Mathematical Domain Reasoning Tasks in Natural Language Tutorial Dialog on Proofs

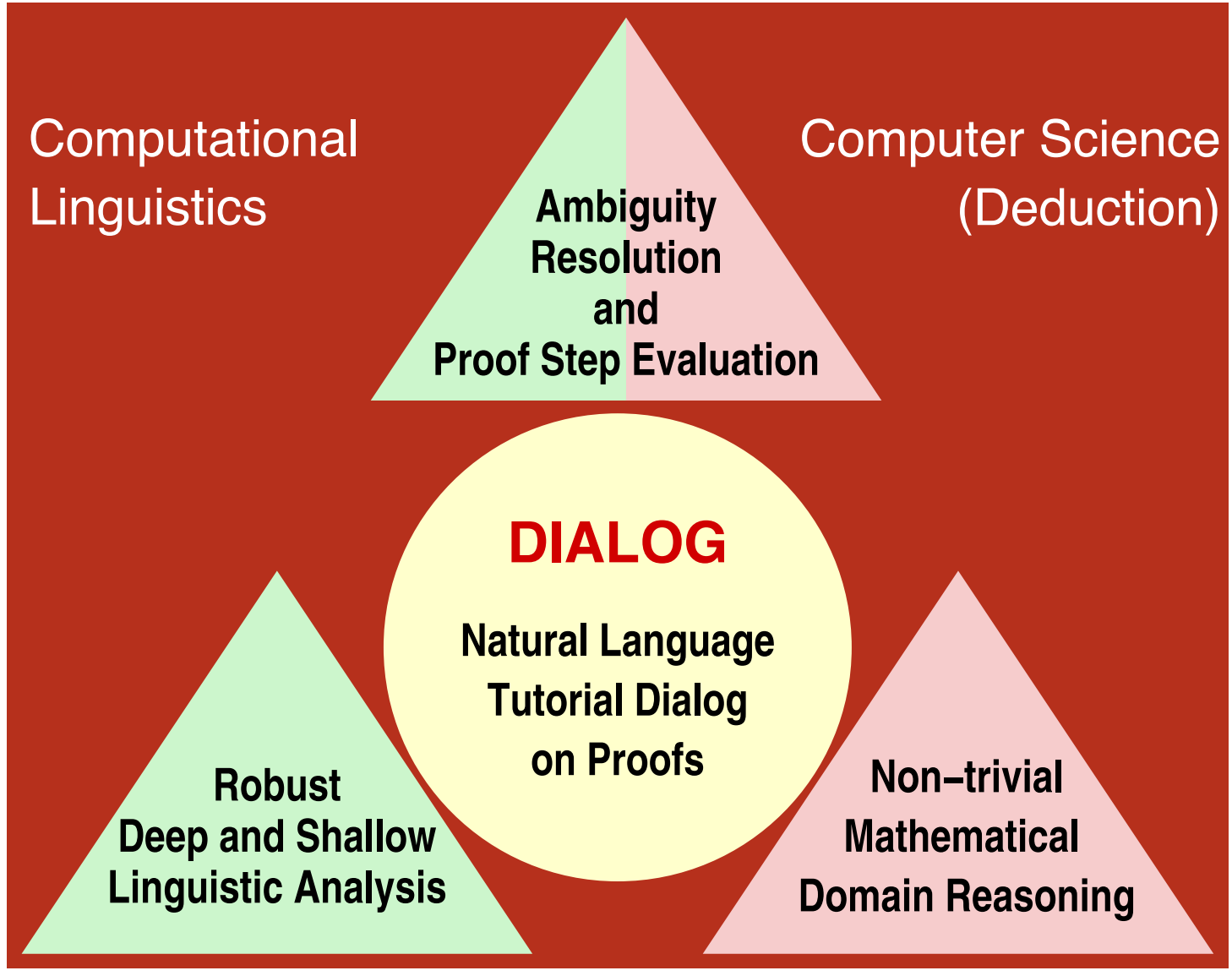
Christoph Benzmüller

Joint work with: Quoc Bao Vo (and the SFB378 DIALOG Project)

Computer Science & Comput. Ling.
Saarland University
Saarbrücken, Germany
http://www.ags.uni-sb.de/~chris/dialog/

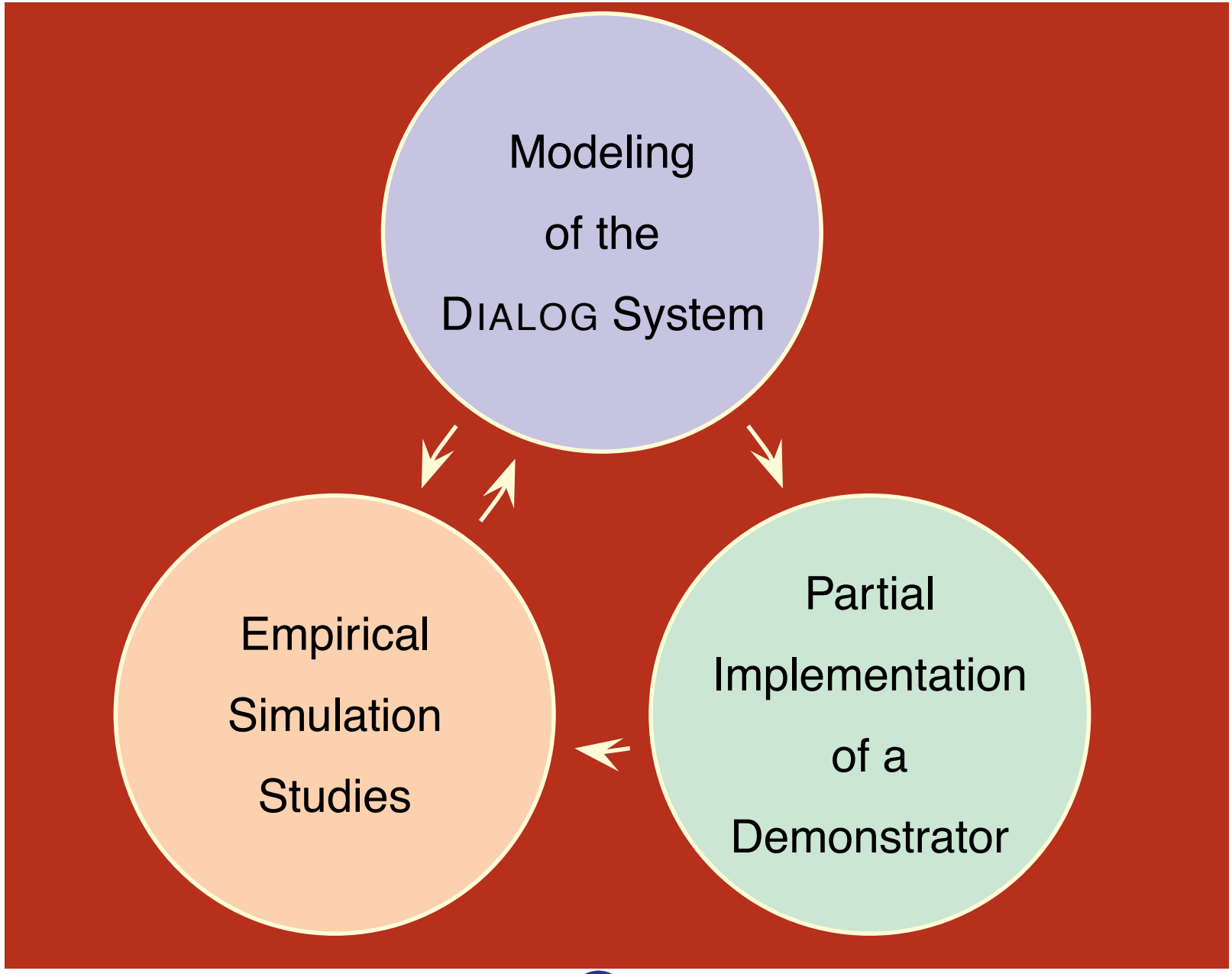
AAAI-05, July 11th, Pittsburgh, USA

The DIALOG Project in the SFB 378



Method: Progressive Refinement

DIALOG Project



WOZ-Experiment → Own Corpus

DIALOG Project

Subject (Student) Room:

Audio Recording

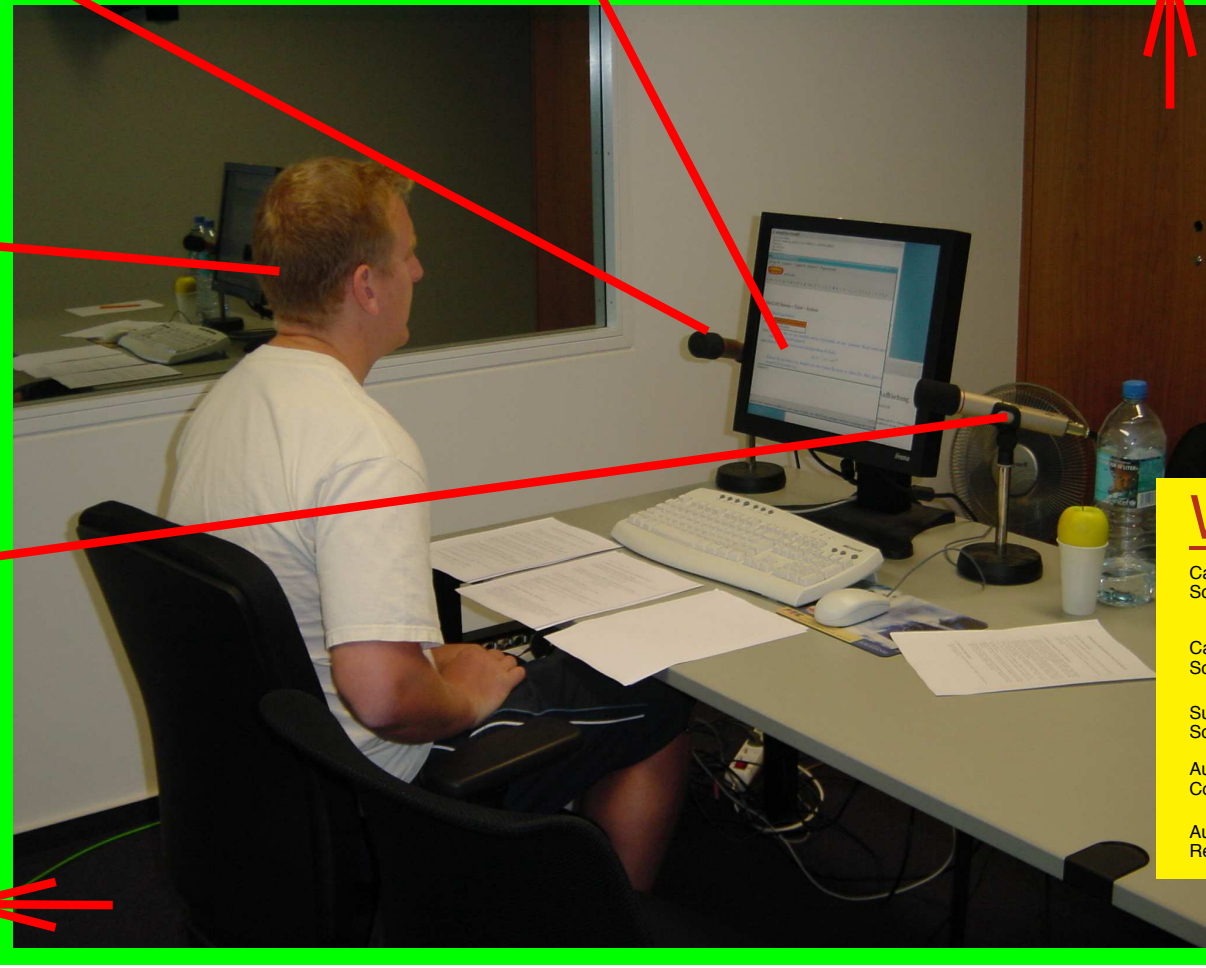
Subject GUI

Cam1

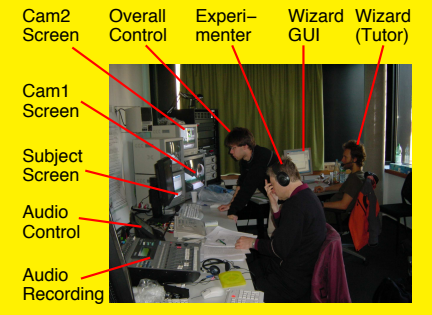
Subject

Audio Control

Cam2



Wizard Room:



WOZ-Experiment → Own Corpus

DIALOG Project

Wizard (Tutor) Room:

Cam2 Screen

Overall Control

Experi–menter

Wizard GUI

Wizard (Tutor)

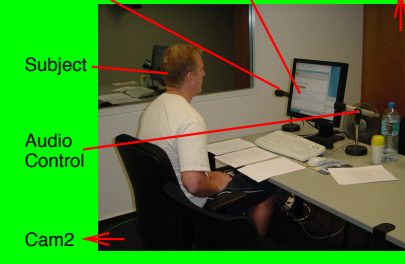
Cam1 Screen

Subject Screen

Audio Control

Audio Recording

Subject Room:



Corpus Example

DIALOG Project

T1: Bitte zeigen Sie: $K((A \cup B) \cap (C \cup D)) = (K(A) \cap K(B)) \cup (K(C) \cap K(D))$!
[Please show: $K((A \cup B) \cap (C \cup D)) = (K(A) \cap K(B)) \cup (K(C) \cap K(D))$!]

S1: nach deMorgan-Regel-2 ist $K((A \cup B) \cap (C \cup D)) = (K(A \cup B) \cup K(C \cup D))$.
[by deMorgan-Rule-2 $K((A \cup B) \cap (C \cup D)) = (K(A \cup B) \cup K(C \cup D))$ holds.]

T2: Das ist richtig!
[This is correct!]

S2: $K(A \cup B)$ ist laut deMorgan-1 $K(A) \cap K(B)$
[$K(A \cup B)$ is $K(A) \cap K(B)$ according to deMorgan-1]

T3: Das stimmt auch.
[That is also right.]

S3: und $K(C \cup D)$ ist ebenfalls laut deMorgan-1 $K(C) \cap K(D)$
[and $K(C \cup D)$ is also $K(C) \cap K(D)$ according to deMorgan-1]

. . .

Get corpus: `http://www.ags.uni-sb.de/~chris/dialog/`
Total figures 1. exp.: 66 dialogs / av. 12 turns / 1115 sentences

Research Challenges

DIALOG Project

Perspective of Mathematical Domain Reasoning (MDR):

- Support for resolution of Ambiguities and Underspecification
- Proof Step Evaluation
  - Soundness: proof step verifiable by formal system?
  - Granularity: size/argumentative complexity of proof step?
  - Relevance: proof step needed/useful in achieving the goal?

Perspective of NL Analysis: [. . . not in this talk . . .]

Perspective of Dialog Management: [. . . not in this talk . . .]

Perspective of Tutoring Proofs: [. . . not in this talk . . .]

Research Challenges

DIALOG Project

Perspective of Mathematical Domain Reasoning (MDR):

- Support for resolution of Ambiguities and Underspecification
- Proof Step Evaluation
  - Soundness: proof verifiable by formal system?
  - Granularity: argumentative complexity of proof step?
  - Relevance: proof step needed/useful in achieving the goal?

Logical vs Tutorial Dimension

Perspective of NL Analysis: [. . . not in this talk . . .]

Perspective of Dialog Management: [. . . not in this talk . . .]

Perspective of Tutoring Proofs: [. . . not in this talk . . .]

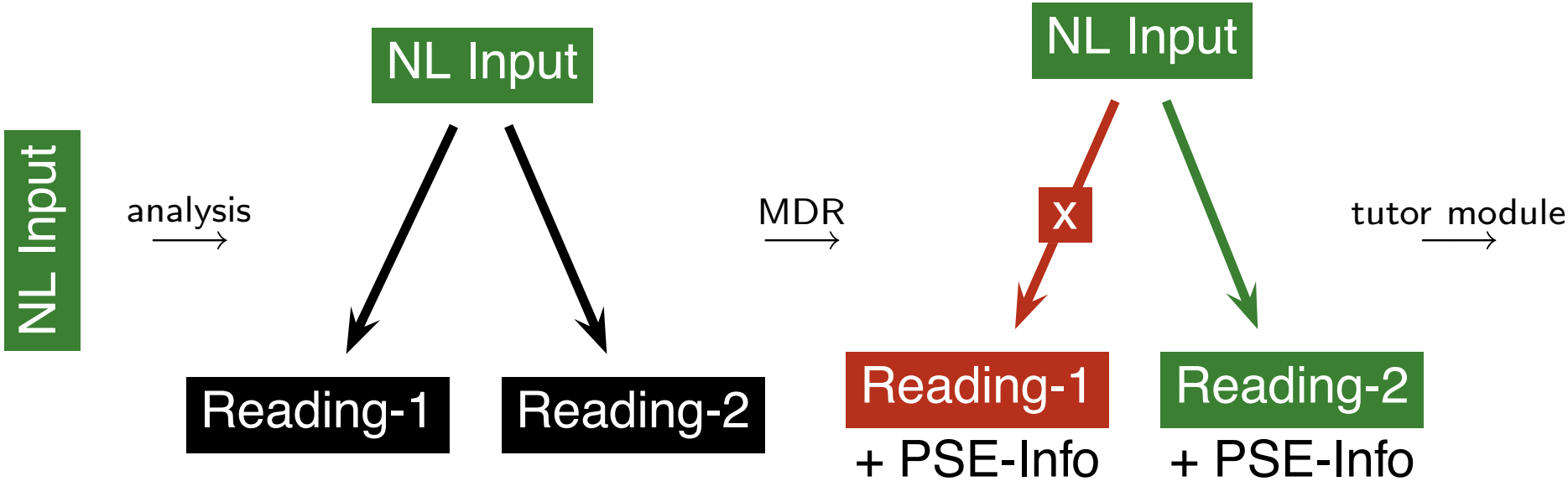
Mathematical Domain Reasoning

DIALOG Project

— declarative abstract level sketches →

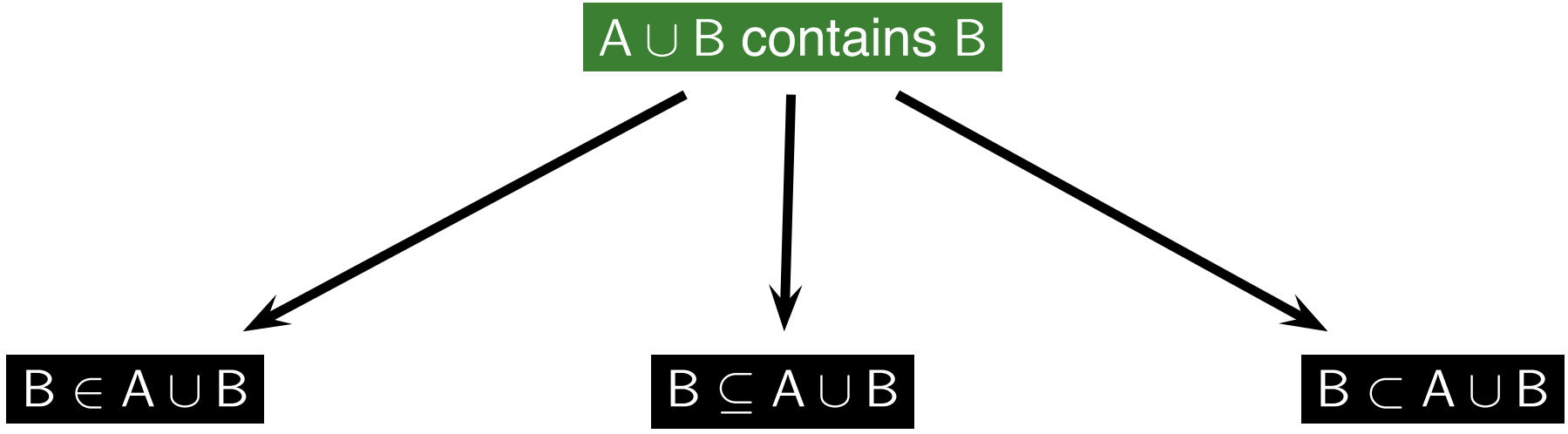
Communication Gap

← procedural calculus level proofs —



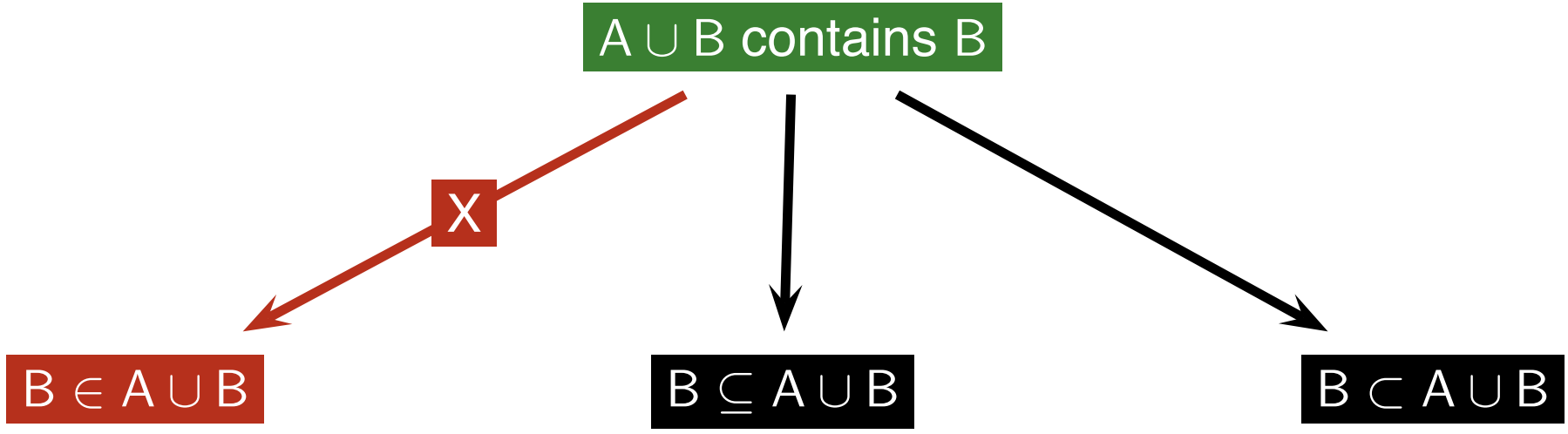
Mathematical Domain Reasoning

DIALOG Project



Mathematical Domain Reasoning

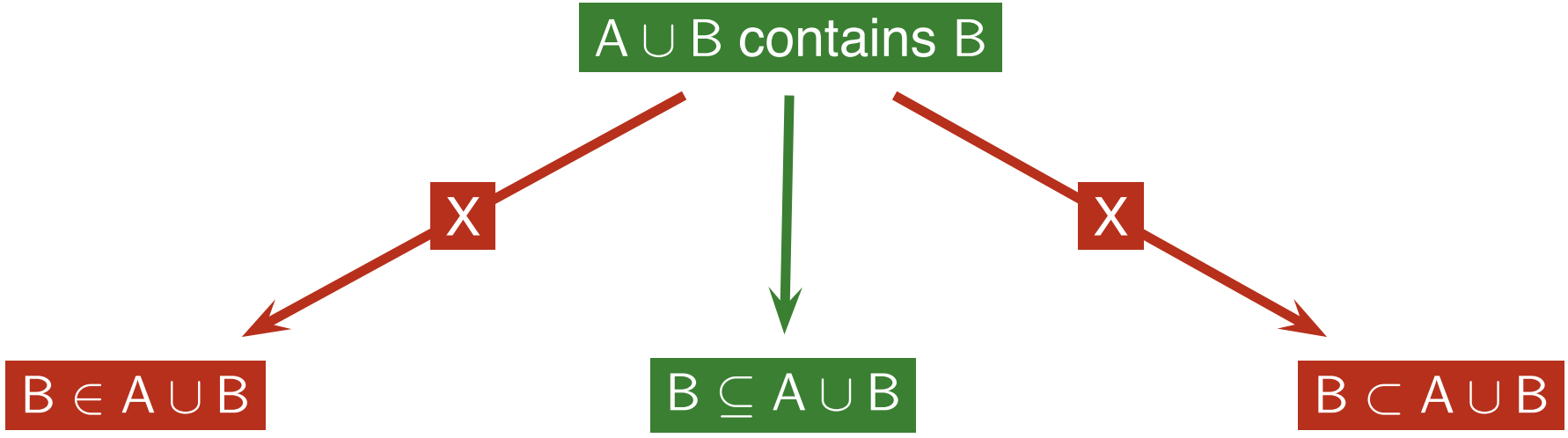
DIALOG Project



type checking

Mathematical Domain Reasoning

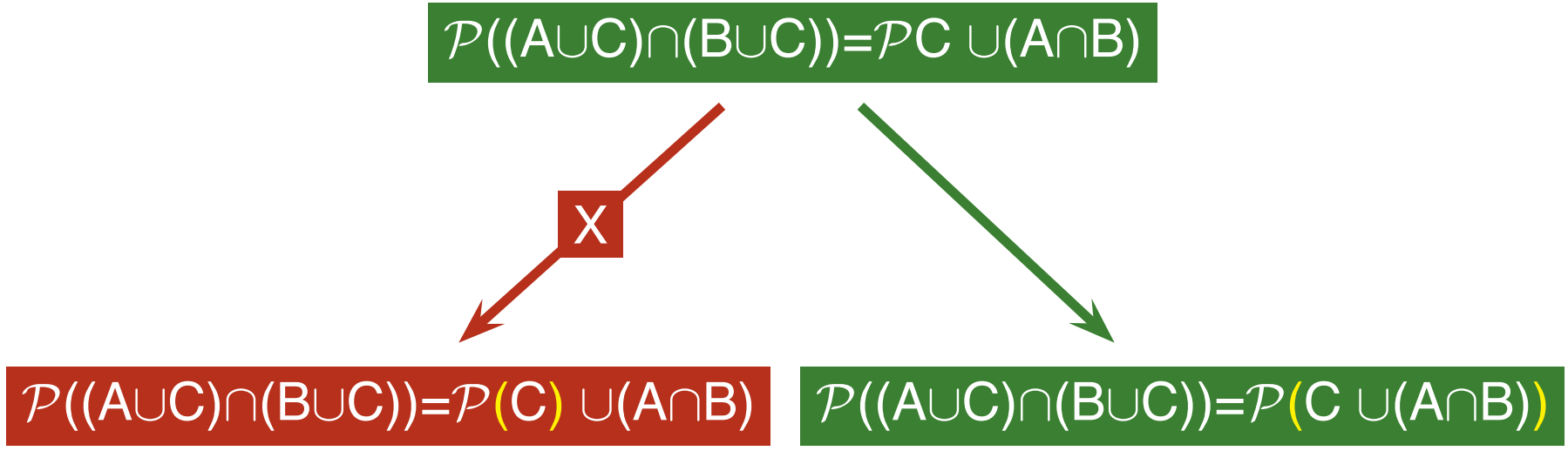
DIALOG Project



theorem proving

Mathematical Domain Reasoning

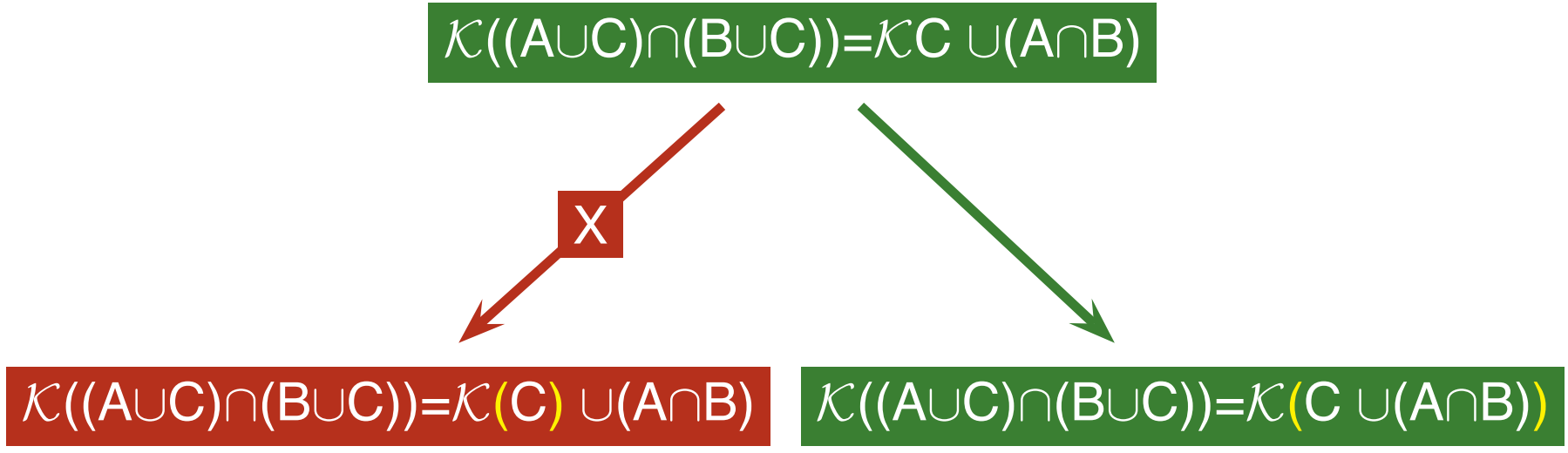
DIALOG Project



type checking

Mathematical Domain Reasoning

DIALOG Project



theorem proving

Proof Step Evaluation

DIALOG Project

(DM-1) . . .
(DM-2) . . .
?
(G) . . .

Given: (DM-1) $\overline{X \cup Y} = \overline{X} \cap \overline{Y}$

(DM-2) $\overline{X \cap Y} = \overline{X} \cup \overline{Y}$

Task: Please show $\overline{(A \cup B) \cap (C \cup D)} = (\overline{A} \cap \overline{B}) \cup (\overline{C} \cap \overline{D})$

New: By deMorgan $\overline{(A \cup B) \cap (C \cup D)} = \overline{(A \cup B)} \cup \overline{(C \cup D)}$.

Proof Step Evaluation

DIALOG Project

(DM-1) . . .
(DM-2) . . .
?
(G) . . .

Given: (DM-1) $\overline{X \cup Y} = \overline{X} \cap \overline{Y}$
(DM-2) $\overline{X \cap Y} = \overline{X} \cup \overline{Y}$

Task: Please show $\overline{(A \cup B) \cap (C \cup D)} = (\overline{A} \cap \overline{B}) \cup (\overline{C} \cap \overline{D})$

New: By deMorgan $\overline{(A \cup B) \cap (C \cup D)} = \overline{(A \cup B)} \cup \overline{(C \cup D)}$.

(DM-1) . . .
(DM-2) . . .
(New) . . . ↓
?
(G) . . .

Soundness: yes
Granularity: 1x(DM-2)
Relevance: yes

(DM-1) . . .
(DM-2) . . .
?
(New) . . . ↑
(G) . . .

Soundness: yes
Granularity: 2x(DM-1)
Relevance: yes

Proof Step Evaluation: How?

DIALOG Project

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$
?
(G) $D \vee E$

New:

We show E.

(1) . . .
(2) . . .
(3) . . .
(4) . . .
?
(G') E
(G) . . .

PSE:

Soundness

Granularity

Relevance

Proof Step Evaluation: How?

DIALOG Project

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$
?
(G) $D \vee E$

New:

We show E.

(1) ...
(2) ...
(3) ...
(4) ...
?
(G') E
(G) ...

PSE:

Soundness

- $(G') \vdash^{?} (G)$
- any proof

Granularity

Relevance

Proof Step Evaluation: How?

DIALOG Project

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$
?
(G) $D \vee E$

New:

We show E.

(1) ...
(2) ...
(3) ...
(4) ...
?
(G') E
(G) ...

PSE:

Soundness

- $(G') \vdash^{?} (G)$
- any proof

Granularity

- size-of($(G') \vdash^{?} (G)$)
- cognitively adequate proofs

Relevance

Proof Step Evaluation: How?

DIALOG Project

Discourse:

(1) $A \wedge B$
(2) $A \Rightarrow C$
(3) $C \Rightarrow D$
(4) $F \Rightarrow B$
?
(G) $D \vee E$

New:

We show E.

(1) ...
(2) ...
(3) ...
(4) ...
?
(G') E
(G) ...

PSE:

Soundness

- $(G') \vdash^{?} (G)$
- any proof

Granularity

- size-of$((G') \vdash^{?} (G))$
- cognitively adequate proofs

Relevance

- $(1), (2), (3), (4) \vdash^{?} (G')$
- detours?, shorter proofs?

Claim

DIALOG Project

Granularity and Relevance

call for

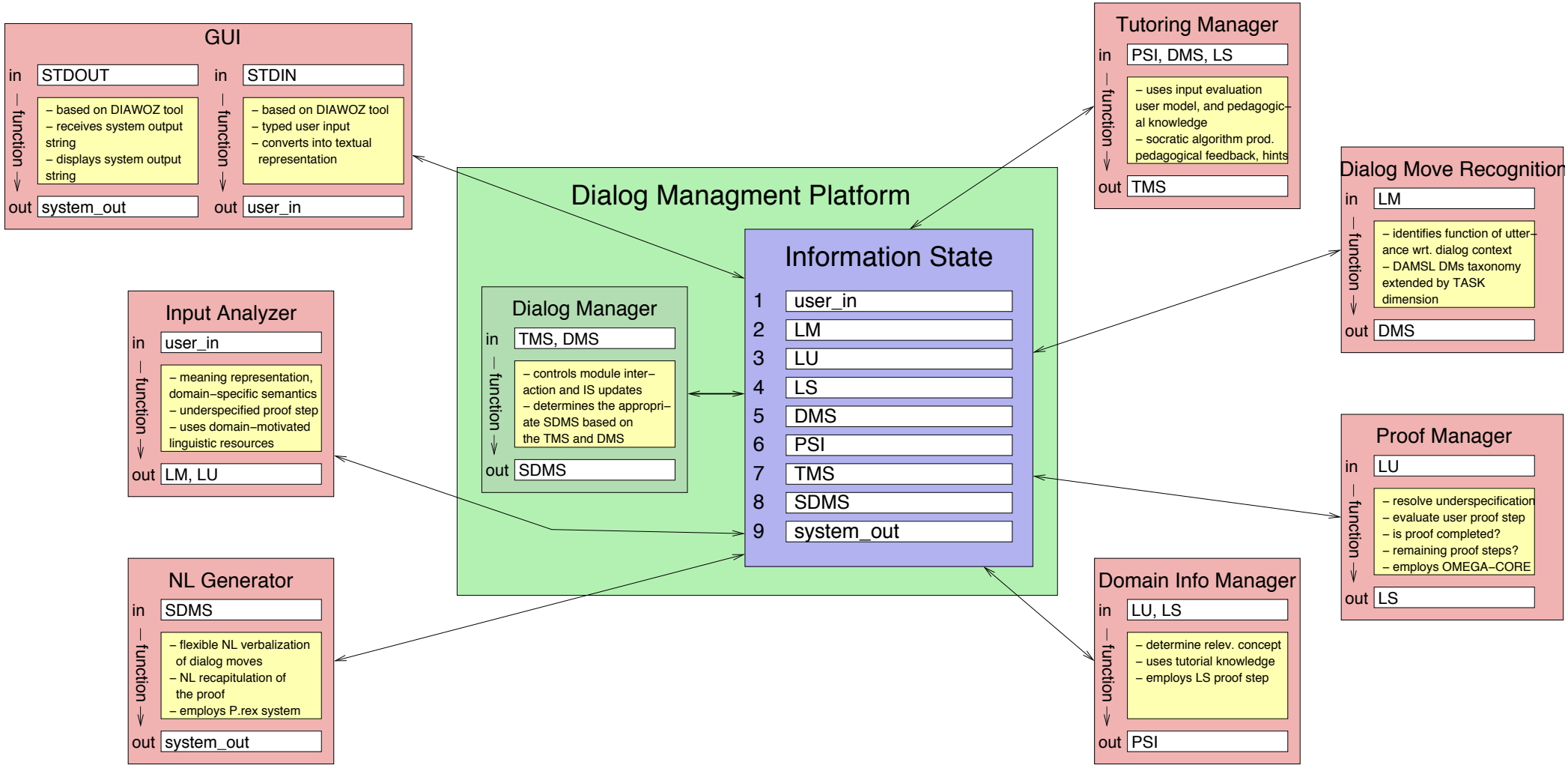
cognitively adequate abstract level proofs

+

enumeration of (some) proof alternatives

- One candidate: knowledge based proof planning [Bundy88]
- Original motivation: widen range of automatable maths
- New motivation: support for proof step evaluation

Implementation: DIALOG Demonstrator



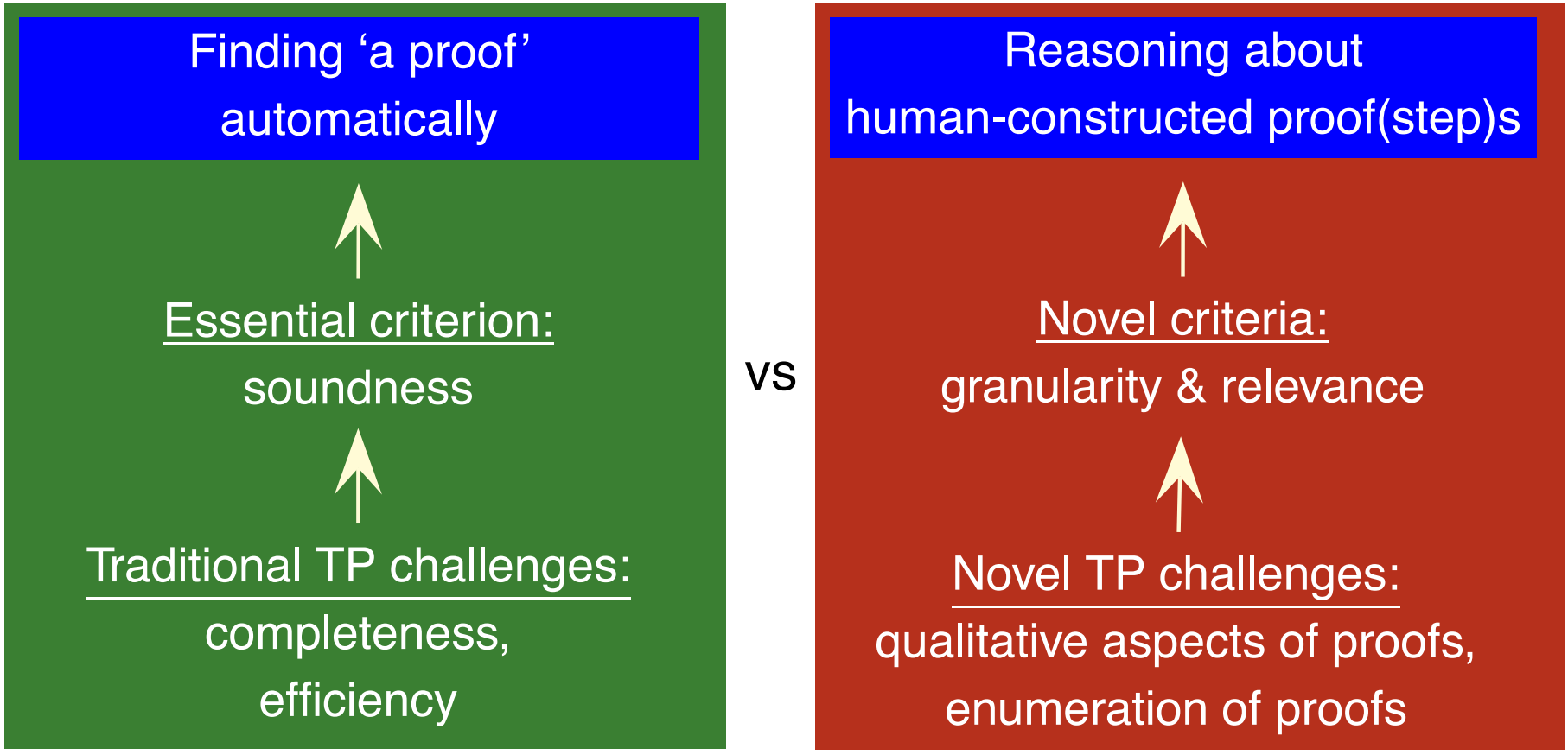
Related Work

DIALOG Project

- Motivation: [Moore93] Flexible tutorial NL dialog supports active learning
- Closest related: [Zinn04] analyzes well structured text-book proofs for soundness
- NL analysis: shallow techniques and keyword spotting probably not suitable
- MDR: Comparison against 'golden standard solutions' [GreaserEtAl00] no suitable
- Dialog modeling: Autotutor [PersonEtAl00], Geometry Tutor [MatsudaVanLehn03], Trindi and Siridus [TraumLarsson03], Beetle [Zinn03]

Conclusion

DIALOG Project



- Lots of ongoing work in all corners of the DIALOG Project