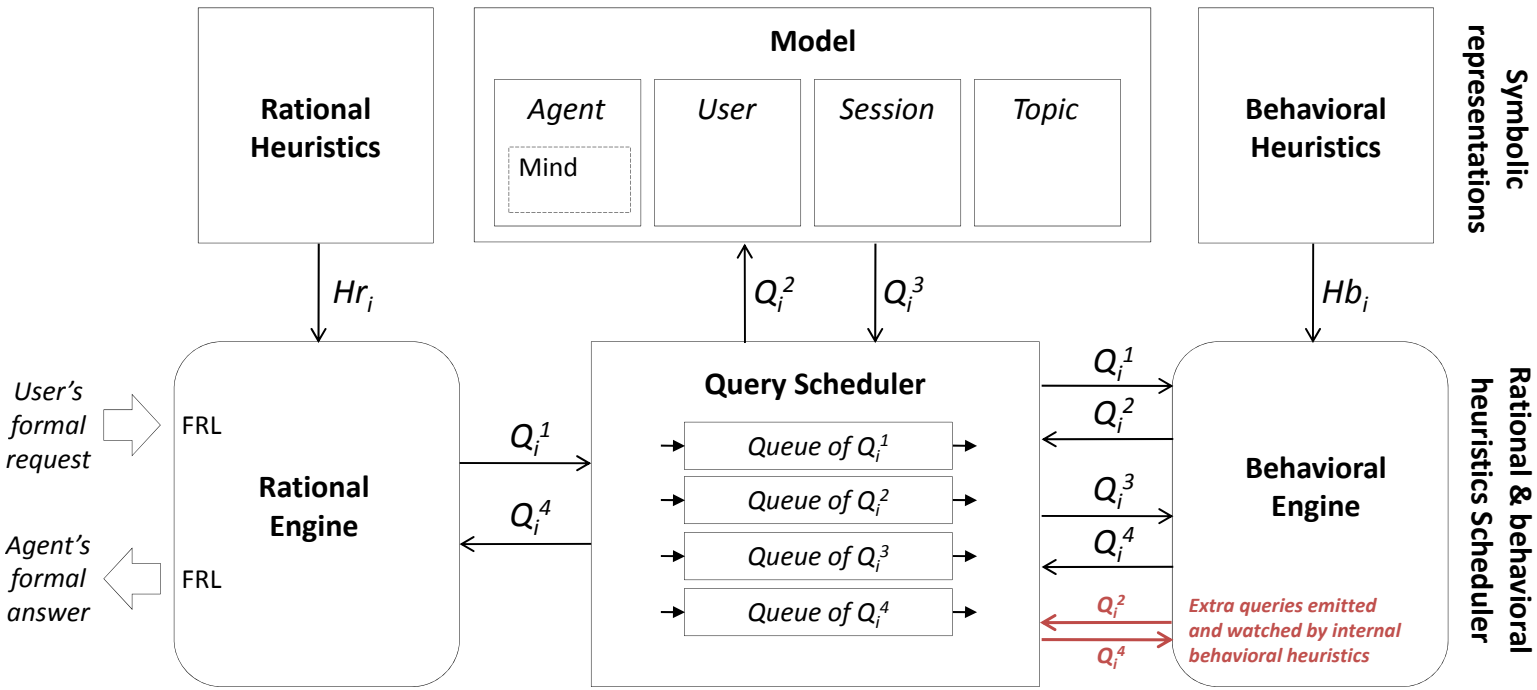
The CB dynamic functionning



The CB dynamic functionning



The CB dynamic functionning



Biased behavior example

A neurotic unhappy agent perceives extra negativity in what the user is saying

Extra negative perception: "I like the color of the title"

H_b : see-life-in-black[Q^1[i_, SET[x_.WorthForUser, 'high']]]:⟶ {

$M_h^< \wedge T_n^{<-=}$ ⟶ {

Q^2[i, SET[x_.WorthForUser, 'high'],

ADD[$\mathscr{R}$.reply, REQUEST$_a$[JUSTIFY$_u$[$\mathscr{M}$.$\mathscr{R}$]]]

Q^2[j, MAP[confidence, λ x.x*0.8]

...},

...}

Biased behavior example

If we combine it with another one like:

Mention unhappiness

H_b : on-entering-unconfidence[Q^3[i_, MAP[confidence, f_]]]:$\rightarrow$ {
$Q_i^{+4} \wedge \backslash M_c^-$ $\rightarrow$ ADD[$\mathscr{R}$.reply, TELL$_a$[$\backslash M_c^-$]]
...}

We can get the answer: "You're making me depressed"

Outline

Conclusion

Summary

- Assistance requires rational reasoning over the task **and** behavioral reasoning about the dialogical session
- The relationship between both is an open issue
- The R&B framework proposed here allows implementation of flexible relationship between rationality and behaviors – for instance, the Cognitive Bias hypothesis

Perspectives

1. Implementing the R&B model (in Mathematica)
2. Combining it with the DIVA toolkit to deploy on the web
3. Experimenting different behaviors with real users

Related works in cognitive models

- CoJACK: addition of human physiological constraints to JACK agents *(Evertsz et al., 2008)*
- Use of degrees in multivalued logics for BDI agents *(Casali et al., 2004)*
- Extra parameters to BDI architectures: to help to model emotional behaviors – fundamental desires, capacities, resources... *(Pereira et al., 2008)*
- Heuristics order: impacts the perception of high level personnality traits *(Dastani, 2002)*