

# Exploring question-driven models of discourse using a task-oriented dialogue corpus

Christopher Potts, Stanford Linguistics

University of Rochester Center for Language Sciences  
October 13, 2011



# Joint work with



David Clausen



Karl Schultz



Alex Djalali



Sven Lauer

# Overview

## Goals

- 1 Explore discourse models driven by Questions Under Discussion.
- 2 Introduce a highly structured task-oriented dialogue corpus.
- 3 Try to inspire others to work with the corpus by reporting on a few suggestive pilot studies.

# Overview

## Goals

- 1 Explore discourse models driven by Questions Under Discussion.
- 2 Introduce a highly structured task-oriented dialogue corpus.
- 3 Try to inspire others to work with the corpus by reporting on a few suggestive pilot studies.

## Plan

- 1 Question Under Discussion (QUD) models
- 2 The Cards corpus: scenario and implementation
- 3 Basic properties of the Cards corpus
- 4 QUD-based domain restriction
- 5 Expert effects and rich common ground
- 6 Looking ahead

# Corpus distribution

`http://CardsCorpus.christopherpotts.net/`

Included:

- The transcripts in CSV format
- Python classes for working with the transcripts
- Examples of the Python classes in action
- R code for reading in the corpus as a data frame

## Questions Under Discussion (QUDs)

- 1 Question Under Discussion (QUD) models
- 2 The Cards corpus: scenario and implementation
- 3 Basic properties of the Cards corpus
- 4 QUD-based domain restriction
- 5 Expert effects and rich common ground
- 6 Looking ahead

## Groenendijk & Stokhof (and Wittgenstein)

**Wittgenstein** We might very well write every assertion in the form of a question followed by an affirmative expression; for instance 'Is it raining? Yes!' Would that mean that behind every claim lies a question?

**Groenendijk  
& Stokhof** Ja!

## Groenendijk & Stokhof (and Wittgenstein)

**Wittgenstein** We might very well write every assertion in the form of a question followed by an affirmative expression; for instance 'Is it raining? Yes!' Would that mean that behind every claim lies a question?

**Groenendijk  
& Stokhof** Ja!

Bart passed.

## Groenendijk & Stokhof (and Wittgenstein)

**Wittgenstein** We might very well write every assertion in the form of a question followed by an affirmative expression; for instance 'Is it raining? Yes!' Would that mean that behind every claim lies a question?

**Groenendijk & Stokhof** Ja!

Did Bart pass?



Bart passed.

## Groenendijk & Stokhof (and Wittgenstein)

**Wittgenstein** We might very well write every assertion in the form of a question followed by an affirmative expression; for instance 'Is it raining? Yes!' Would that mean that behind every claim lies a question?

**Groenendijk  
& Stokhof** Ja!

Did Bart pass?



Bart passed.

BART<sub>F</sub> passed.

Bart PASSED<sub>F</sub>.

The ROCKSTARS<sub>F</sub> wore LEATHER<sub>F</sub>.

## Groenendijk & Stokhof (and Wittgenstein)

**Wittgenstein** We might very well write every assertion in the form of a question followed by an affirmative expression; for instance 'Is it raining? Yes!' Would that mean that behind every claim lies a question?

**Groenendijk & Stokhof** Ja!

Did Bart pass?



Bart passed.

Who passed?



BART<sub>F</sub> passed.

What did Bart do?



Bart PASSED<sub>F</sub>.

Who wore what?



The ROCKSTARS<sub>F</sub> wore LEATHER<sub>F</sub>.

# Questions

I needn't commit to a particular view of questions. I require only:

- 1 questions present alternatives;
- 2 questions are not linguistic objects, though some natural language sentences might identify some of them; and
- 3 questions can be partially ordered by some notion of resolution.

## Discourse models

### Questions under Discussion (QUDs)

Discourse is structured by abstract, implicit QUDs, which shape the common ground, determine what is informative and relevant to communicate, and help define the domain (Groenendijk & Stokhof, Ginzburg, Roberts, Büring).

### Decision problems

van Rooij showed how to induce questions from a utility function:

$$a^* = \{w \in W \mid U(a, w) \geq U(a', w) \text{ for } a' \in A\}$$

### Planning

The QUD theory does not provide a theory of planning, but there should be a mutually beneficial relationship between the two.

## Pragmatically required over-answering

**Context:** Homer calls a hotel.

Homer: Is Lisa Simpson in Room 10?

Clerk A: She's in room 20.

Clerk B: #No.

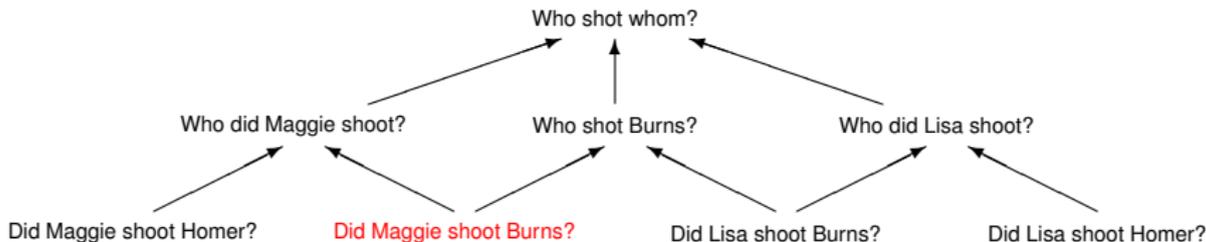


(Ginzburg, Roberts)

# Anaphora

Homer: Did Maggie shoot Burns?

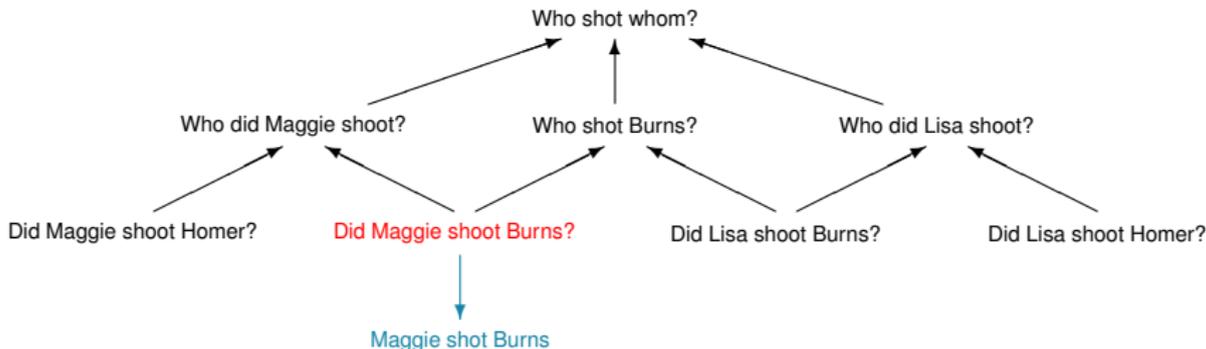
Wiggum: She did.



# Anaphora

Homer: Did Maggie shoot Burns?

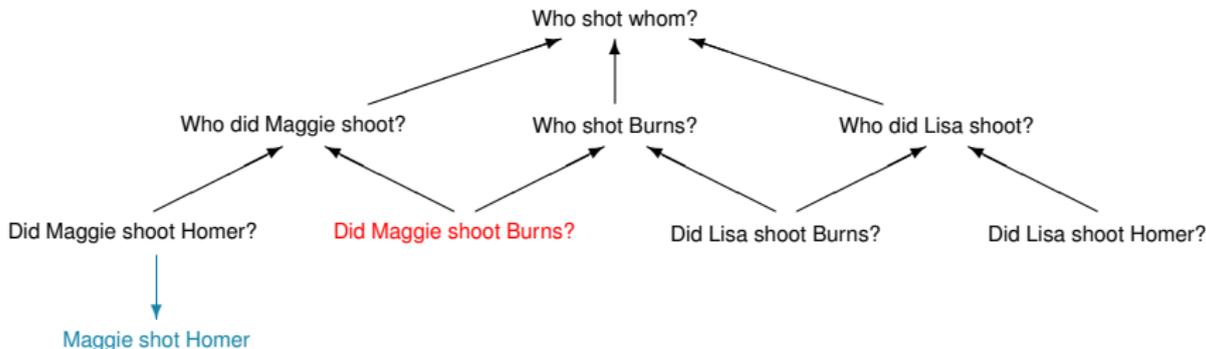
Wiggum: She did. **Maggie shot Burns**



# Anaphora

Homer: Did Maggie shoot Burns?

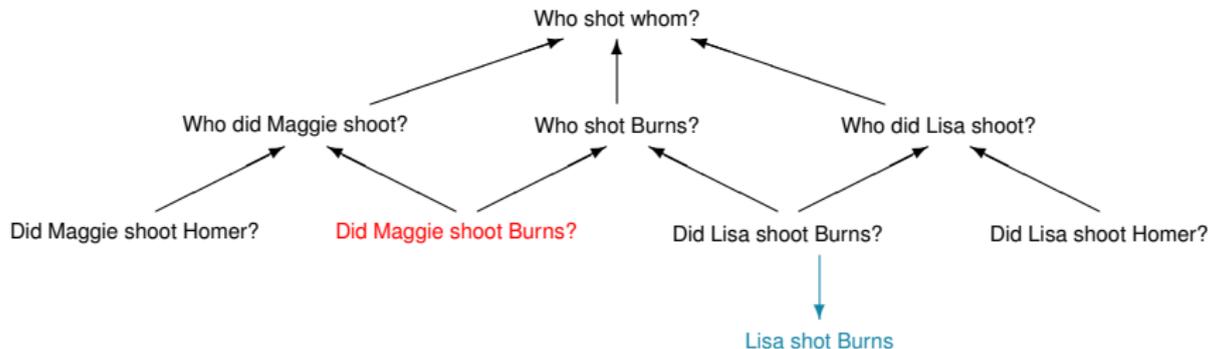
Wiggum: She did. **Maggie shot Homer**



# Anaphora

Homer: Did Maggie shoot Burns?

Wiggum: She did. Lisa shot Burns



## Domain restriction

*I didn't see any.*

## Domain restriction

- Are there typos in my slides?

*I didn't see any.*

## Domain restriction

- Are there typos in my slides?
- Are there bookstores downtown?

*I didn't see any.*

## Domain restriction

- Are there typos in my slides?
- Are there bookstores downtown?
- Are there cookies in the fridge?

*I didn't see any.*

## Domain restriction

- Are there typos in my slides?
- Are there bookstores downtown?
- Are there cookies in the fridge?
- ...

*I didn't see any.*

# Granularity

*Where are you from?*

(Groenendijk & Stokhof, Ginzburg)

# Granularity

*Where are you from?*

- *Connecticut.*

(Issue: birthplaces)

(Groenendijk & Stokhof, Ginzburg)

# Granularity

*Where are you from?*

- *Connecticut.* (Issue: birthplaces)
- *The U.S.* (Issue: nationalities)

(Groenendijk & Stokhof, Ginzburg)

# Granularity

*Where are you from?*

- *Connecticut.* (Issue: birthplaces)
- *The U.S.* (Issue: nationalities)
- *Stanford.* (Issue: affiliations)

(Groenendijk & Stokhof, Ginzburg)

# Granularity

*Where are you from?*

- *Connecticut.* (Issue: birthplaces)
- *The U.S.* (Issue: nationalities)
- *Stanford.* (Issue: affiliations)
- *Planet earth.* (Issue: intergalactic meetings)

(Groenendijk & Stokhof, Ginzburg)

## Mention-some/mention-all

*Where can we buy supplies?*

(Beck & Rullmann, van Rooij & Schultz)

# Mention-some/mention-all

*Where can we buy supplies?*

## Mention-all

- **Context:** We're writing a comprehensive guide to the area.
- **Resolvedness condition:** An exhaustive listing of the (reasonable) shopping places.

## Mention-some/mention-all

*Where can we buy supplies?*

### Mention-all

- **Context:** We're writing a comprehensive guide to the area.
- **Resolvedness condition:** An exhaustive listing of the (reasonable) shopping places.

### Mention-some

- **Context:** We're low on food and water.
- **Resolvedness condition:** Mentioning the best (closest, safest, etc.) place, or a few good options.

# Mention-some/mention-all

*Who has a light?*

(Beck & Rullmann, van Rooij & Schultz)

# Mention-some/mention-all

*Who has a light?*

## Mention-all

- **Context:** Speaker needs to ensure that no one in the group is going to get stopped by airport security.
- **Resolvedness condition:** List of everyone who has a light.

## Mention-some/mention-all

*Who has a light?*

### Mention-all

- **Context:** Speaker needs to ensure that no one in the group is going to get stopped by airport security.
- **Resolvedness condition:** List of everyone who has a light.

### Mention-some

- **Context:** Speaker needs to light her cigar.
- **Resolvedness condition:** Just name one (friendly, willing, nearby) person with a lighter.

## Structured domain restriction

*What cards do you have?*

## Structured domain restriction

*What cards do you have?*

### Wide domain

- **Context:** Speaker dealt the cards and noticed that some were missing.
- **Resolvedness condition:** List everything you're holding.

# Structured domain restriction

*What cards do you have?*

## Wide domain

- **Context:** Speaker dealt the cards and noticed that some were missing.
- **Resolvedness condition:** List everything you're holding.

## Narrowed, structured domain

- **Context:** Speaker folds. He wants to know what beat him.
- **Resolvedness condition:** Just name the good cards.

## Definite plurals

*Are the windows open?*

(Malamud)

## Definite plurals

*Are the windows open?*

### Existential

- **Context:** We're leaving town and want the house secure.
- **Resolvedness condition:** Is any window open?

# Definite plurals

*Are the windows open?*

## Existential

- **Context:** We're leaving town and want the house secure.
- **Resolvedness condition:** Is any window open?

## Universal

- **Context:** The sills will be painted.
- **Resolvedness condition:** Is every window open?

## Discourse particles

German *überhaupt*

(Rojas-Esponda)

The speaker wishes to move to a higher QUD than the current one.

German *ja*

(Kratzer & Matthewson)

“the speaker is firmly committed to the descriptive content *p*, and doesn't consider the question whether or not *p* to be on the agenda for either the current or any future inquiry”.

Japanese sentence-final *yo*

(Chris Davis)

The speaker indicates a belief that the utterance resolves the addressee's immediate issue.

## Summary

- Good evidence that QUDs are a factor in resolving context dependence.
- Relatively little quantitative or corpus exploration of this idea (but see work by Cooper, Fernandez, Ginzburg, Lappin, Larsson).
- Such evidence would help us to better understand where and how QUDs come into play, and how speakers represent discourses, issues, and lexical and constructional meanings.

# The Cards corpus: scenario and implementation

- 1 Question Under Discussion (QUD) models
- 2 The Cards corpus: scenario and implementation**
- 3 Basic properties of the Cards corpus
- 4 QUD-based domain restriction
- 5 Expert effects and rich common ground
- 6 Looking ahead

# Amazon Mechanical Turk HIT (Human Intelligence Task)

- **Title:** Collaborative Search Game with Chat
- **Description:** Two-player collaborative video game involving dialogue/chat with other Turkers.
- **Payment:** \$1.00, and up to \$0.50 cents for rich, collaborative problem-solving using meaningful dialogue.
- **Restrictions:** US IP addresses; at least 95%. approval rating

# Instructions

## Instructions

1. You'll first be presented with an experimental consent form; the transcripts from these games will be used to study collaborative problem solving.
2. Once you click through, you'll go into "Wait for Participant" mode until another player shows up. **It might take a while for another player to show up! We suggest opening a new window and doing other things until someone arrives**
3. When play begins, you'll be presented with a task description: you and your partner are tasked with finding six consecutive cards of the same suit.
4. When you've solved the game, click the "Task Complete" button to register that the task is complete. You can continue playing and chatting at that point, or you can click **Complete HIT and return to MTurk** at the top of the screen.
5. You will receive a bonus of up to \$0.50 for good collaboration with meaningful communication.
6. Notes:
  - Once the game begins, you can move around with the arrow keys or the on-screen buttons.
  - The chat window is near the top of the screen.
  - You can't see your partner, and your partner can't see you!
  - The yellow boxes mark cards in your "line of sight".
7. **Caution: Players who do not complete the task properly will not be paid. It is vital that you read and understand the task description and follow the instructions.**
8. Feel free to play multiple times — especially useful if you get good at the game.

## Annotated game screen

# Gameboard

**TYPE HERE**

**Yellow boxes mark cards in your line of sight.**

**You are on 2D**

**Task description: Six consecutive cards of the same suit**

Received: hi  
Sent: I have the JH  
Received: I have the SH

Type text here:  
[Disable Sound] I'm on 2D, which isn't too useful. There are cards to my right and below, though. I'll check them out.

Gather six consecutive cards of a particular suit (decide which suit together). Each of you can hold only three cards at a time, so you'll have to coordinate your efforts. You can talk all you

P1 turns remaining: 546  
P2 turns remaining: 599

up  
Click a card to pick it up:  
2D

left  
Click a card to drop it from your hand:  
JH  
right

down

[Indicate Task Complete]

**The cards you are holding**

**Move with the arrow keys or these buttons.**

# Consent form

## **Title of project: Pragmatic enrichment and contextual inference**

Principal investigator: Dr. Christopher Potts, Department of Linguistics, Stanford University

### **Purpose of Study**

In this research, we are investigating how people communicate when working together to resolve a joint goal. We collect transcripts of dialogues of players playing a simple game that requires participants to strategize together to solve tasks.

### **Procedures**

In this study, you will play a game with another unseen player over a network. You will see instructions on the screen, and you may communicate with the other player by keyboard. The game consists of a map and several hidden objects. You have a limited number of moves to collect the objects according to the instructions you receive. You must coordinate with your teammate to achieve the goal. The time of a single game is variable depending on the players. A typical game lasts between 10 and 20 minutes.

### **Risks**

There are no known risks involved in this experimental procedure.

### **Benefits**

## Scenario

*Gather six consecutive cards of a particular suit (decide which suit together), or determine that this is impossible. Each of you can hold only three cards at a time, so you'll have to coordinate your efforts. You can talk all you want, but you can make only a limited number of moves.*



# Transcripts: game play

Agent	Time	Action type	Contents
			⋮
Player 1	566650	PLAYER_MOVE	7,11
Player 2	567771	CHAT_MESSAGE_PREFIX	which c's do you have again?
Player 1	576500	CHAT_MESSAGE_PREFIX	i have a 5c and an 8c
Player 2	577907	CHAT_MESSAGE_PREFIX	i jsut found a 4 of clubs
Player 1	581474	PLAYER_PICKUP_CARD	7,11:8C
Player 1	586098	PLAYER_MOVE	7,10
			⋮

## MTurk set-up

- Data collection in June 2010 and August 2011.
- PHP wrapper to Pragbot written by Victoria Schwanda.
- Server-side configuration by Chriz Czyzewicz.
- Collection times: 5 batches each lasting about 5 hours, spread out over two work weeks.
- At peak times: 30 transcripts per hour.
- Total cost: about \$2,000.

## Turker Nation discussion

Initial post announcing the new HITs, our goals for them, and warning about wait-times and potential bugs.

[Click here to go to the NEW Turker Nation forum](#) :: [Requesters Boardroom](#) :: [Everyone Else](#) :: [Collaborative Search Game with Chat](#)

Page 1 of 11 » Jump to page  Go

[remove bookmark](#) [share](#) [print](#)

 Author	Topic: Collaborative Search Game with Chat (Read 1,444 times)
<p><b>stanfordpraglab</b> Grizzled Turker ☆☆☆☆☆ member is <b>online</b></p> <p></p> <p>Joined: Jun 2010 Posts: 126</p>	<p> <b>Collaborative Search Game with Chat</b> « Thread Started on Aug 10, 2011, 10:21am » <a href="#">modify</a> <a href="#">delete</a></p> <hr/> <p>Turker Nation!</p> <p>I'm writing to let you all know that we've some posted some more HITs called Collaborative Search Game with Chat</p> <p>In the game, you navigate around in a maze-world. You and your partner try to gather six consecutive cards of the same suit before you run out of moves. You and your partner decide which suit and which sequence.</p> <p>You can see where cards are (yellow boxes) when they enter your "line of sight". Navigation is with the keyboard or with the buttons on the right of the screen. When you're on a card, its name displays on the right and you can pick it up by clicking on it.</p> <p>You can hold only three cards at once. so you have to coordinate on your strategv.</p>

# Turker Nation discussion

Responding to player queries while the HITs were running.

<p><b>stanfordpraglab</b> Grizzled Turker ☆☆☆☆ member is <b>online</b></p> <p></p> <p>Joined: Jun 2010 Posts: 126</p>	<p><b>Re: Collaborative Search Game with Chat</b> ◀ Reply #9 on Aug 10, 2011, 11:04am ▶</p> <p> <a href="#">modify</a>  <a href="#">delete</a></p> <hr/> <p><b>Aug 10, 2011, 10:50am, moniquitta wrote:</b></p> <div data-bbox="378 467 1163 515" style="border: 1px solid black; padding: 5px;"><p>Would this be weird but can we like state TN here in the beginning so we know we are playing someone from here?</p></div> <p>This is totally fine by us --- even adds a new community dimension to the data. ---Chris</p> <p><a href="#">Report to Mod</a> - <a href="#">Link to Post</a> - <a href="#">Back to Top</a>  Logged</p>
<p><b>moniquitta</b> Turkaholic ☆☆☆☆☆ member is <b>offline</b></p> <p></p> <p>Joined: Feb 2011 Gender: Female ♀ Posts: 722</p>	<p><b>Re: Collaborative Search Game with Chat</b> ◀ Reply #10 on Aug 10, 2011, 11:05am ▶</p> <hr/> <p>Sweet awesome. OH and good to see you on here working with us.</p> <p><a href="#">Report to Mod</a> - <a href="#">Link to Post</a> - <a href="#">Back to Top</a>  Logged</p>

# Turker Nation discussion

Responding to player queries while the HITs were running.

<p><b>stanfordpraglab</b> Grizzled Turker ☆☆☆☆ member is <b>online</b></p>	<p> <b>Re: Collaborative Search Game with Chat</b> « Reply #80 on Aug 11, 2011, 10:55am »</p> <p> <a href="#">modify</a>  <a href="#">delete</a></p>
<p> Joined: Jun 2010 Posts: 126</p>	<p>Hi Turkers,</p> <p>You will absolutely get approved if you make a genuine attempt to solve the game and run out of moves. Hope this version isn't too hard.</p> <p>Stanford Praglab</p> <p><a href="#">Report to Mod</a> - <a href="#">Link to Post</a> - <a href="#">Back to Top</a>  <a href="#">Logged</a></p>

# Turker Nation discussion

The players helped us spot trouble.

<p><b>timefactor</b> Turker ☆☆☆ member is offline</p>	<p> <b>Re: Collaborative Search Game with Chat</b> « Reply #152 on Aug 11, 2011, 4:28pm »</p>
<p> Joined: Jun 2011 Gender: Male ♂ Posts: 13</p>	<p><b>Aug 11, 2011, 4:24pm, sleepstar wrote:</b> wow.. high paying, fast HITs.. go figure the stupid scammers would show up. ughhhhhhhh.</p> <p>Seriously. This game is not that hard and wouldn't even take much longer than just blindly submitting. This batch was disappointing since three of my games were these losers who didn't play and I was too scared to submit them.</p> <p><a href="#">Report to Mod</a> - <a href="#">Link to Post</a> - <a href="#">Back to Top</a>  <a href="#">Logged</a></p>

# Turker Nation discussion

And resolve trouble.

<p><b>bid</b> Turkaholic ☆☆☆☆☆ member is offline</p> <p></p> <p>Joined: Jan 2011 Gender: Male  Posts: 769</p>	<p><b>Re: Collaborative Search Game with Chat</b> « Reply #158 on Aug 11, 2011, 4:45pm »</p> <hr/> <p><b>Aug 11, 2011, 4:35pm, moniquitta wrote:</b> I wonder if a qual would help.</p> <p>I wondered the same thing. I suppose it depends on how many of these there will be. Of course, with a qual, they could weed out some, shall we say less than dedicated participants, increase the quality of data being obtained and (can't believe I am saying this) make it increasingly challenging to succeed.</p> <p><a href="#">Report to Mod</a> - <a href="#">Link to Post</a> - <a href="#">Back to Top</a>  <a href="#">Logged</a></p>
---	--

## Feedback

*That was actually a pretty fun hit.*

*The game with chat was great and like to see more HITs from you.*

*These HITs were really enjoyable. Hopefully you will put more on the site. You state that we can keep doing them, but right now if I click on your HIT, it tells me there are no more available for me. Is there something I can do to try again? Thanks.*

*I waited 1.22 before someone showed up. They never talked to me and didn't finish the job before leaving. Am I still out because they didn't cooperate?*

## Other task-oriented dialogue corpora

Corpus	Task type	Domain	Task-orientation	Dialogues	Format
Switchboard	discussion	wide-open	very loose	2,400	aud/txt
SCARE	search	3d world	tight	15	aud/vid/txt
Cards	search	2d grid	tight	745	txt
TRAINS	routes	map	tight	120	aud/txt
Map Task	routes	map	tight	128	aud/vid/txt

(See also Blaylock & Allen, 'Generating artificial corpora for plan recognition')

## Other task-oriented dialogue corpora

Corpus	Task type	Domain	Task-orientation	Dialogues	Format
Switchboard	discussion	wide-open	very loose	2,400	aud/txt
SCARE	search	3d world	tight	15	aud/vid/txt
Cards	search	2d grid	tight	745	txt
TRAINS	routes	map	tight	120	aud/txt
Map Task	routes	map	tight	128	aud/vid/txt

(See also Blaylock & Allen, 'Generating artificial corpora for plan recognition')

Chief selling points for Cards:

- Pretty large.
- Controlled enough that similar things happen often.
- Very highly structured — the only corpus whose release version allows the user to replay all games with perfect fidelity.

# The Cards corpus: basic properties

- 1 Question Under Discussion (QUD) models
- 2 The Cards corpus: scenario and implementation
- 3 Basic properties of the Cards corpus**
- 4 QUD-based domain restriction
- 5 Expert effects and rich common ground
- 6 Looking ahead

# Sample run

[START](#) ← [BACK](#) [FORWARD](#) → [FINISH](#) (keyboard navigation: ← back | → forward | ↓ start | ↑ finish)

Player 1 holding: 8S, 7S, 5S  
 Player 2 holding: 6S, 9S  
 Current event Player 2, 771670: PLAYER\_MOVE15,20

1. Player 1 (36836): OK, we need to choose a suit - spades, etc. Any preference?
2. Player 2 (63662): I just got here. Doesn't matter to me.
3. Player 1 (73356): Ok, how about spades?
4. Player 2 (79177): Sounds good.
5. Player 1 (98577): Use arrow keys to move around maze, look for yellow boxes that have S
6. Player 1 (122058):
7. Player 1 (136792): Picked up 8 S
8. Player 2 (143520): Great!
9. Player 1 (170895): We need consecutives: 5, 6, 7, or 9, 10, 11
10. Player 1 (286170): picked up 10 Spades; we need the 9S
11. Player 2 (302728): I haven't found any yet. I'm still looking.
12. Player 2 (350053): I keep getting red squares. Do you know what they mean?
13. Player 1 (362998): No, I was wondering.
14. Player 2 (538256): I have a 6 and 9S
15. Player 1 (552754): picked up 7 spades. I can only hold 3 cards at a time.
16. Player 2 (568965): So what do we have now?
17. Player 1 (570774): Wow!
18. Player 1 (588214): We either need 9S or 5S
19. Player 2 (624182): OK, I just saw the 2 and 3.
20. Player 2 (630221): I'll keep going.
21. Player 1 (688140): 5S is at lower right corner - I can't pick up
22. Player 1 (711580): Should I drop 10?
23. Player 2 (711608): Outside the maze?
24. Player 1 (750054): Our mazes might be different. Maybe only MY maze has it at lower right. I'll drop 10S

## By the numbers

- 744 transcripts
- Game length mean: 414.44 actions (median 325.50, sd 261.88)
- Actions:
  - Card pickup: 11,027
  - Card drop: 7,202
  - Move: 255,734
  - Utterance: 23,532
    - Utterance length mean: 5.84 words (median 5, sd 5.08)
    - Total word count: 137,323
    - Total vocabulary: 4,004 (3,453 if card references are normalized)

# Constants and total randomness

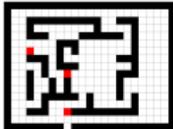
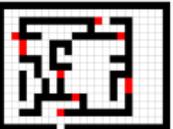
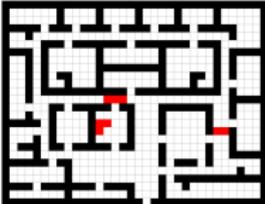
## Constants

- Task description ('six consecutive cards of the same suit')
- Line of sight: 3
- Max cards in hand: 3

## Randomness

- Players' initial positions
- All card positions

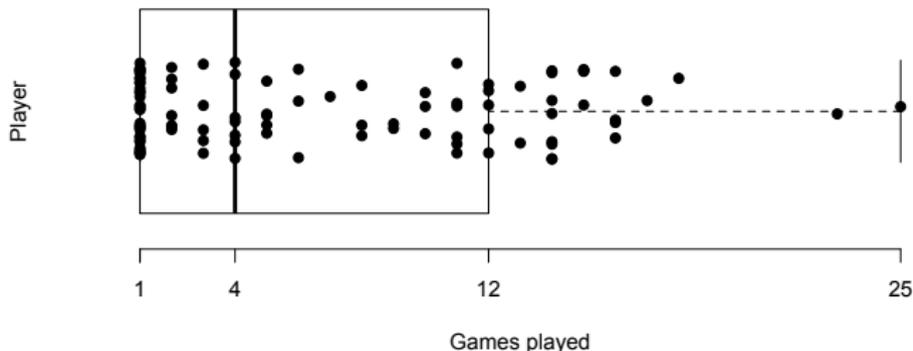
## Points of variation

		 ★	 ★		Total
Max turns	100	214	0	0	214
	500	0	0	44	137
	600	180	144	70	394
Total	394	144	114	93	745

★ = can be infeasible depending on initial positions

## Player annotations

- 324 of the transcripts include player ids.
- 202 distinct players



**Figure:** Number of games played by each player. The data points have been jittered randomly along the y-axis to make their clustering evident.

## Language in context

Each transcript is a data structure that is intuitively a list of temporally-ordered states

$$(\text{context, event})$$

The context includes

- local information (the state of play at that point)
- historical information (the events up to that point)
- global information (limitations of the game, the task, etc.)

When the event is an utterance, we can interpret it *in context*.

This is what pragmatics is all about, but it is very rare to have a dataset that truly lets you do it.

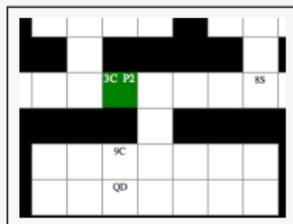
# QUD-based domain restriction

- 1 Question Under Discussion (QUD) models
- 2 The Cards corpus: scenario and implementation
- 3 Basic properties of the Cards corpus
- 4 QUD-based domain restriction**
- 5 Expert effects and rich common ground
- 6 Looking ahead

# Domain restriction

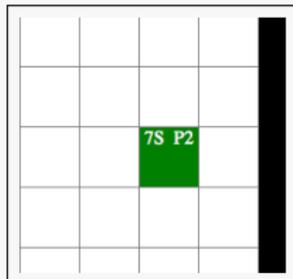
## Example

Player 1: lets do spades  
 Player 1: I have the as, qs, and ks  
 [...]  
 Player 1: ok, i found js  
 Player 2: Ok. I haven't found anything...lol



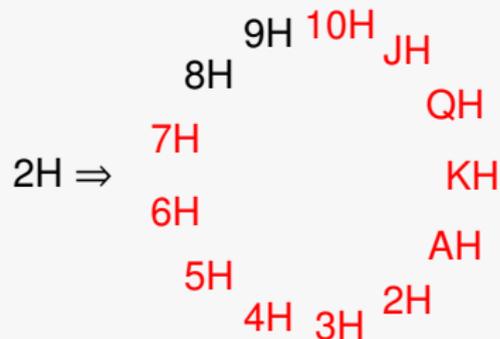
## Example

Player 1: what do u wont to look for?  
 Player 2: hearts?  
 Player 1: ok good let's get to busness  
 Player 1: i found KH!  
 [...]  
 Player 2: find ah  
 [...]  
 Player 1: u find anything else?  
 Player 2: no



## Relevance and the task

Example (Gather six consecutive cards of a particular suit.)



## Relevance and the task

Example (Gather six consecutive cards of a particular suit.)

**Context:** We're holding {4H, 5H}

2H ⇒

9H 10H JH  
8H QH  
7H KH  
6H AH  
5H  
4H 3H 2H

## Relevance for card manipulations

### Theory #1: costly dropping

For each player hand  $H$  (0-3 cards),  $\text{value}(H)$  is the minimum number of pick-up and drop moves to a solution:

$$\text{value}(\emptyset) = 6$$

$$\text{value}(5H) = 5$$

$$\text{value}(5H, 6H) = 4$$

$$\text{value}(5H, 2S) = 6$$

$$\text{value}(5H, 2S, 3D) = 7$$

## Relevance for card manipulations

### Theory #1: costly dropping

For each player hand  $H$  (0-3 cards),  $\text{value}(H)$  is the minimum number of pick-up and drop moves to a solution:

$$\begin{array}{ll} \text{value}(\emptyset) = 6 & \text{value}(5H, 2S) = 6 \\ \text{value}(5H) = 5 & \text{value}(5H, 2S, 3D) = 7 \\ \text{value}(5H, 6H) = 4 & \end{array}$$

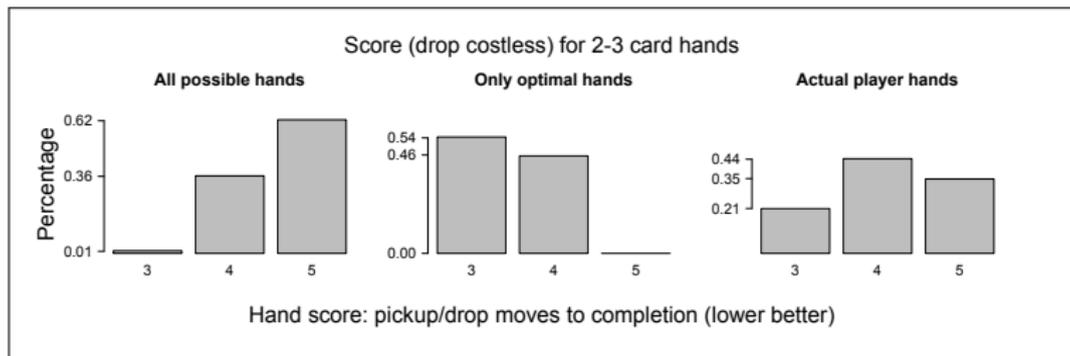
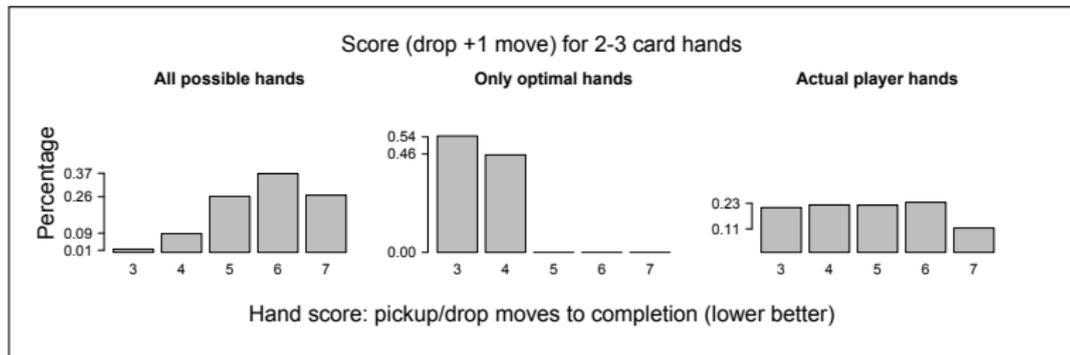
### Theory #2: costless dropping

For each player hand  $H$  (0-3 cards),  $\text{value}(H)$  is the minimum number of pick-up moves to a solution:

$$\begin{array}{ll} \text{value}(\emptyset) = 6 & \text{value}(5H, 2S) = 5 \\ \text{value}(5H) = 5 & \text{value}(5H, 2S, 3D) = 5 \\ \text{value}(5H, 6H) = 4 & \end{array}$$

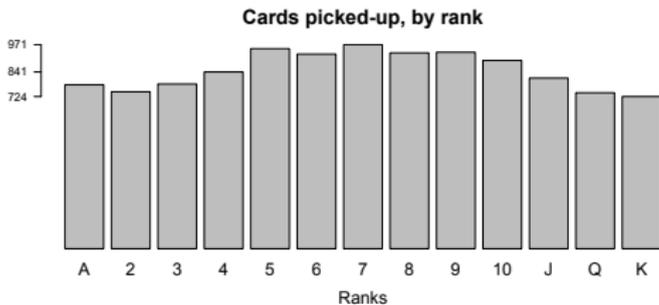
## Player card manipulations: what's their theory?

Random samples of 7,500 moves with the same 2-card/3-card balance as the empirical distribution:



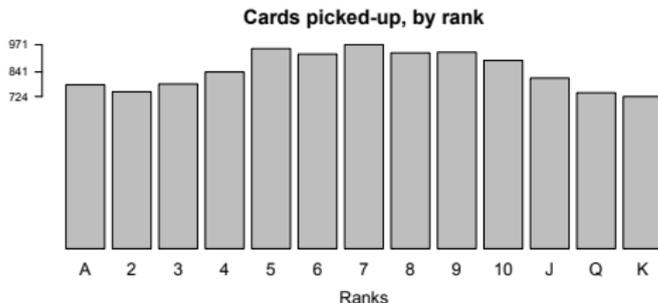
# Player card manipulations: what's their theory?

## Other factors



# Player card manipulations: what's their theory?

## Other factors



## Conjectures

- Structured strategies and collaborative decision making: “once we settle on hearts, we’re reluctant to switch”
- 8H perceived as more relevant to 2H than 6S.

## Underspecified referential expressions

### Goal

To use the (evolving) task to make educated guesses about what underspecified card-oriented nominals pick out.

Player 2: Look for 2.

Player 1: and the 3?

### Hypothesis

For any nominal referring expression, the intended referent will be the one that is (i) consistent with the information specified; and (ii) would bring the players closest to a solution *given the cards they are holding in the context of utterance*.

## Underspecified referential expressions

### Goal

To use the (evolving) task to make educated guesses about what underspecified card-oriented nominals pick out.

**The players are holding {4H,KH}**

Player 2: Look for 2.

Player 1: and the 3?

### Hypothesis

For any nominal referring expression, the intended referent will be the one that is (i) consistent with the information specified; and (ii) would bring the players closest to a solution *given the cards they are holding in the context of utterance*.

# Annotations

Nominals referring to cards:

`[_FEATURES text ]_{DENOTATION}`

- I have `[_3H 3H ]_{3H}`
- Need `[_8 8 ]_{8H}`
- I'll drop `[_9|DEF|SG the 9 ]_{9H}`
- try `[_H|INDEF|PL h ]_{2H, 3H, 4H, 5H, 6H, 7H, 8H, 9H, 10H, JH, QH, KH}`
- got `[_X|PRO|SG it ]_{9H}`
- i'll look around to see if i can find `[_X|INDEF|PL any you can pick up ]_{2H, 3H, 4H, 5H, 6H, 7H, 8H, 9H, 10H, JH, QH, KH}`

## Results (for 10 transcripts)

Resolving the reference of the singular definite card references:

- [-(SUIT|RANK)DEF|SG the (SUIT|RANK) ]-{CARD}
- [- (SUIT|RANK) ]-{CARD}

---

Literal	103	(37%)
Requiring enrichment	172	(63%)
Total	275	

---

---

Correct inference	164	(95%)
Incorrect inference	8	(5%)
Total	172	

---

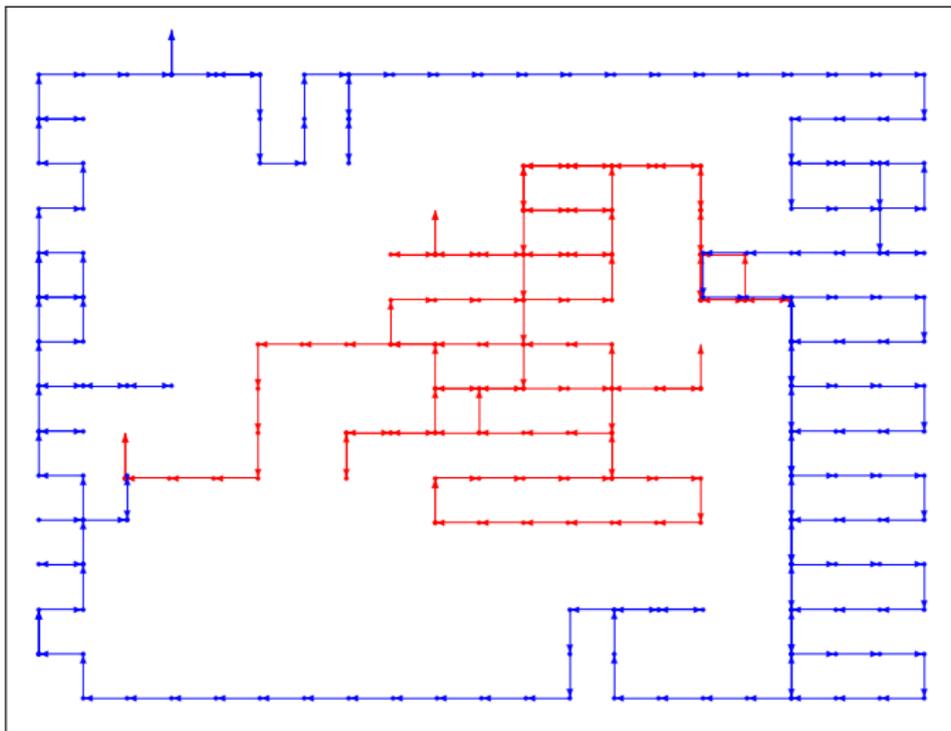
- The mistakes are localized: strings of card references that the system botches uniformly.
- Most inferences involve guessing the suit based on the rank, which is easier than guessing the rank based on the suit.

# Expert effects and rich common ground

- 1 Question Under Discussion (QUD) models
- 2 The Cards corpus: scenario and implementation
- 3 Basic properties of the Cards corpus
- 4 QUD-based domain restriction
- 5 Expert effects and rich common ground**
- 6 Looking ahead

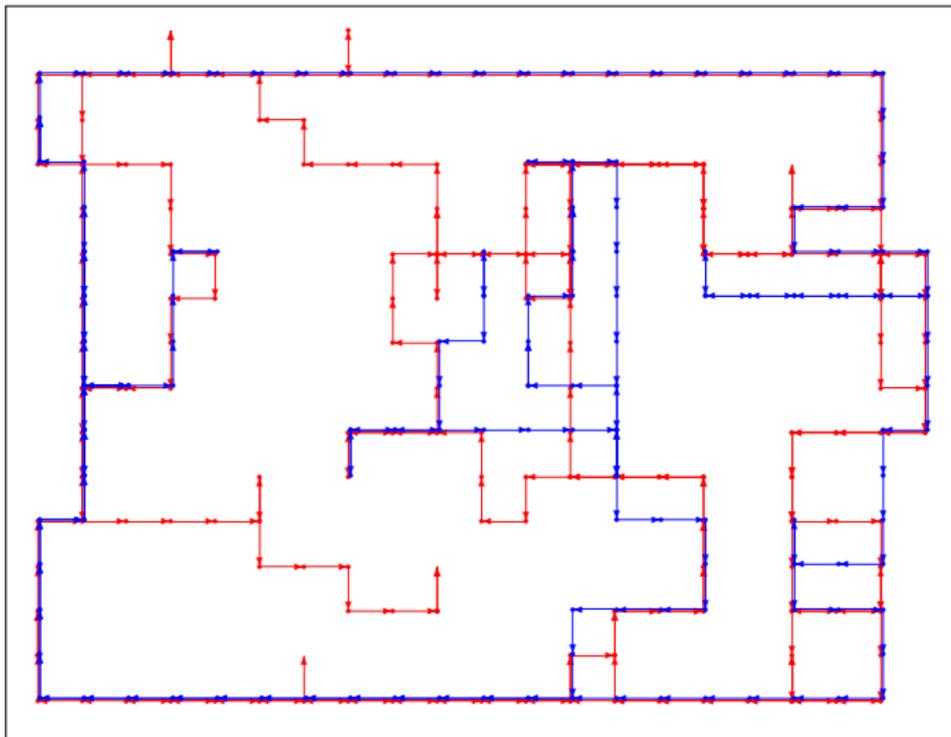
# Paths

## Experts



# Paths

## Novices



## Novice strategy

Player 1: Hello. Are you here?

Player 2: yes

Player 2: do you see any cards

Player 1: Yes. I see a yellow spot. Those are our cards. We'll only be able to see the ones that are in our view

Player 1: until we move with our arrows.

Player 2: i see 3 of them

Player 1: We only have a certain number of moves, so we should decide how we're going to do this before we use them, do you think?

Player 2: sure

Player 1: Ok. So, we have to pick up six cards of the same suit, in a row...

Player 1: each of us can hold three, so...

Player 1: I think I should get my three, then you should get your three or vice versa

Player 2: ok

Player 2: you go ahead

Player 1: What suit should we do?

Player 1: And which six cards do you want to try for?

Player 2: whatever you want

Player 1: I'm Courtney, by the way- nice to meet you.

Player 2: i'm becky....nice to meet you too

Player 1: Hi Becky. How about we go for hearts? And take 234567  
[...]

## Journeyman strategy

These players have explored and are now forming a strategy:

Player 1 I have 9 clubs and K clubs

Player 1 want to look for clubs?

Player 2 ok

[...]

The players then find various clubs, checking with each other frequently, until they gain an implicit understanding of which specific sequences to try for (either 8C-KC or 9C-AC):

Player 1 so you are holding Jc and Kc now?

Player 2 i now have 10d JC and KC

Player 2 yes

Player 1 drop 10d and look for either 8c or Ace of clubs

# Expert strategy

Player 2: hi

Player 1: hi--which side r u on?

Player 2: right side

Player 2: u?

Player 1: left/middle

Player 1: ok i gathered everything in my area

Player 2: i think i have all of them also

Player 1: how bout 5C - 10C?

Player 2: ok

Player 1: i have 5C, 8C, 9C, and you should have 6C, 7C, 10C

Player 2: got them

# Depth annotations

## Depth 1

How do I interact with the game world?

What are the meanings of the various technical terms?

## Depth 2

What is the goal of the game generally?

## Depth 3

What is the configuration of the game board?

What is the expertise of my fellow player?

## Depth 4

What is the goal of this game specifically?

## Depth 5

How do we achieve this goal generally?

## Depth 6

What cards do we need to achieve this goal specifically?

## Depth 7

Have we completed the game?

P2: what suit do we want? (4)

P1: I hit a KD. (3)

I think we should see what we get,  
and keep the most promising suit (3)

P2: i have a JD (3)

P1: That works

P2: so we are looking for Ds? (4)

P1: I vote Ds. (4)

P2: okay i have 10D , 9D and JD (6)

P1: 7D (6)

P1: okay do you think my cards work? (5)

P1: So we're looking for 8D,  
and 6D or QD (6)

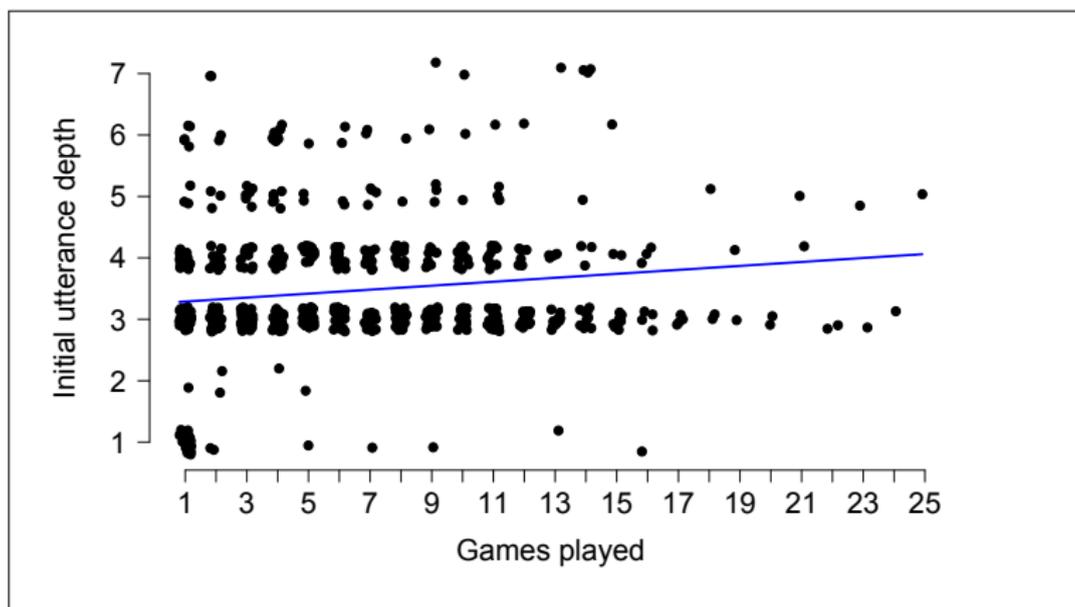
P2: You should be good (I'm slow at  
this...) (6)

## Initial depth and rich common ground

- We expect experts to be more likely than novices to assume that their partners will accommodate rich contextual knowledge.
- This predicts that experts will initiate discourse at a deeper level in the question graph than novices.
- Furthermore, these effects should be amplified if it is mutual knowledge between the two players that they are both experts.

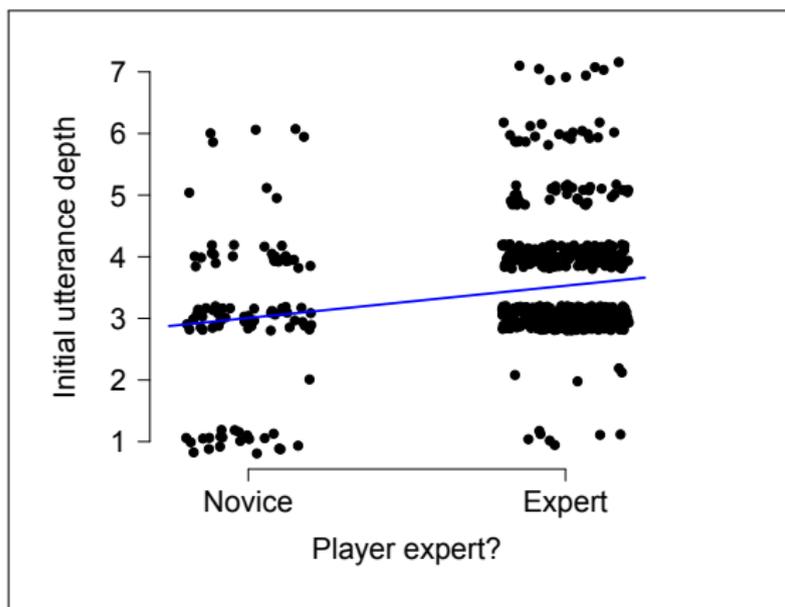
## Results

Greater number of games played (greater expertise) reliably correlates with greater initial utterance depth, though the effect is weak, presumably because one game suffices to become an expert.



## Results

Let **Expert** be true of a player–game pair  $(P, T)$  iff  $T$  is not  $P$ 's first game. **Expert** is a very highly correlated with greater initial utterance depth in  $T$ .





# Looking ahead

- 1 Question Under Discussion (QUD) models
- 2 The Cards corpus: scenario and implementation
- 3 Basic properties of the Cards corpus
- 4 QUD-based domain restriction
- 5 Expert effects and rich common ground
- 6 Looking ahead**

## Variations

- 1 **Asymmetry**: player A has 300 moves but limited line-of-sight; player B has 20 moves but can see everything.
- 2 **Shared resources**: the players share a store of 500 moves, so inefficient movement incurs a social cost (hat-tip to Florian).
- 3 **Bonuses**: based on move counts and other notions of efficiency that might maximize verbal interaction.

## Additional annotations

- 1 Card-reference annotations for the whole corpus.
- 2 Depth annotations for full transcripts.
- 3 Annotating for semantic ellipsis – e.g., “2S” to `seeking(2S)`, `found(2S)`, etc.
- 4 ...

## General theoretical goals

- Distinguish the models of discourse on offer.
- Distinguish theories of issues/questions/goals.
- Inform proposals about how context-dependence is resolved.
- Get semprag people excited about working with corpora.

<http://CardsCorpus.christopherpotts.net/>