Scaling up Partially Observable Markov Decision Processes for Dialogue Management

Jason D. Williams

22 July 2005

Machine Intelligence Laboratory
Cambridge University Engineering Department

Outline

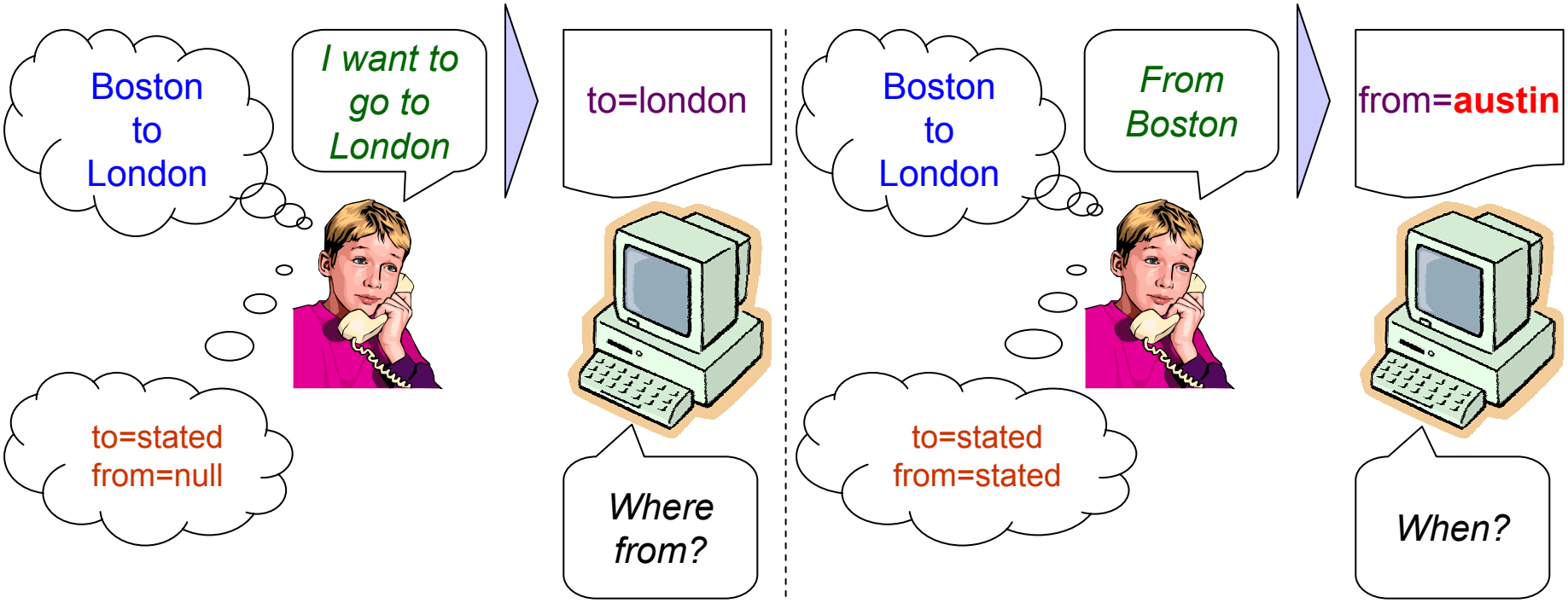
- Dialogue management as a POMDP
- Scaling up with the “summary POMDP” method
- Example problem & results
- Conclusions & future work

Material in this talk is joint work with Steve Young

Jason D. Williams, Pascal Poupart, and Steve Young (2005). *Factored Partially Observable Markov Decision Processes for Dialogue Management.* 4th Workshop on Knowledge and Reasoning in Practical Dialog Systems, International Joint Conference on Artificial Intelligence (IJCAI), August 2005, Edinburgh.

Jason D. Williams, Pascal Poupart, and Steve Young (2005). *Partially Observable Markov Decision Processes with Continuous Observations for Dialogue Management.* In Proceedings of the 6th SigDial Workshop on Discourse and Dialogue, September 2005, Lisbon.

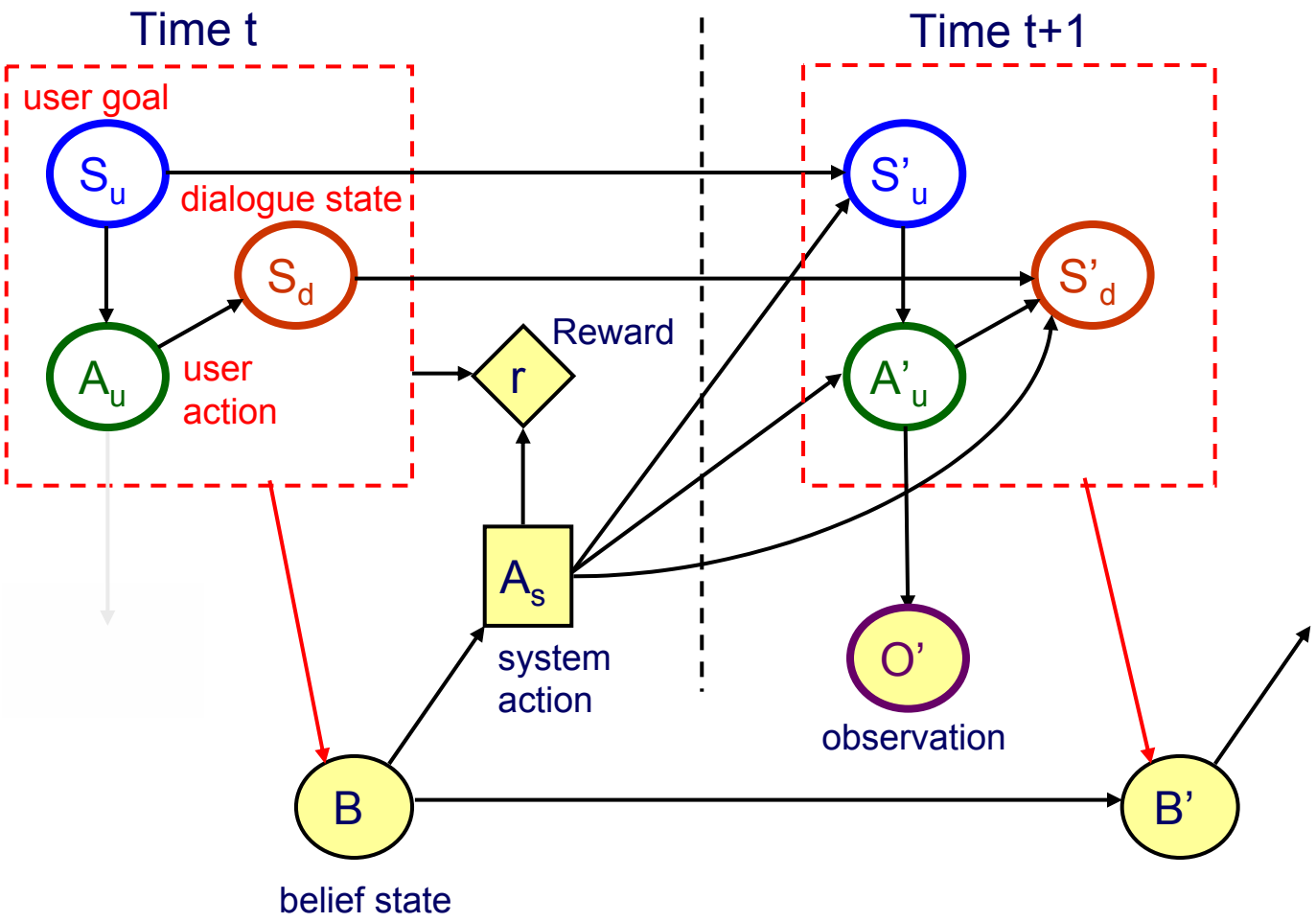
Idealized Human-Computer Dialog



- Current state = dialogue modelling
- System action selection = dialogue management

- Dialogue state is *unobserved:*
  - User's goal
  - User's (real) action
  - Conversation state
- Inferences via *observation:*
  - May contain **errors**

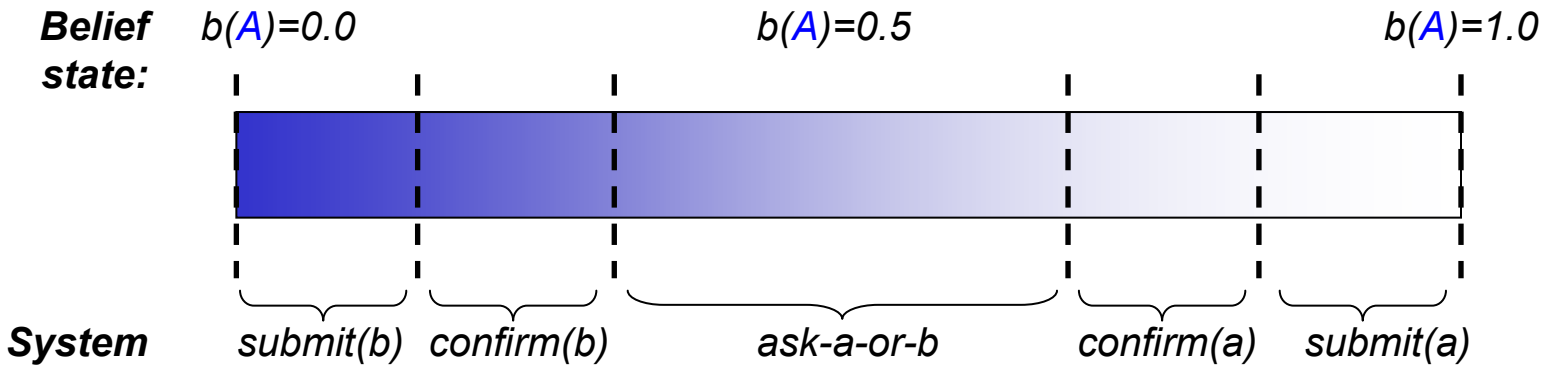
Decompose state variable into 3 “models”



Designer specifies *r;* mapping $A_s = \Pi(b(S))$ seeks to maximize cumulative reward

A policy as a partitioning

- A policy is a mapping: *situation* ➔ *action*
- One representation is a *partitioning* of belief space
- 2-dimensional example:
  - Simple state space with 2 user goals: A & B
  - Belief space can be written as b(A) (b/c b(B) = 1 - b(A))
  - Policy shows action to take for each point in belief space



$$A_s = \Pi(b(S))$$

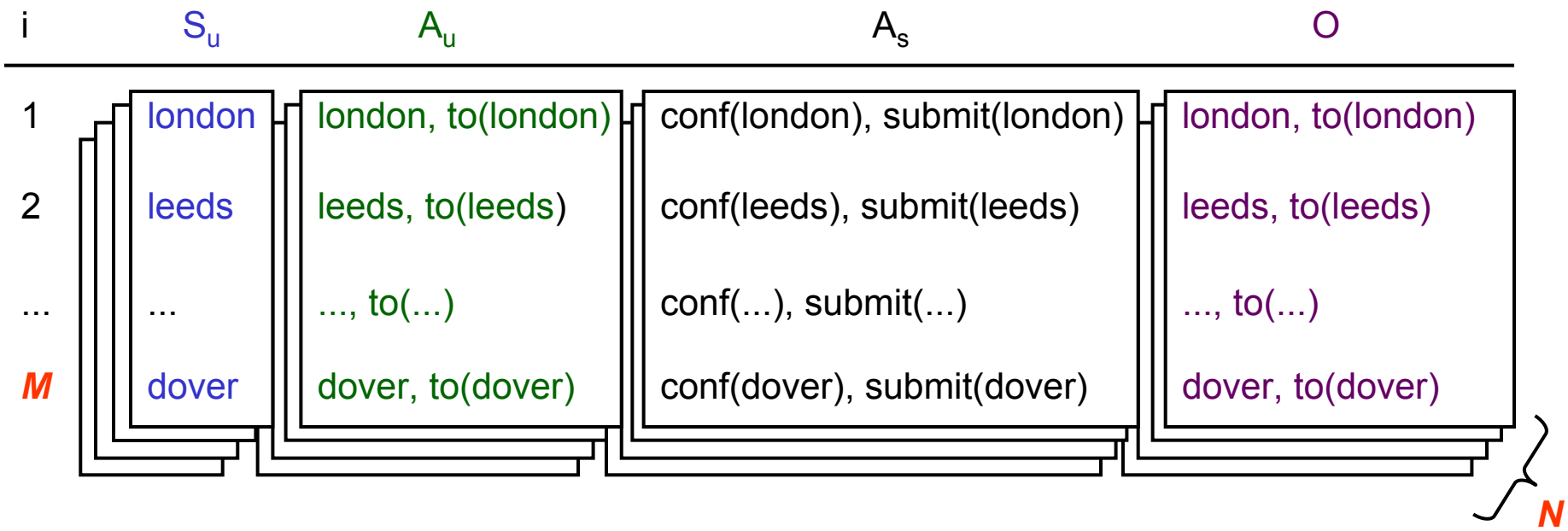
This partitioning is produced by a POMDP optimization method

Outline

- Dialogue management as a POMDP
- Scaling up with the “summary POMDP” method
- Example problem & results
- Conclusions & future work

POMDPs scale poorly

Belief space, actions, and observations cover *N* slots and *M* slot values

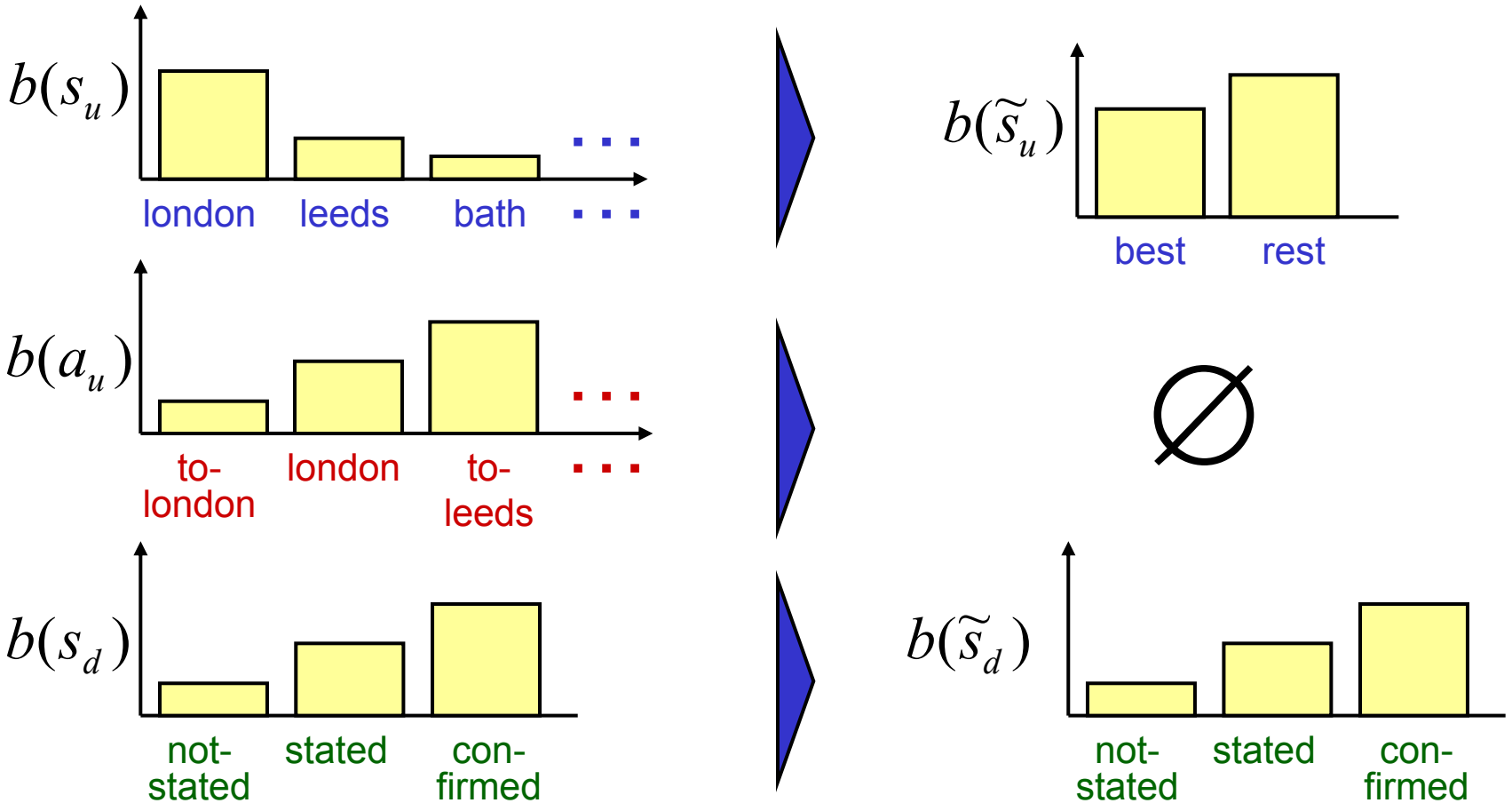


For *M*=1000 and *N*=2, there are ~10^6 states, actions, and observations

Growth factor is $O(M^N)$

Can perform belief monitoring with factoring, but how to optimize?

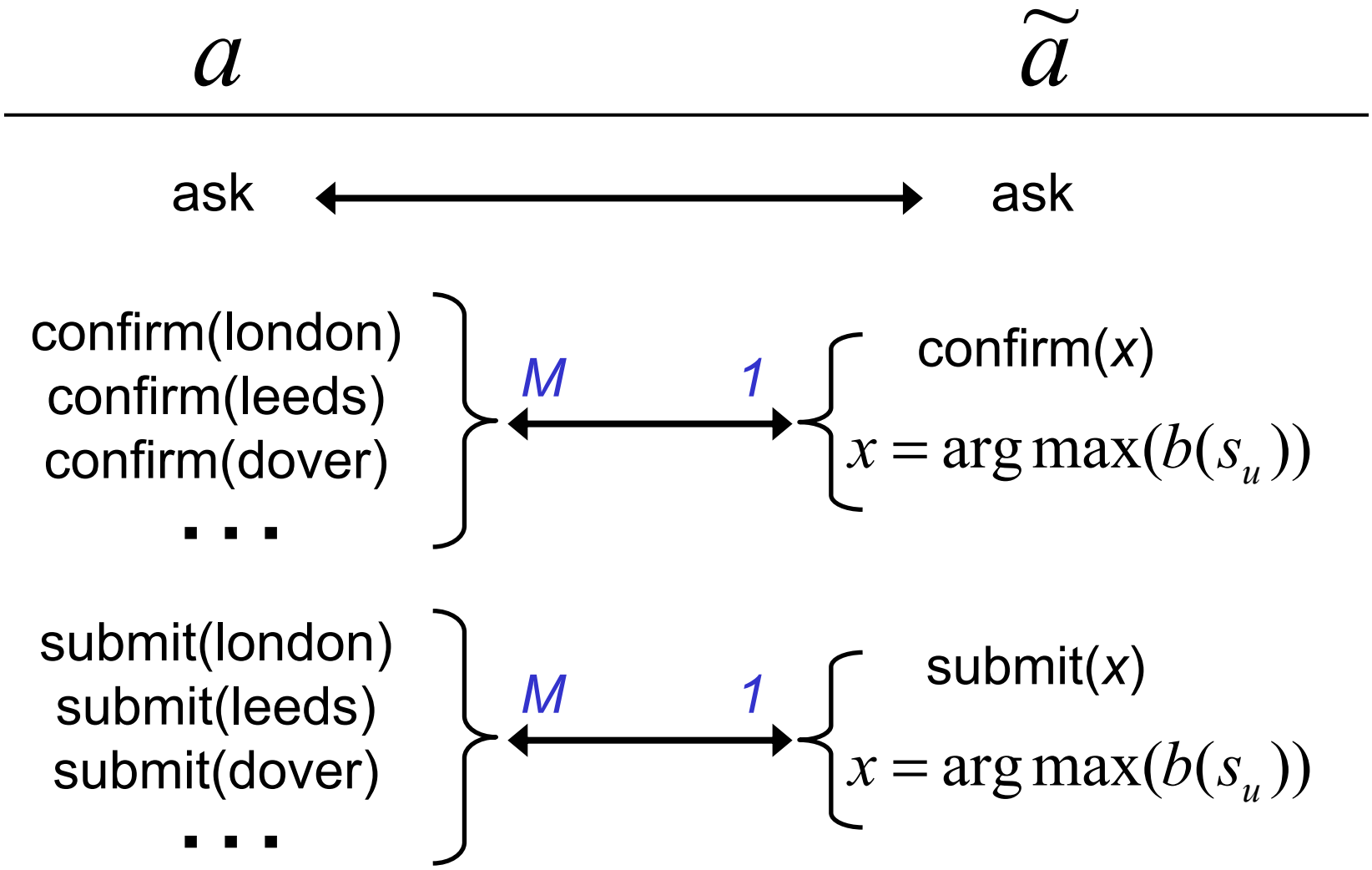
Intuition of the "Summary POMDP"



$$|S_u| \cdot |A_u| \cdot |S_d| = 1000 \cdot 2000 \cdot 3$$

$$|\tilde{S}_u| \cdot |\tilde{S}_d| = 2 \cdot 3$$

Actions in the summary POMDP

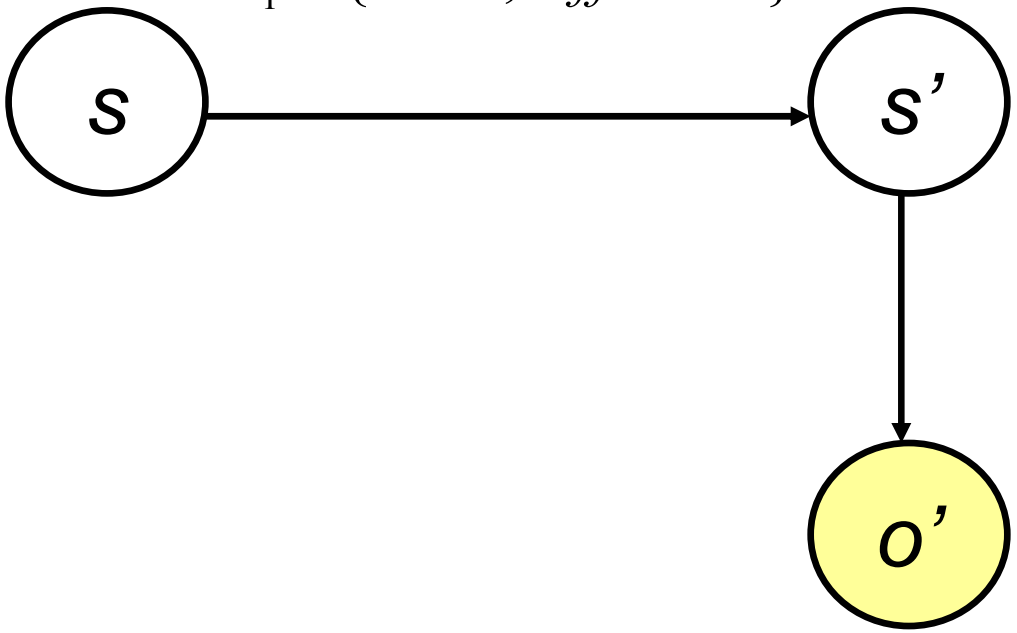


Observations in the summary POMDP

Observations in summary POMDP give an compact indication of how $b(\tilde{s})$ is changing based on ***s*** and ***o***. It has 2 parts.

$$\arg\max(b(s_u)) == \arg\max(b(s'_u))?$$

$$\tilde{o}_1 \in \{same, different\}$$



Is $\arg\max(b(s_u))$ "consistent" with *o*?

$$\tilde{o}_2 \in \begin{cases} consistent \\ inconsistent \\ no-info \end{cases}$$

“Summary POMDP” method

1

$$\left.\begin{array}{c} p(s' \mid s,a) \\ p(o' \mid s',a) \\ r(s,a) \end{array}\right\} \xrightarrow{\text{sample}} \left\{\begin{array}{c} p(\tilde{s}' \mid \tilde{s},\tilde{a}) \\ p(\tilde{o}' \mid \tilde{s}',\tilde{a}) \\ r(\tilde{s},\tilde{a}) \end{array}\right.$$

2

$$\left.\begin{array}{c} p(\tilde{s}' \mid \tilde{s},\tilde{a}) \\ p(\tilde{o}' \mid \tilde{s}',\tilde{a}) \\ r(\tilde{s},\tilde{a}) \end{array}\right\} \xrightarrow{\text{optimize}} \pi(b(\tilde{s}))$$

3

run

maintain $b(s)$

$$\left(\begin{array}{c} b(\tilde{s}_d) \leftarrow b(s_d) \\ b(\tilde{s}_u) \leftarrow \max(b(s_u)) \\ \tilde{a} \leftarrow \pi(b(\tilde{s})) \\ a \leftarrow (\tilde{a}, \arg\max(b(s_u))) \end{array}\right)$$

Outline

- Dialogue management as a POMDP
- Scaling up with the “summary POMDP” method
- Example problem & results
- Conclusions & future work

1000-place dialog problem

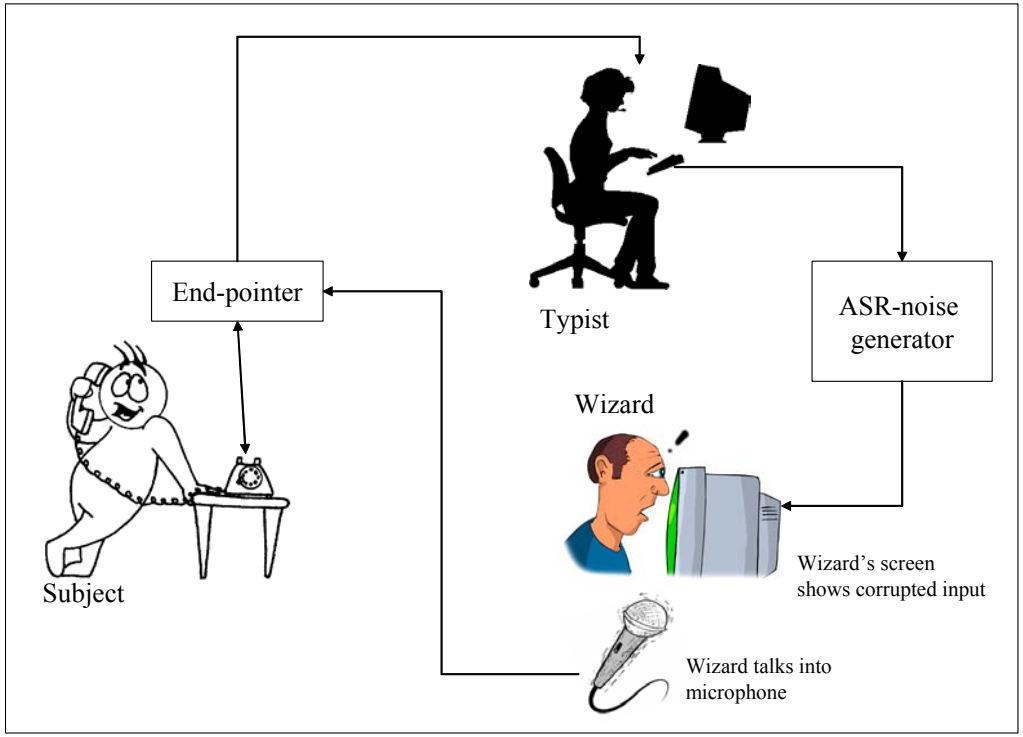
User is travelling from *x* to *y* in a world with 1000 cities

- s_u *:* The user's desired *from* and *to* cities – e.g., *(a,b)*
- a_u, o *: yes, no, null, a, from(a), to(a), {from(a),to(b)},* etc.
- s_d *:* Each slot can be *not stated*, *unconfirmed*, or *confirmed*
- a_s *: ask-from, ask-to, conf-from(a), conf-to(b), submit(a,b)*

- **User goal model**: User has a fixed goal throughout the dialogue
- **User action model**: Estimated from real dialogue data
- **Conversation model**: Deterministically "promotes" slots
- **Speech recognition model** : Substitutions & del's with prob. p_{err}
- **Reward fn** : +50 correct submit, -50 wrong submit, -5 fail, -3 for confirming something which is *not stated*, etc., -1 per-turn otherwise

- **Solve with *Perseus* (PBVI)**

Simulated speech recognition channel



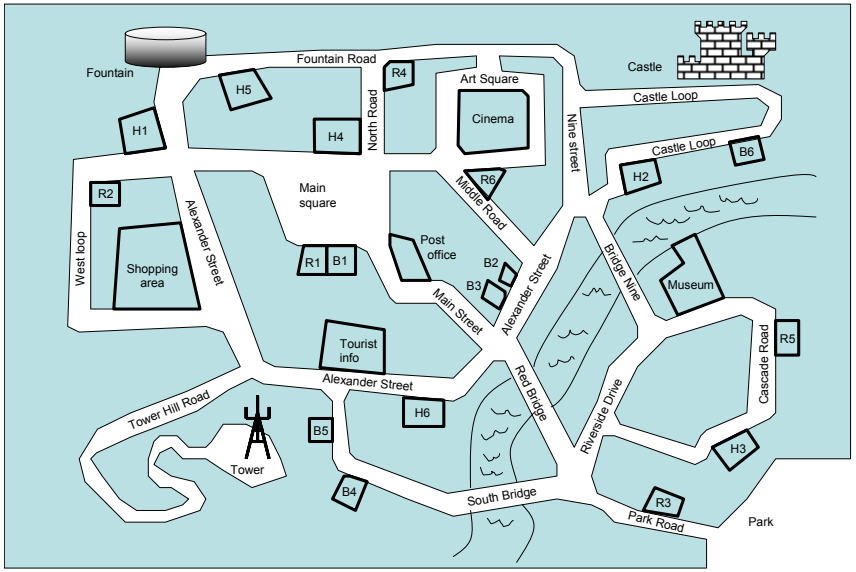
- 2 subjects interact through a *simulated* speech recognition channel
- Typical end-pointing model
- Utterances are transcribed by a typist
- Confusions generated using lexicon, phonetic confusion matrix, and language model

Matthew Stuttle, Jason D. Williams, and Steve Young. *A Framework for Wizard-of-Oz Experiments with a Simulated ASR-Channel*. ICSLP, 2004, South Korea.

“SACTI-1” dialog data corpus

- “Tourist information” domain
- Wizard has more info than user; user given specific task

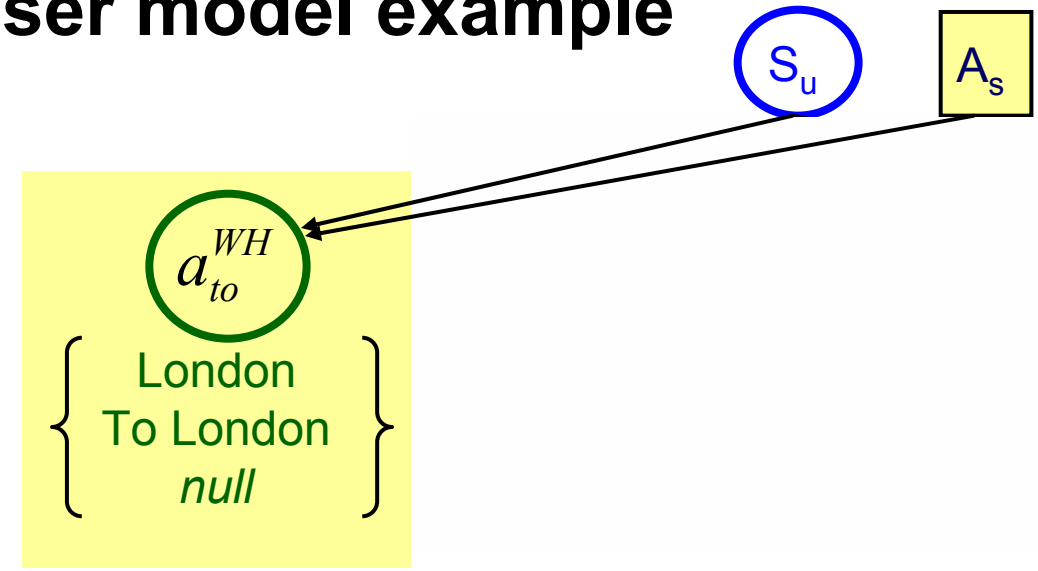
- 36 users / 12 wizards
- 144 dialogs
- 4071 turns
- 4 different WER targets
- Task completion
- Likert-scale questions



Conversational characteristics broadly similar to those observed in human/machine dialog.

Jason D. Williams and Steve Young. *Characterizing Task-Oriented Dialog using a Simulated ASR Channel*. ICSLP, 2004, South Korea.

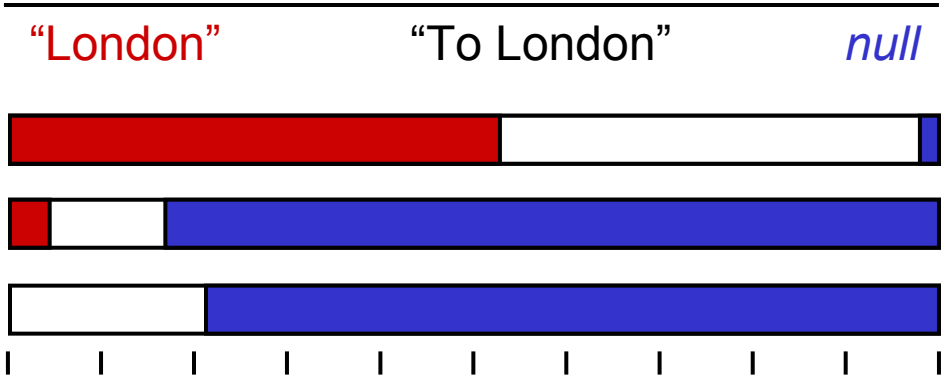
User model example



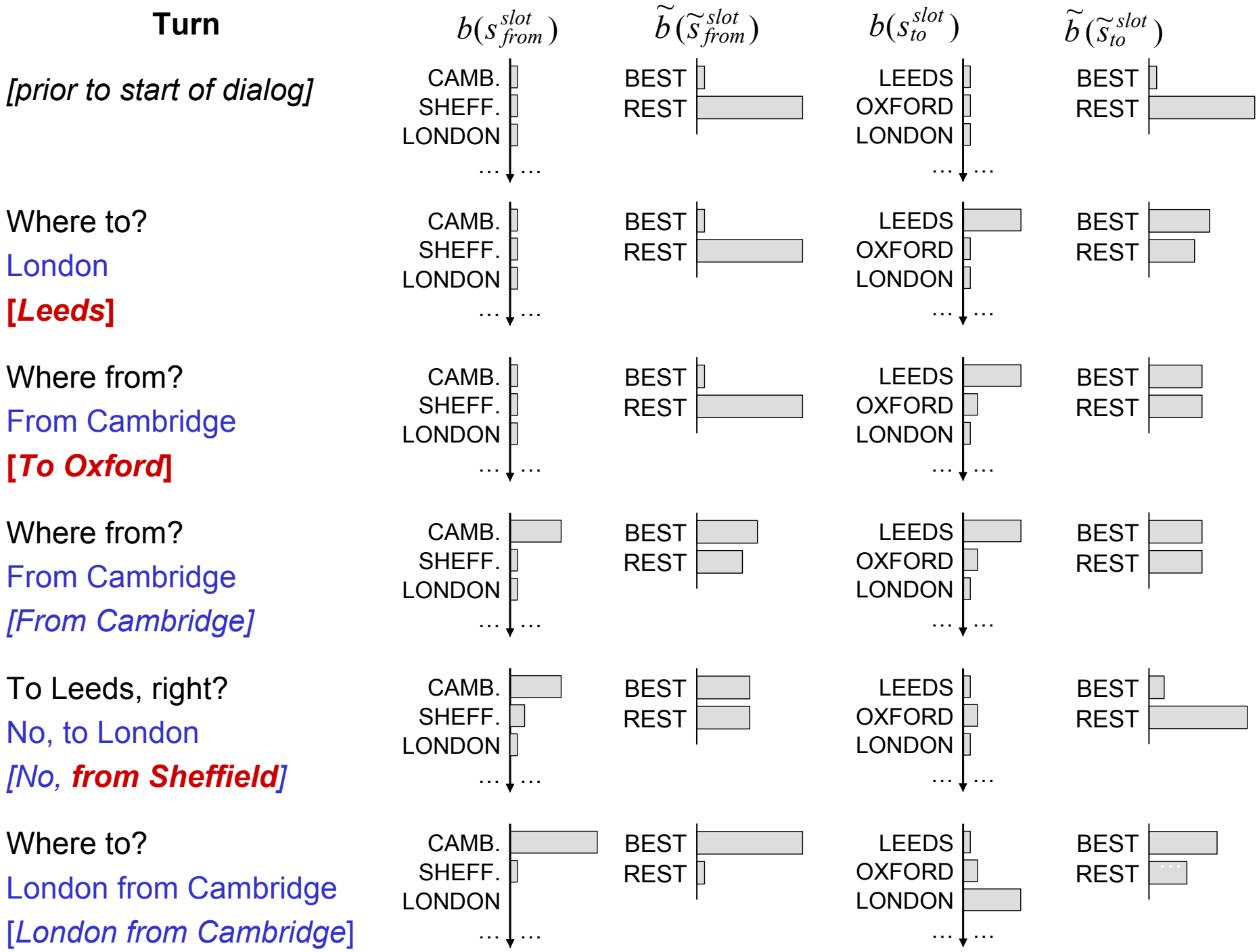
P(A$_u$-WH | A$_s$, S$_u$=to(london))

A$_s$
Where are you going to?
To London, is that right?
Where are you going from?

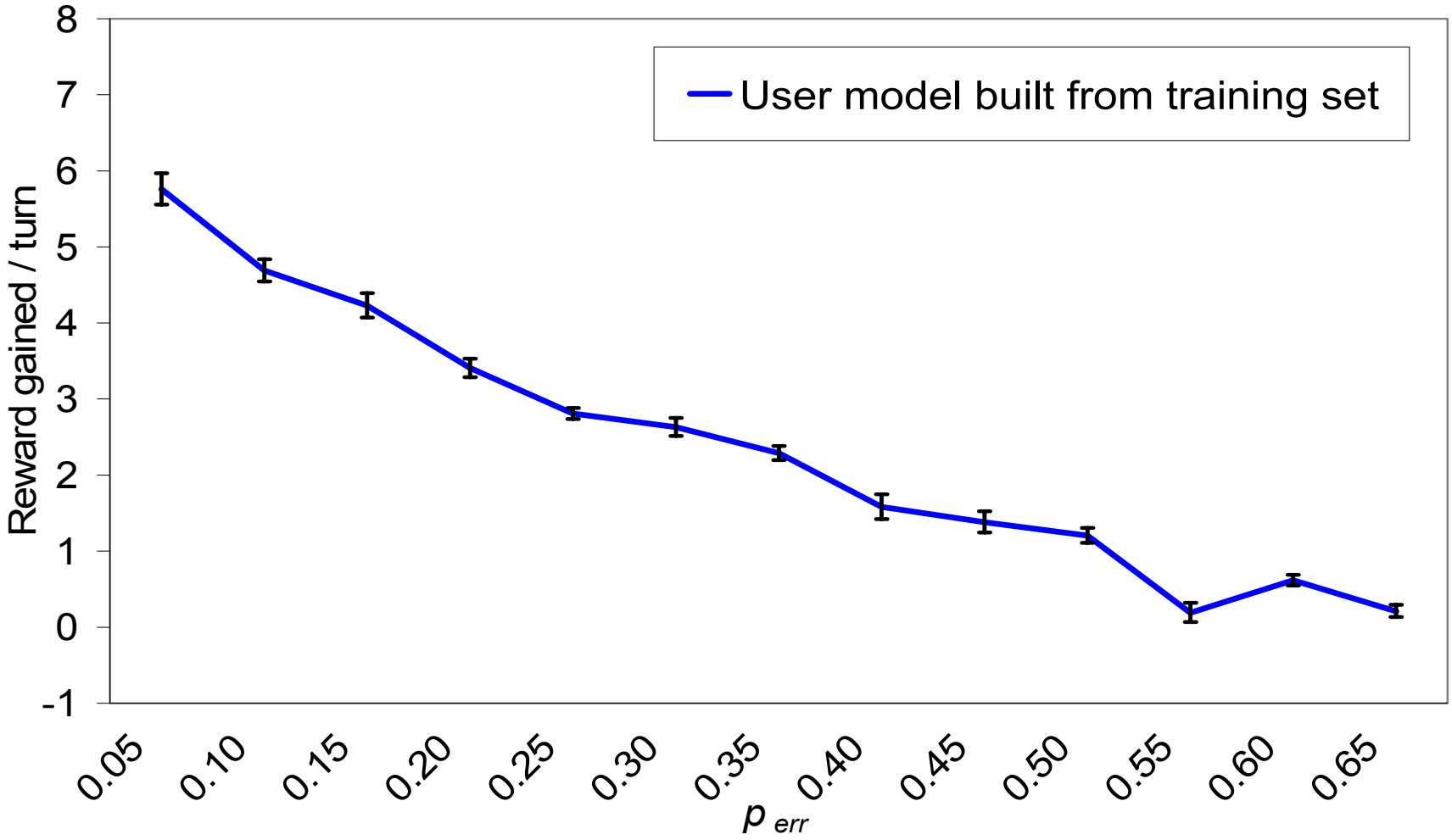
P mass of A$_u$ ("to/WH" portion only)



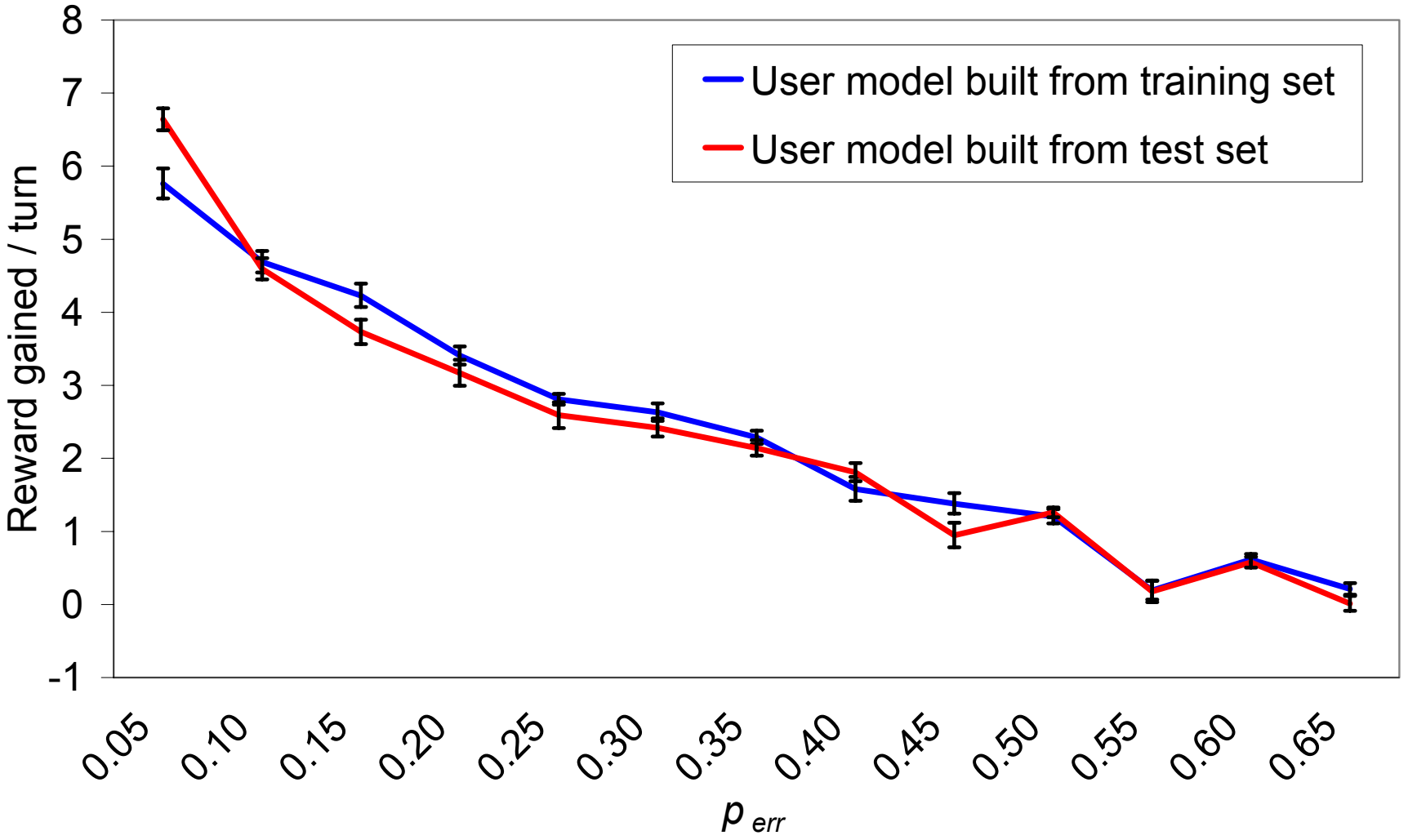
Data was divided, and two model sets were created:
Training model set & Testing model set


Turn
$b(s^{slot}_{from})$
$\tilde{b}(\tilde{s}^{slot}_{from})$
$b(s^{slot}_{to})$
$\tilde{b}(\tilde{s}^{slot}_{to})$
[prior to start of dialog]
Where to?
London
[Leeds]
Where from?
From Cambridge
[To Oxford]
Where from?
From Cambridge
[From Cambridge]
To Leeds, right?
No, to London
[No, from Sheffield]
Where to?
London from Cambridge
[London from Cambridge]
CAMB.
SHEFF.
LONDON
…
BEST
REST
LEEDS
OXFORD
LONDON

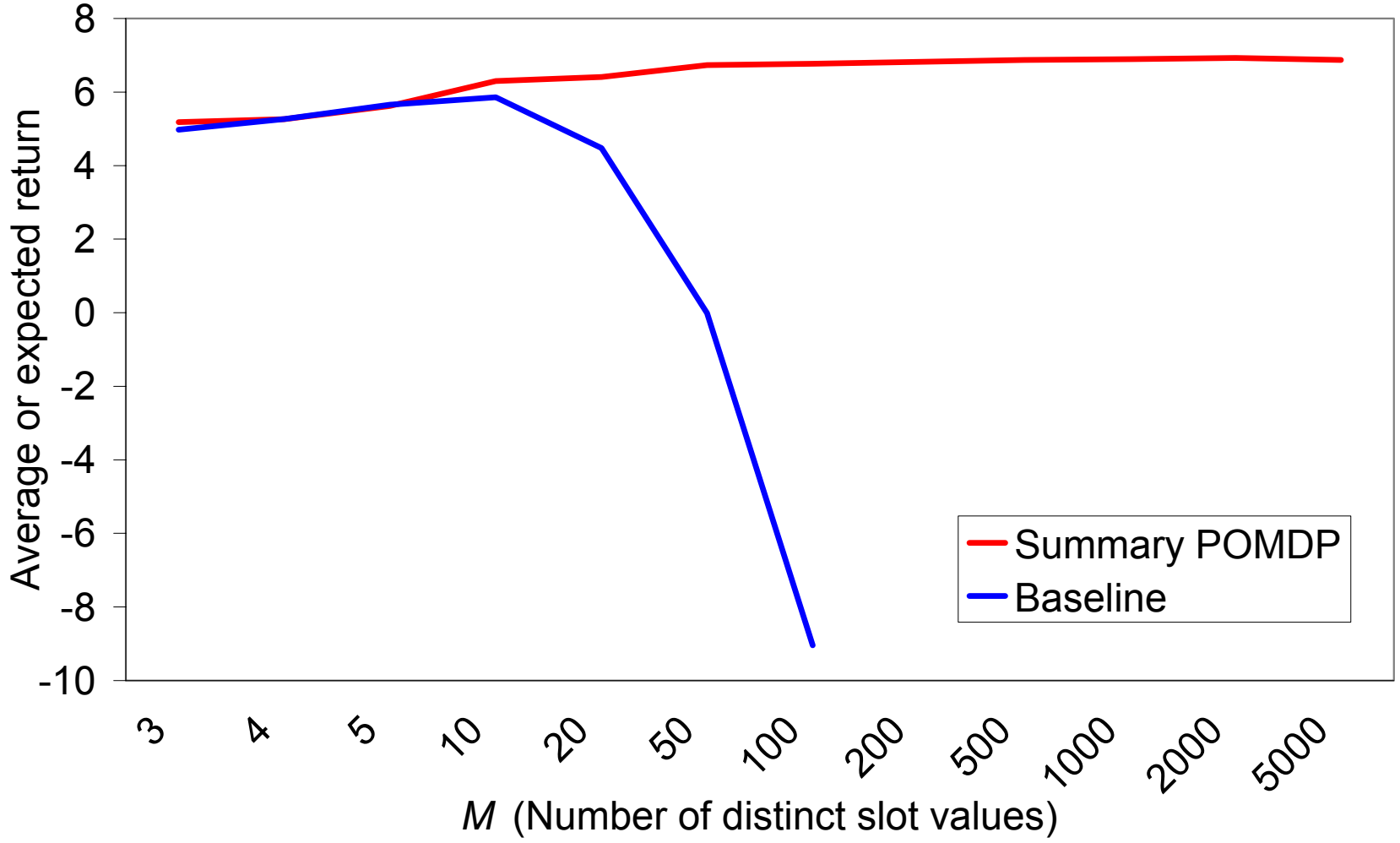
Performance vs. p_{err}



Performance vs. p_{err}



Scalability vs. direct optimization



Outline

- Dialogue management as a POMDP
- Scaling up with the “summary POMDP” method
- Example problem & results
- Conclusions & future work

Conclusions & future work

- For slot-based dialog management, POMDPs…
  - Outperform MDP baseline
  - Outperform 3 handcrafted baselines
  - Can be scaled as a “summary POMDP” to many slot values

- Interesting future avenues
  - Scale *number of* slots to 10s
  - Move beyond slot-based approaches
  - Apply to real system

Thanks!

Jason D. Williams

jdw30@cam.ac.uk

Joint work with

Steve Young

sjy@eng.cam.ac.uk