Modeling Dialogue

Building Highly Responsive Conversational Agents

ESSLLI 2016

David Schlangen, Stefan Kopp
with Sören Klett
CITEC // Bielefeld University

Universität Bielefeld

CITEC
Cognitive Interaction Technology
Center of Excellence
Bielefeld University

Who we are

- Stefan Kopp, Professor for Computer Science, Faculty of Technology, Uni. Bielefeld (stefan.kopp@uni-bielefeld.de)
- Head of research group *Social Cognitive Systems* at CITEC, U. Bielefeld
- Research interests:
  - understanding social minds and their interaction
  - adaptive and responsive conversational agents
  - multimodal communication
- http://scs.techfak.uni-bielefeld.de

Who we are

- Sören Klett, Ph.D. student at *Social Cognitive Systems group* at Uni. Bielefeld, (`sklett@techfak.uni-bielefeld.de`)
- research on user-adaptive decision-making in dialogue systems
- developed and prepared toolkit you will be using in this course, here to provide technical support

Who we are

- David Schlangen, Professor for Applied Computational Linguistics, Uni Bielefeld. (david.schlangen@uni-bielefeld.de)
- Lead *Dialogue Systems Group* at Bielefeld / CITEC.
- Research Interests:
  - “understanding understanding”
    - highly responsive dialogue systems / incremental processing
    - grounded semantics
- http://www.dsg-bielefeld.de

Who are you?

- show of hands:
  - undergrad, master, post-grad, beyond
  - familiarity with dialogue theory?
  - Timo & Arne's class in week 1?
  - Experience with building dialogue systems / conv. agents?

Modeling Dialogue

Building Highly Responsive Conversational Agents

David Schlangen, Stefan Kopp
with Sören Klett
CITEC // Bielefeld University

Universität Bielefeld

CITEC
Cognitive Interaction Technology
Center of Excellence
Bielefeld University

Modeling Dialogue

Building Highly Responsive Conversational Agents

Responsive Agents

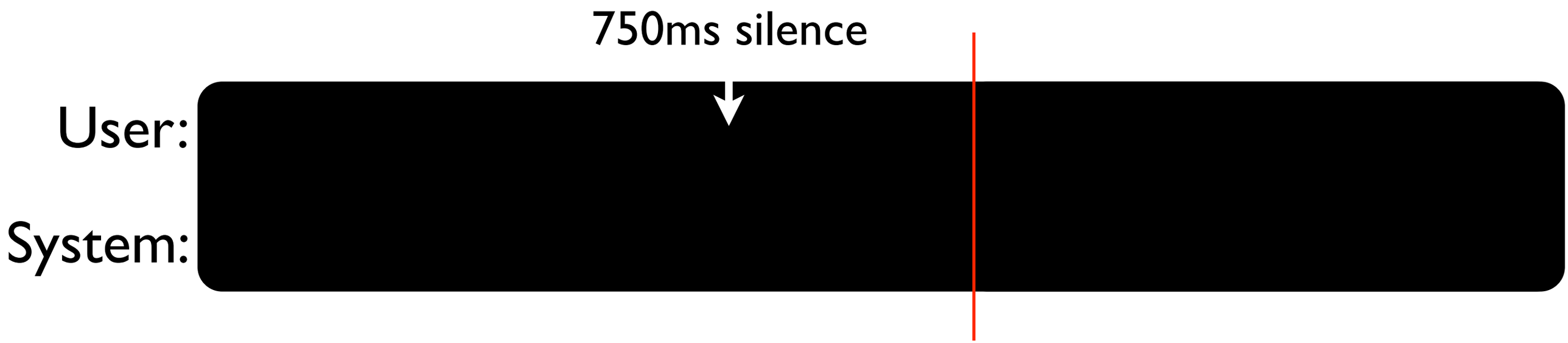
- working definition:
  - are responsive to the needs of the dialogue partner(s), *at all times*
  - minimize <u>time</u> between *event* and *response*

“Traditional” Approach

- only optimize <u>coherence</u> between *event* and *response*

- event and response are full speech acts.

the status quo: *non-incremental* processing



A:

B:

A:

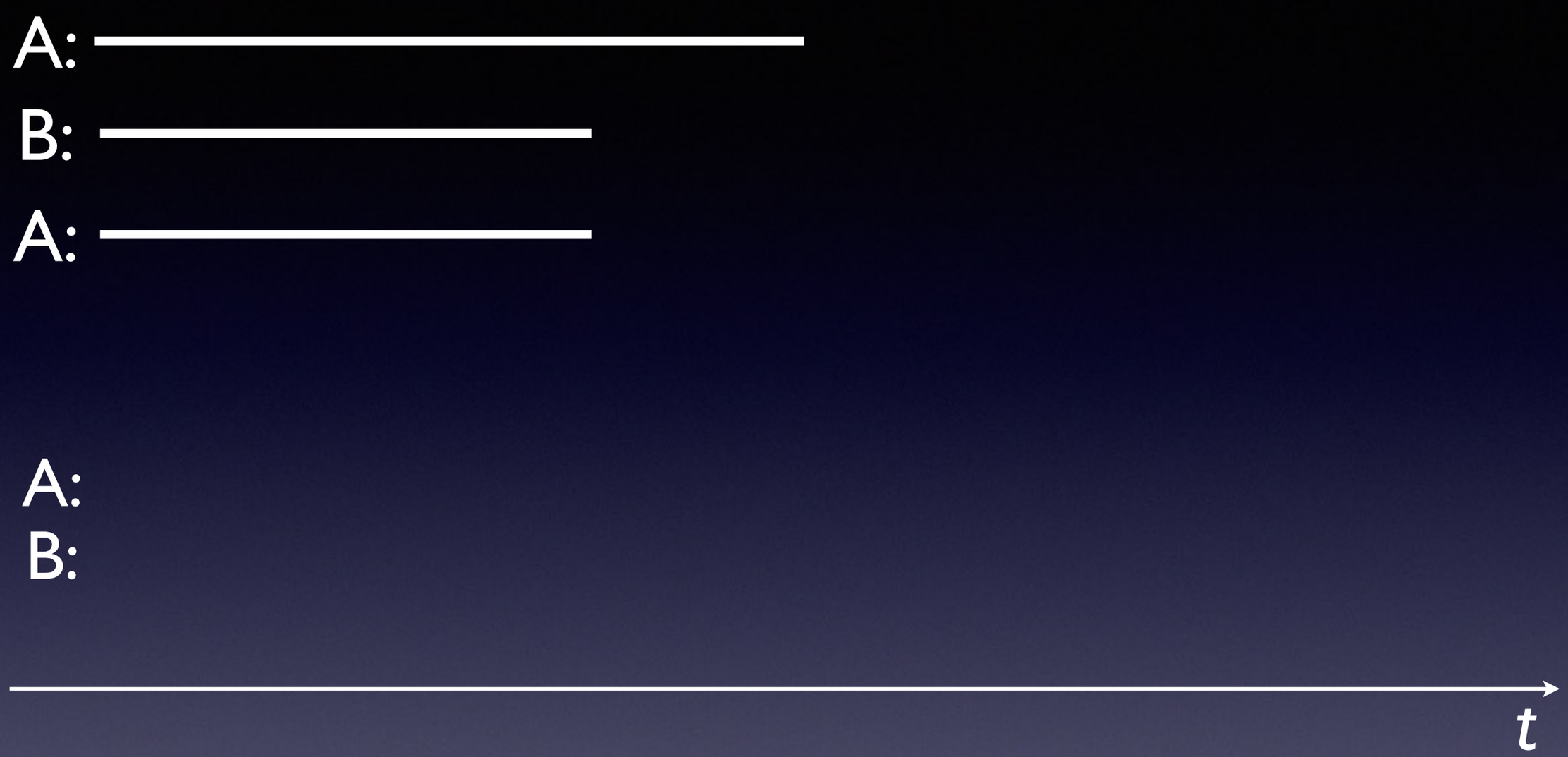
B:

A:

B:

A:

B:


A:
B:
A:
A:
B:
t


A:
B:
t

Responsive Agents

- working definition:
  responsive to needs of dialogue partner(s)
  minimize <u>time</u> between *event* and *response*

- Qs:
  - why?
  - how?
  - what needs?
  - what type of events?
  - which types of responses?
  - who / what creates these events?
  - does an event have to have occurred to respond to it?
  - what are the optimization criteria?

Overview of Course

- Day 1: Motivation, Phenomena, State of the Art
- Day 2: Technical Challenges, Approaches
- Day 3: Introduction to Task & Technical Framework
- Day 4: Hands-On Exercises
- Day 5: Reports, Discussion

Modeling Dialogue

Building Highly Responsive Conversational Agents

Day 1: Motivation, Phenomena, Theoretical Terms

David Schlangen, Stefan Kopp
with Sören Klett
CITEC // Bielefeld University

Universität Bielefeld

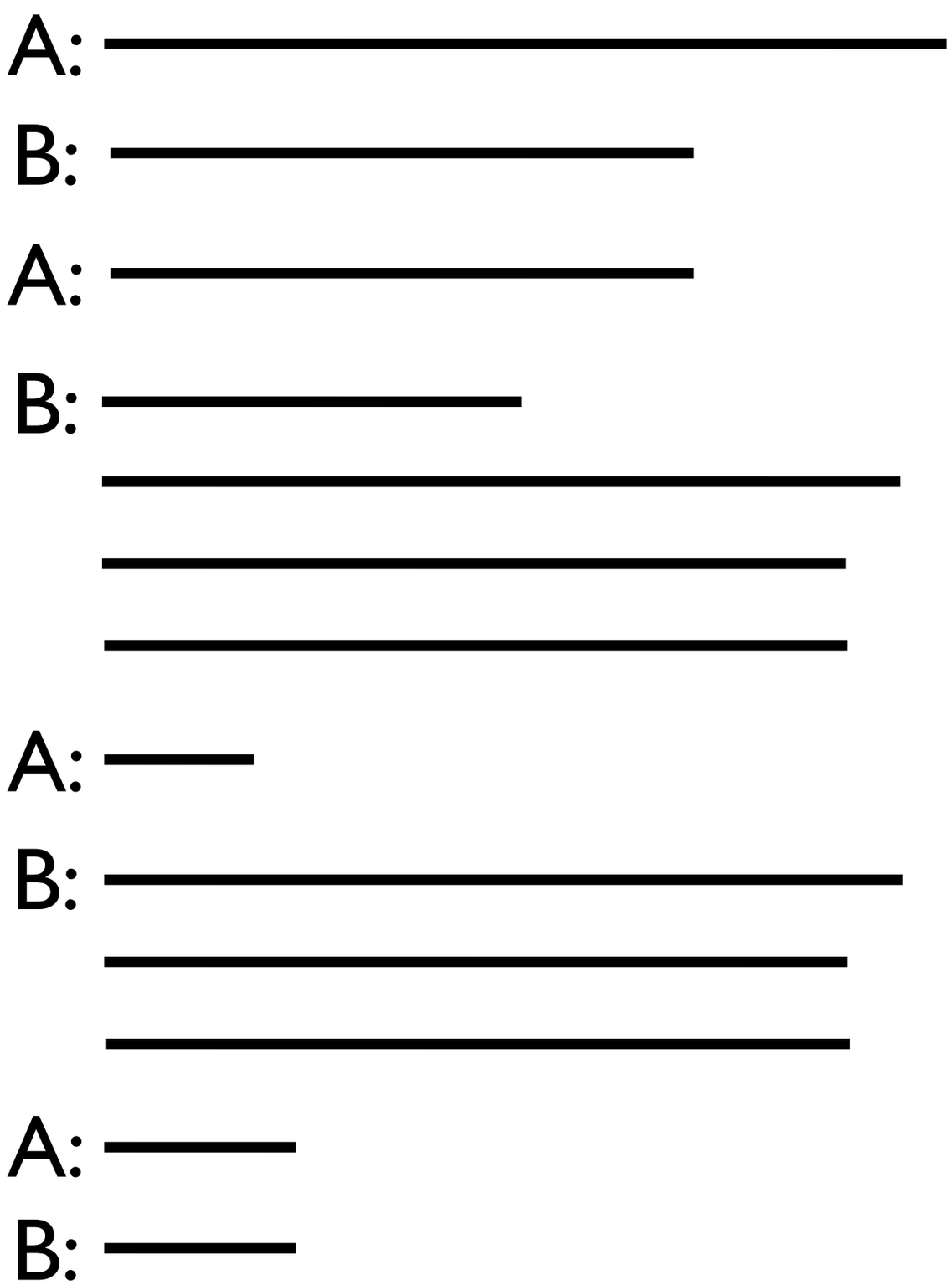
CITEC
Cognitive Interaction Technology
Center of Excellence
Bielefeld University

Overview of Day 1

- What does responsiveness mean here?
- What do people do in dialogues?
- Dialogue as coordinated, joint action / as process.
  - Grounding, Turn-Taking, etc.
- State of the art in responsive conversational agents

Example Datum

- Pentomino/Noise Corpus, 2006; (Fernández & Schlangen 2006; Zarrieß *et al.* LREC 2016)
- 3:05 — 5:02 in `20161123_run1_pento`
- using the wonderful ELAN annotation tool (https://tla.mpi.nl/tools/tla-tools/elan/)



In what sense responsive to needs of partner?

Orderly sequence of contributions?

P so basically okay draw your eye from the bottom of the backwards L?

reference in installments

E yeah? okay?

P go to the left the first square you come to?

E yeah? okay! alright I got it.

P that's where the bottom of the long twin-tower piece goes.

levels of understanding

E okay

E alright I got it yeah I'm putting it in there right now

E it is in there.

acknowledgement of acknowledgement

P good

P there is the straight line from the top down?
E yeah

P fit it all the way to the bottom and it should be: ehm

interruption, realises
E pff oh I have to flip it then own misunderstanding
P then you must flip it yeah

E yeah
P so the angle would be eh pointing I guess to the right

E okay I got that..
P the open part you got that? now then

E wait i'm sticking it in there right now okay
P okay

P (and then it + the top of the T) fits (into: + next to) the first piece

self correction

P where the L is the backwards L

E the top of the T fits next to the first piece?

P yeah

P first piece that you put in was the backwards L?

E yeah yeah

P all the way on the bottom right?

P and then the top of the T fits into lets say the lap of the L

E eh unfortunately not.

P no?

E <laughter/> no! it will overlap with the first piece.

P okay.

P (and then it + the top of the T) fits (into: + next to) the first piece

lack of uptake → expansion

P where the L is the backwards L

E the top of the T fits next to the first piece?
P yeah

P first piece that you put in was the backwards L?

E yeah yeah
P all the way on the bottom right?

P and then the top of the T fits into lets say the lap of the L

E eh unfortunately not.
P no?

E <laughter/> no! it will overlap with the first piece.
P okay.

P (and then it + the top of the T) fits (into: + next to) the first piece

P where the L is the backwards L

E the top of the T fits next to the first piece?

P yeah

P first piece that you put in was the backwards L?

E yeah yeah

P all the way on the bottom right?

P and then the top of the T fits into lets say the lap of the L

E eh unfortunately not.

laughter events

P no?

E <laughter/> no! it will overlap with the first piece.

P okay.

A second example

- (Kimbara 2007, U. Chicago)
- multimodal co-completion

Observations

- reference in installments
- signal level of understanding
- (invited?) interruption; continuation
- self corrections (= self interruption)
- expand until successful
- completion by partner

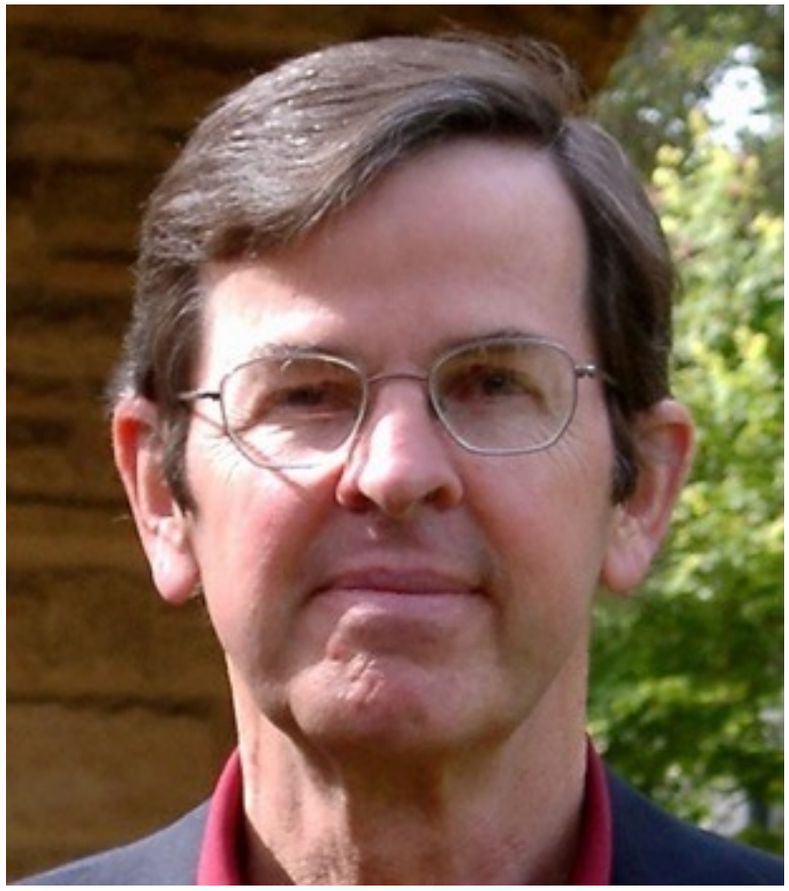
But *why* do people do that, and why should we model that in practical systems?

Overview of Day 1

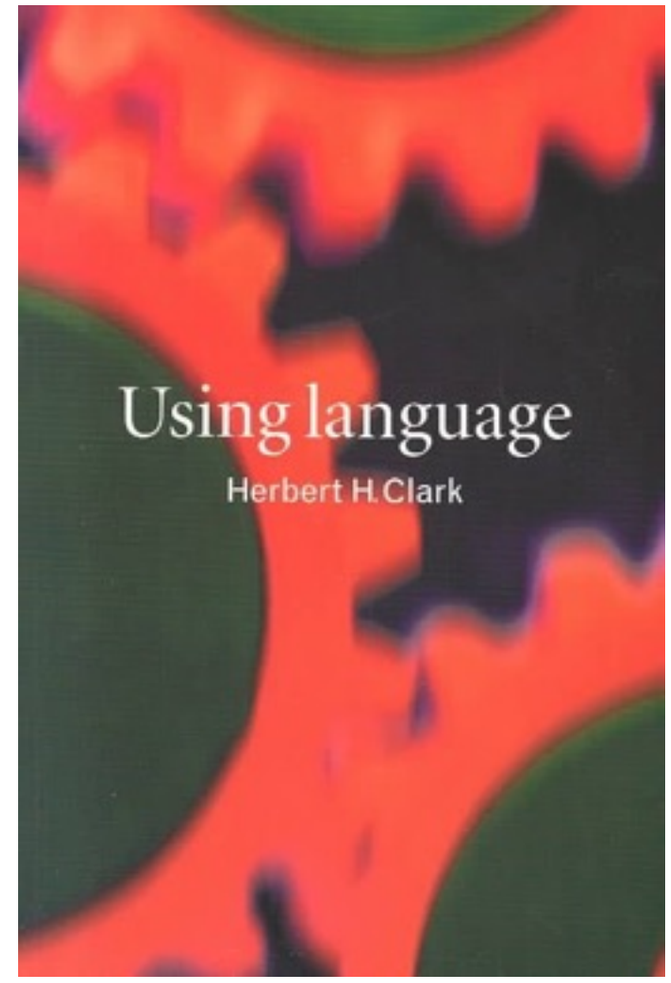
- What does responsiveness mean here?
- What do people do in dialogues?
- Dialogue as coordinated, joint action / as process.
  - Grounding, Turn-Taking, etc.
- State of the art in responsive conversational agents

Spoken Dialogue

- Uses evanescent medium.
- Consists of spontaneously and autonomously produced contributions.
- Participants want to understand and be understood.
- Need to coordinate what they are doing.



Herb Clark



(Clark, 1996)

synthesising much of what was originally researched in the field of conversation analysis (Sacks, Schegloff, Jefferson & others, 1960s ff)

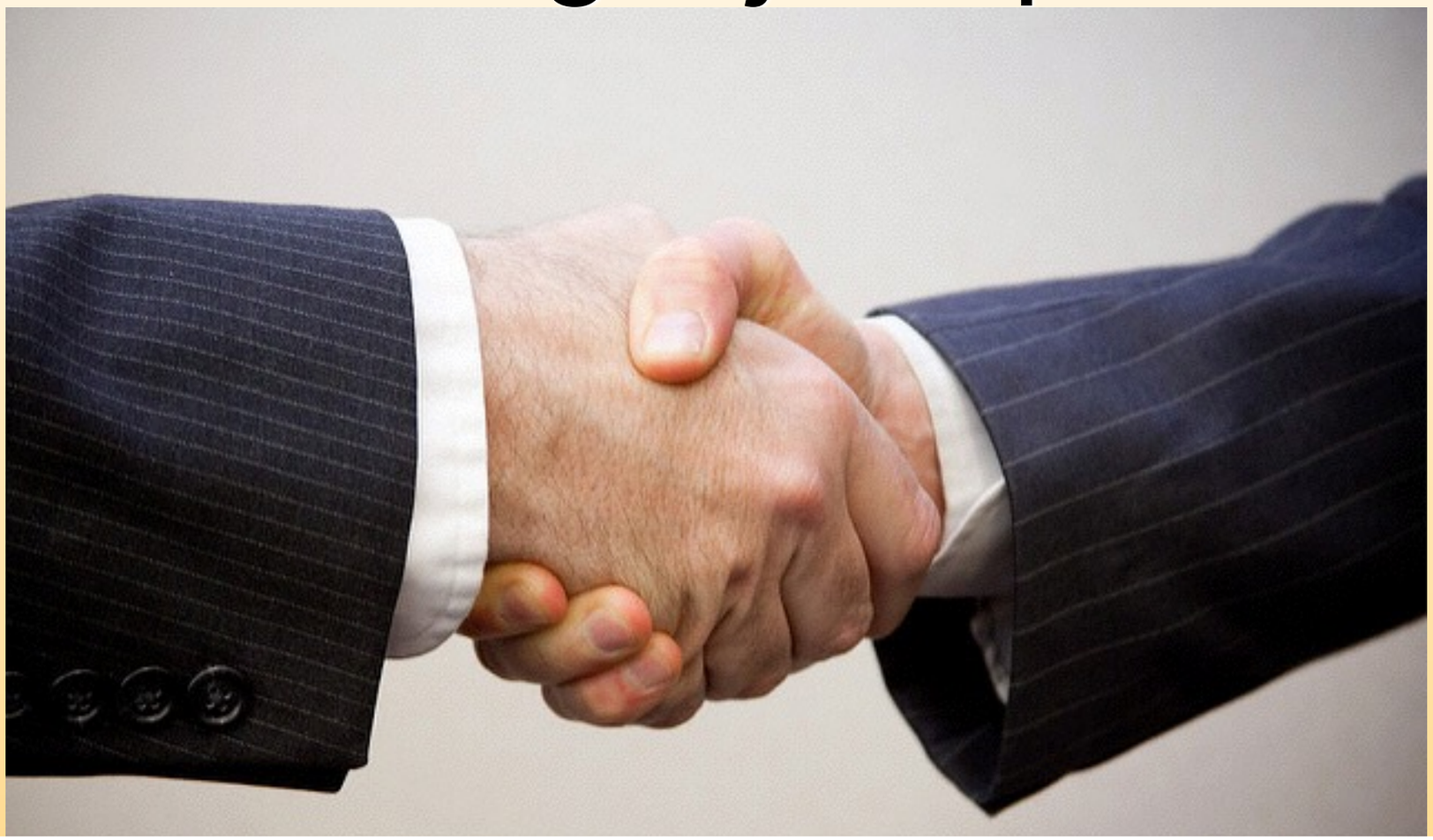
Dialogue as joint process

- From dialogue as exchange of propositions to dialogue as joint process aimed at creating mutual understanding about joint projects.
    - joint action in dialogue
    - temporal coordination

Dialogue as joint process

- From dialogue as exchange of propositions to dialogue as joint process aimed at creating mutual understanding about joint projects.
  - joint action in dialogue
  - temporal coordination

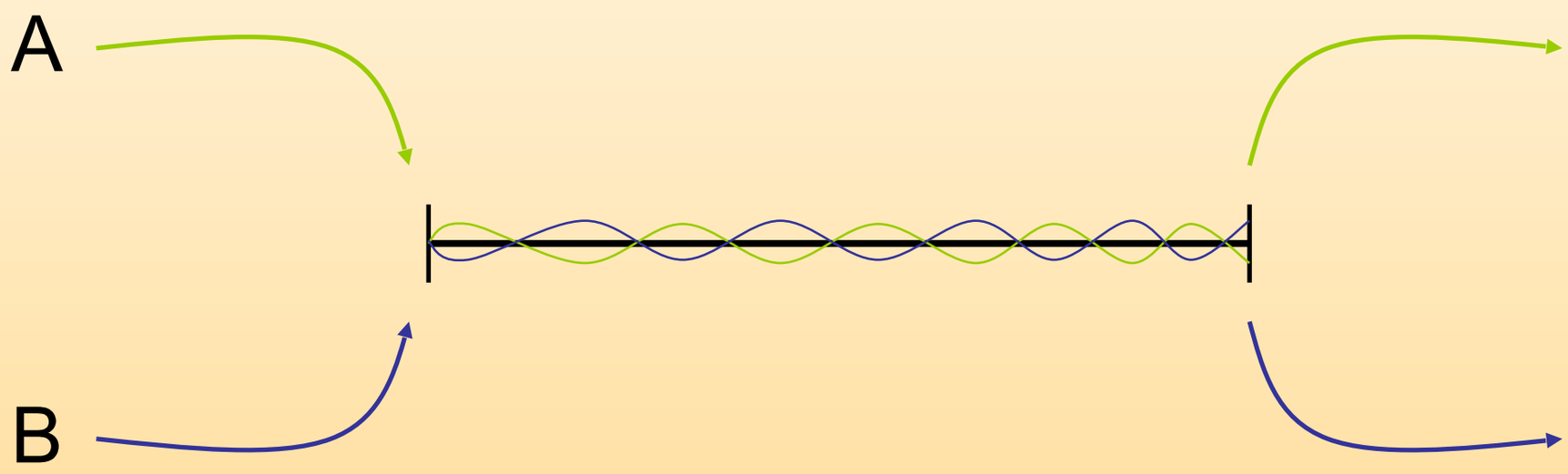
coordinating a joint process



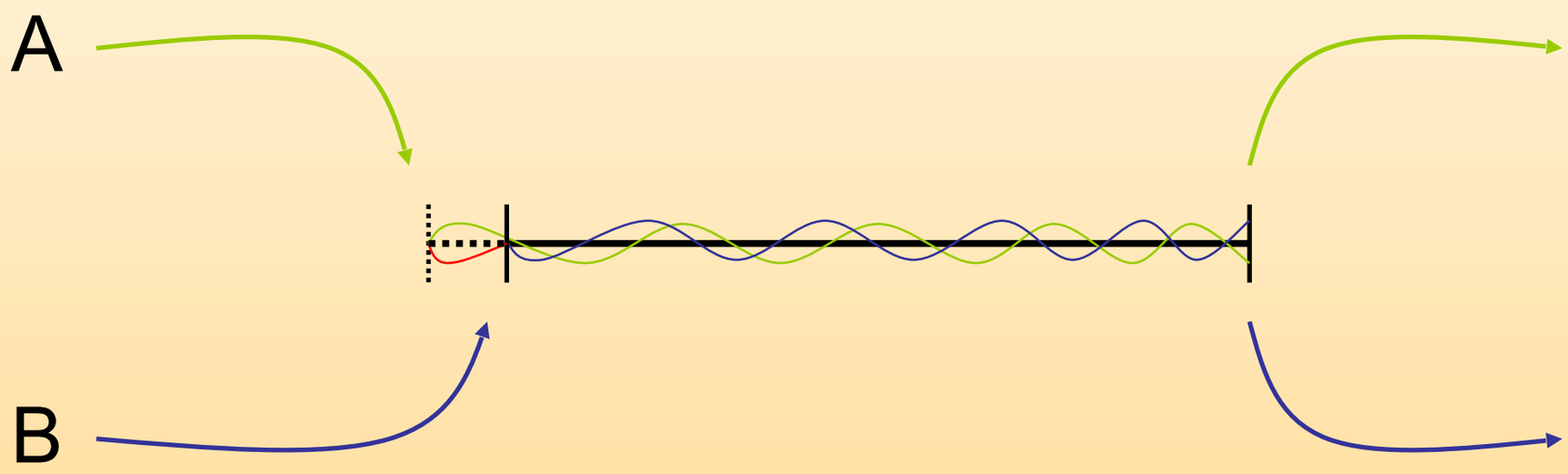
- what needs to be coordinated here?
  - beginning / entry, main part, end / exit

https://www.flickr.com/photos/124247024@N07/13903385550 www.flazingo.com

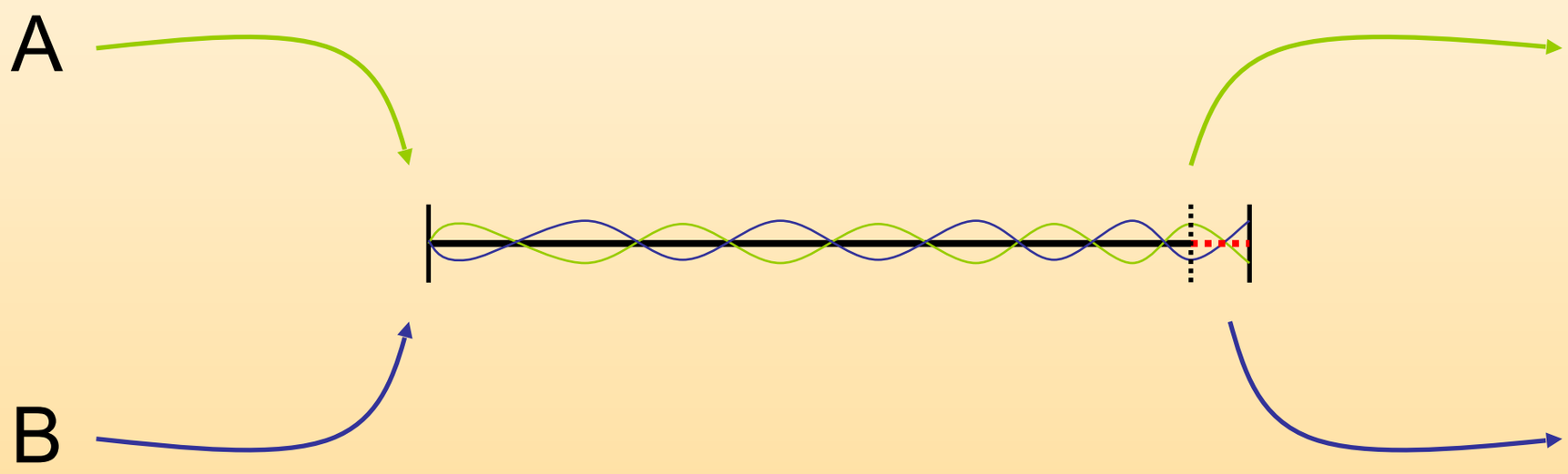
coordinating a process



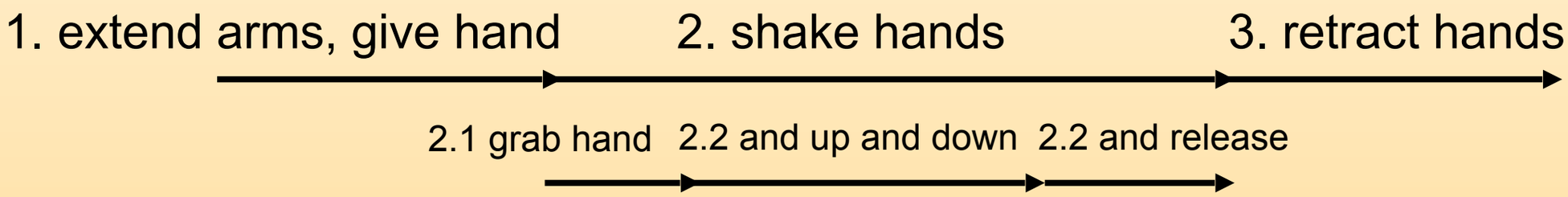
coordinating a process



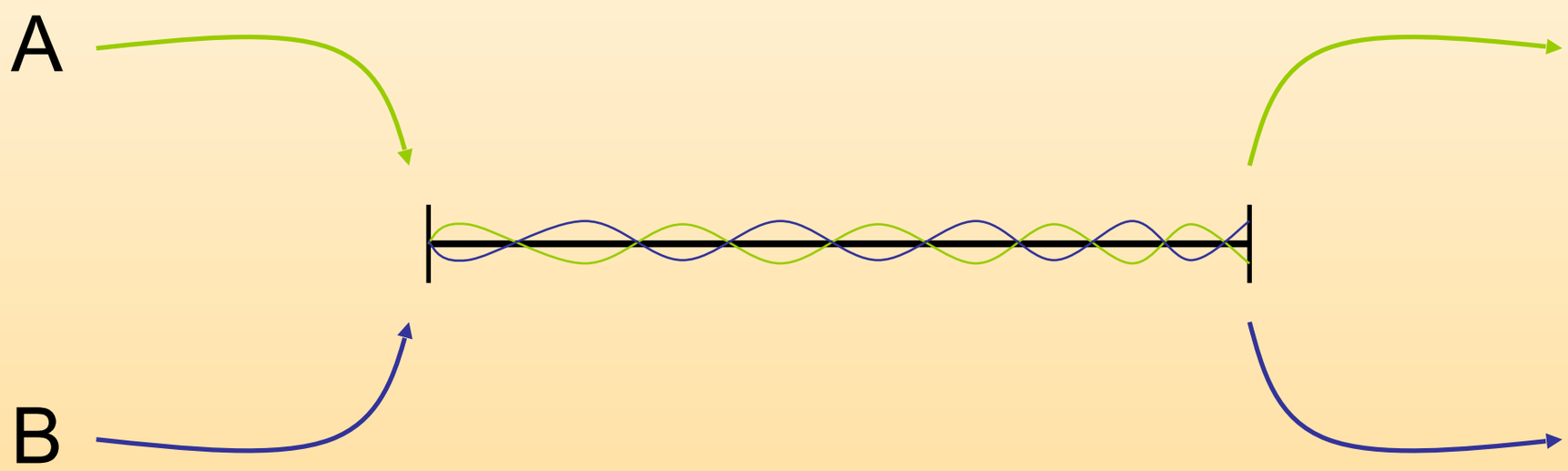
coordinating a process



shaking hands

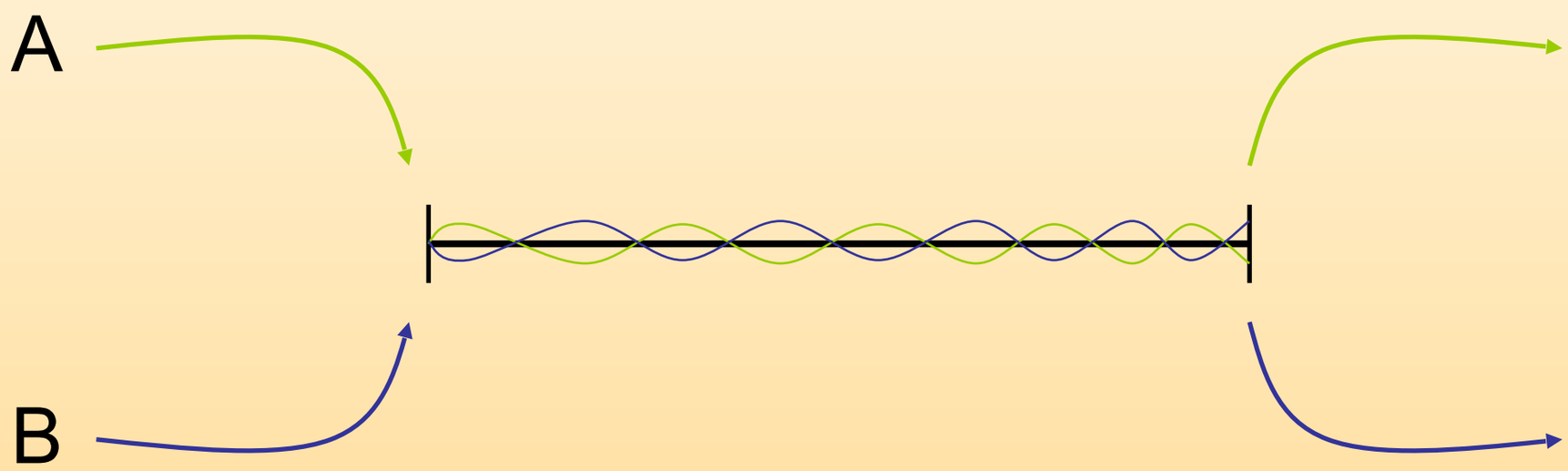


coordinating a process



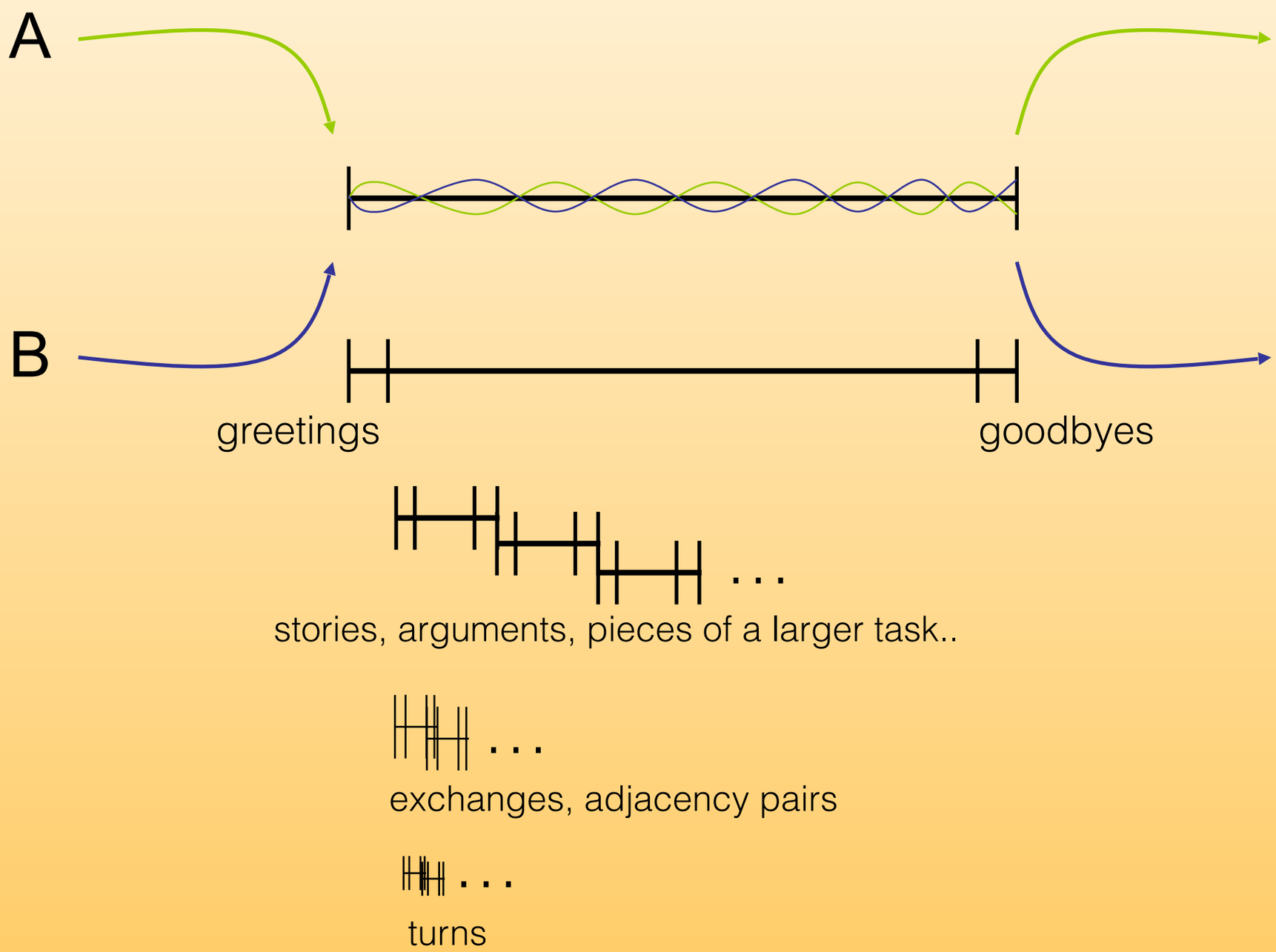
- what needs to be coordinated, and how?
  - beginng / entry:
    - as successor of previous action sequence
  - main part
    - who's doing what?
  - end / exit: when to stop

coordinating a process



- coordination devices:
  - one party leads (e.g., dancing)
  - external beat (e.g., dancing, playing music)
  - convention (e.g., shaking hands)
  - predictability (e.g., language?)

dialogue as a process



P so basically okay draw your eye from the bottom of the backwards L?

E yeah? okay?

P go to the left the first square you come to?

E yeah? okay! alright I got it.

P that's where the bottom of the long twin-tower piece goes.

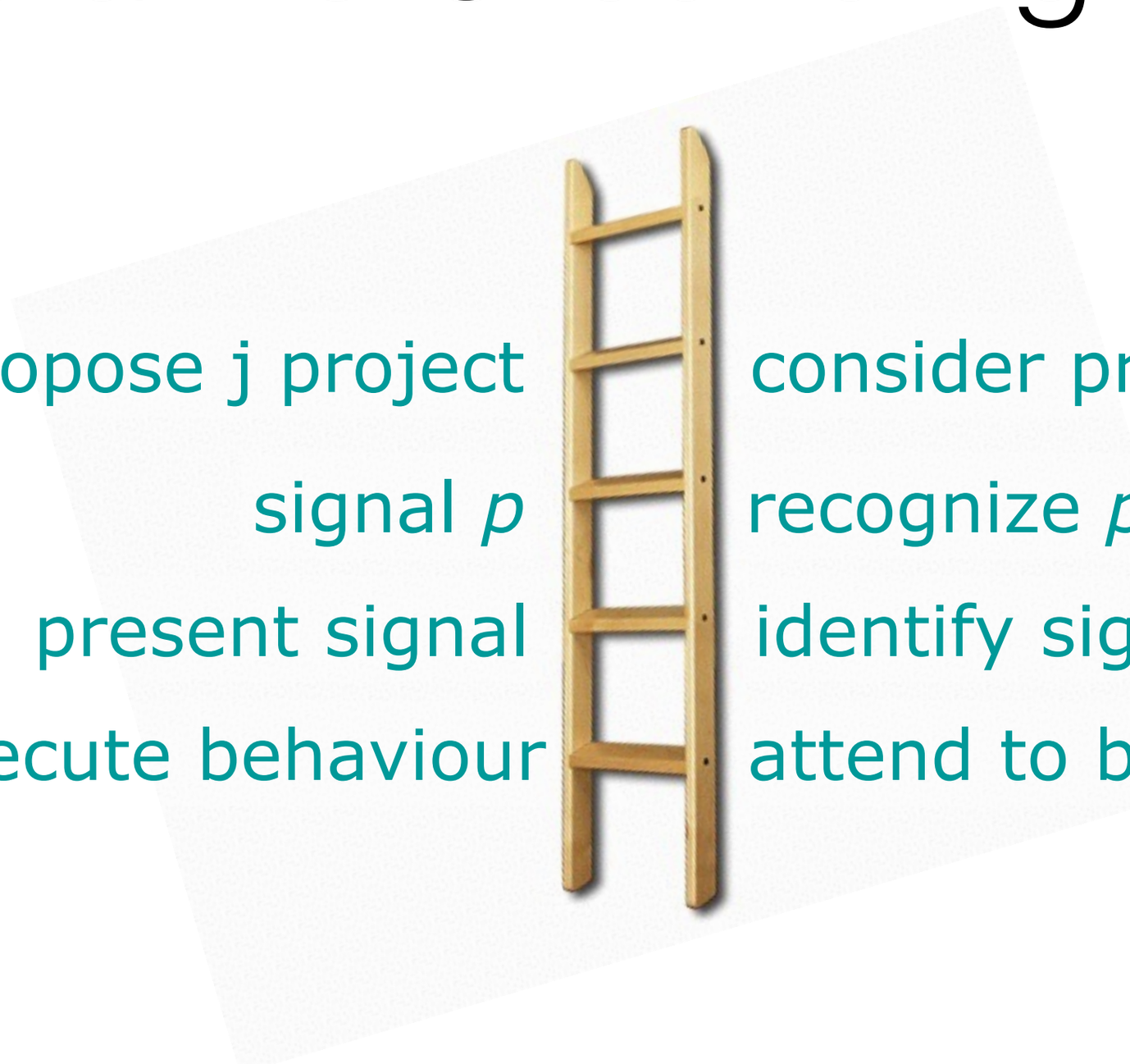
P (and then it + the top of the T) fits (into: + next to) the first piece

P where the L is the backwards L

Dialogue as joint process

- From dialogue as exchange of propositions to dialogue as joint process aimed at creating mutual understanding about joint projects.
  - joint action in dialogue
  - temporal coordination

H. Clark's Grounding Model



propose j project	consider proposal
signal *p*	recognize *p*
present signal	identify signal
execute behaviour	attend to behaviour

(Clark 1996; Clark & Wilkes-Gibbs 1986)

She is getting the elevator to come

She is calling the elevator

She is activating the “up” button

She is pressing the “up” button

She is pressing her finger against the “up” button

She is getting the elevator to come

She is calling the elevator

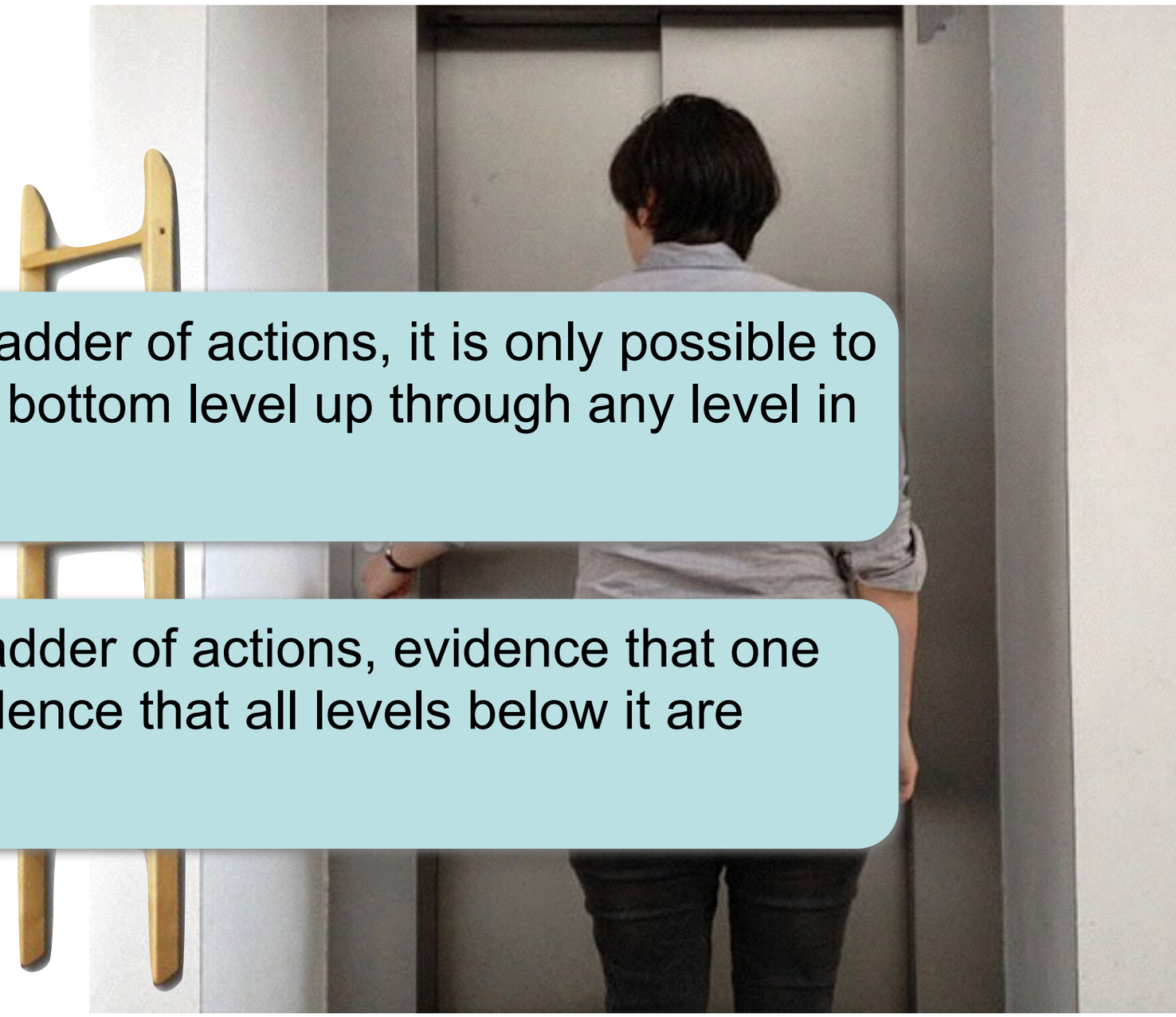
She

butto

She

butto

She is pressing her finger against the call button

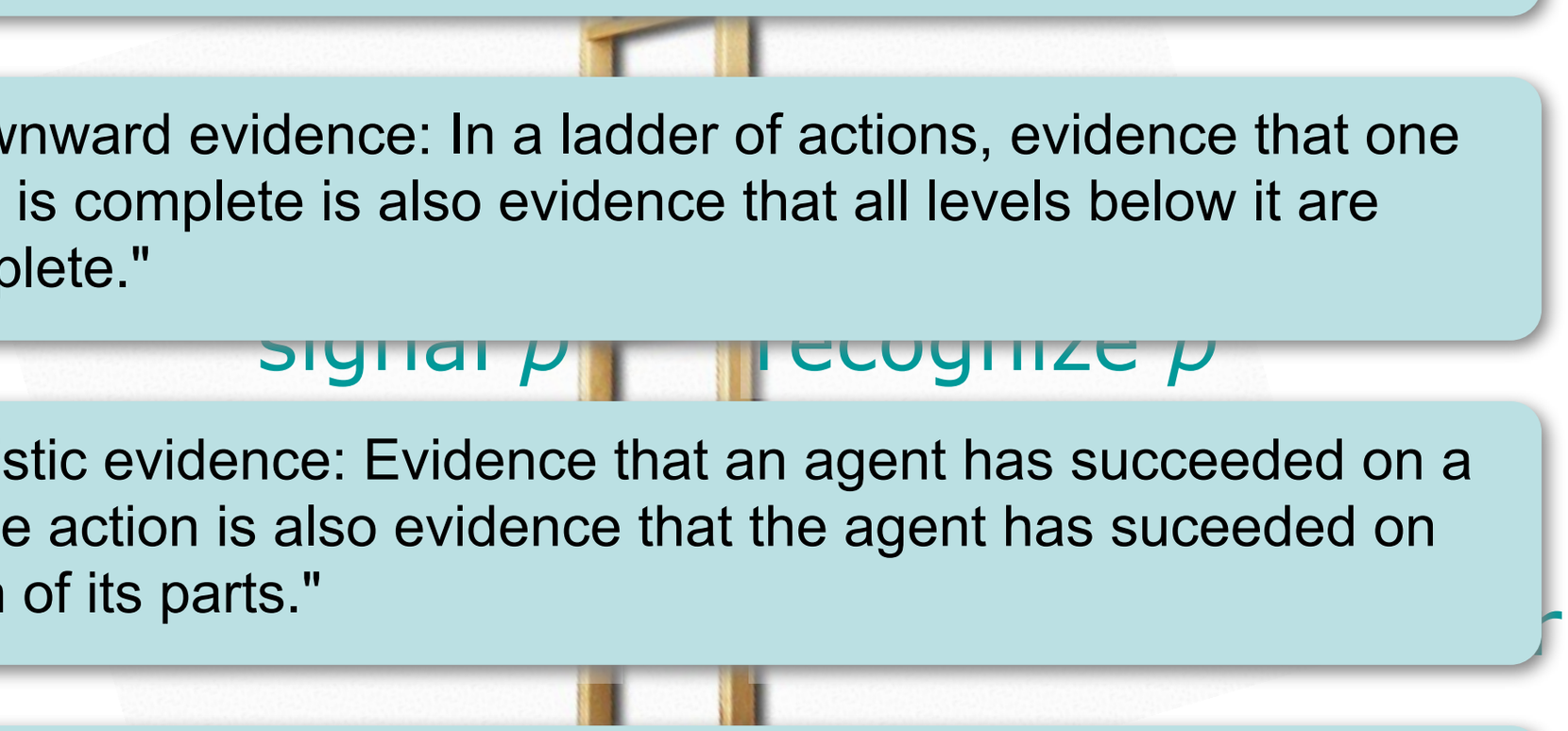


"Upwards Completion: In a ladder of actions, it is only possible to complete an action from the bottom level up through any level in the ladder."

"Downward evidence: In a ladder of actions, evidence that one level is complete is also evidence that all levels below it are complete."

H. Clark's Grounding Model

"Upwards Completion: In a ladder of actions, it is only possible to complete an action from the bottom level up through any level in the ladder."



"Downward evidence: In a ladder of actions, evidence that one level is complete is also evidence that all levels below it are complete."

"Holistic evidence: Evidence that an agent has succeeded on a whole action is also evidence that the agent has suceeded on each of its parts."

"Principle of joint closure: The participants in a joint action try to establish the mutual belief that they have succeeded well enough for current purposes."

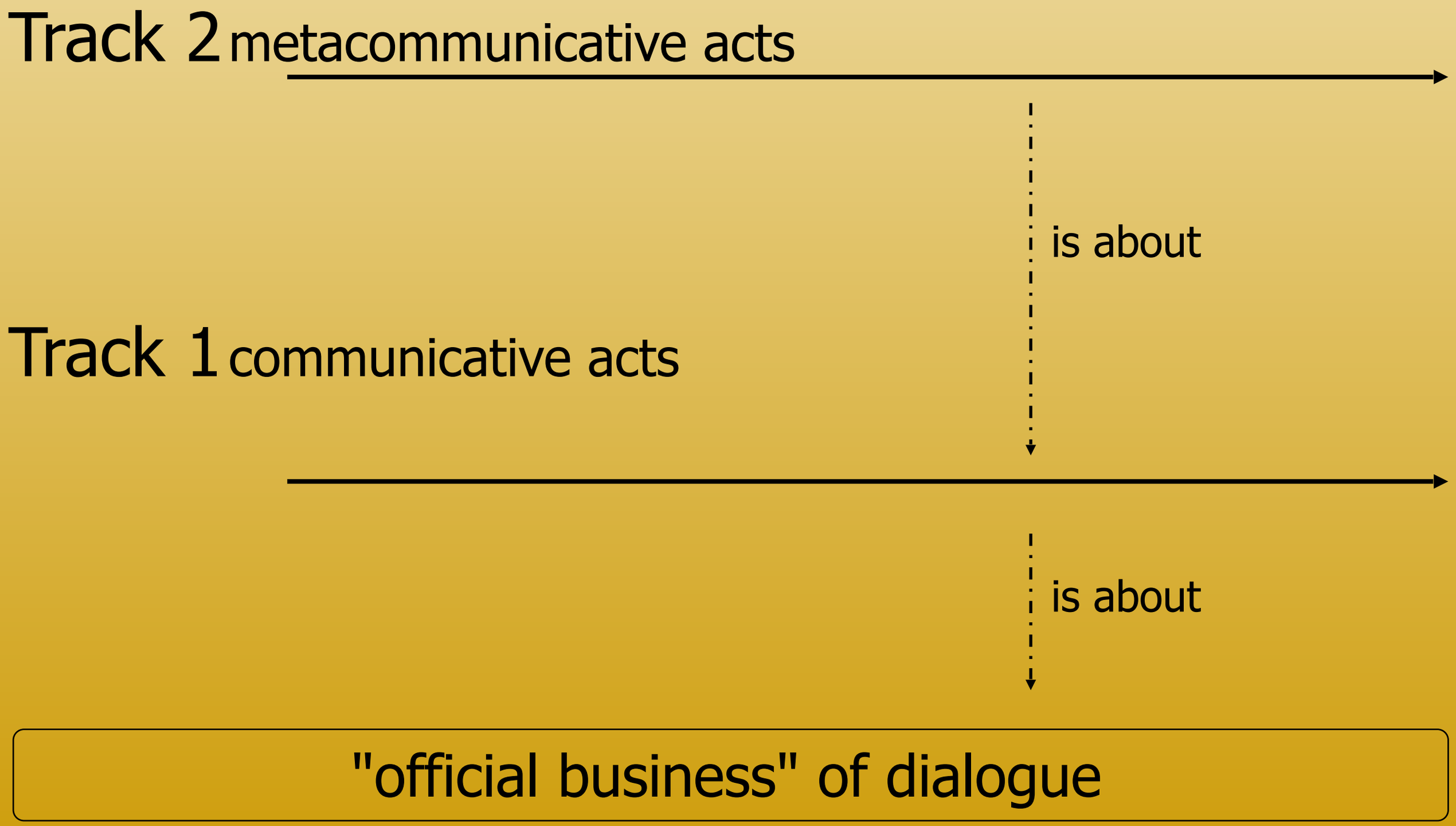
Grounding

- Clark's (1996) 4-level model (cf. also (Allwood 1995))

Level	*Speaker -- Hearer*
4	proposal & consideration
3	meaning & understanding
2	presentation & identification
1	execution & attention

- give evidence for understanding on all levels (with downwards entailment)
- types of evidence: continued attention, relevant next contribution, acknowledgement, demonstration, display

Conversational tracks



Grounding

Track 2 *Do you understand this?*

Track 1 "Who came to the party?"

"official business" of dialogue

Grounding

Track 2 *Do you understand this?* --- *Yes*

Track 1 "Who came to the party?" ---- "Peter."

"official business" of dialogue

evidence of success

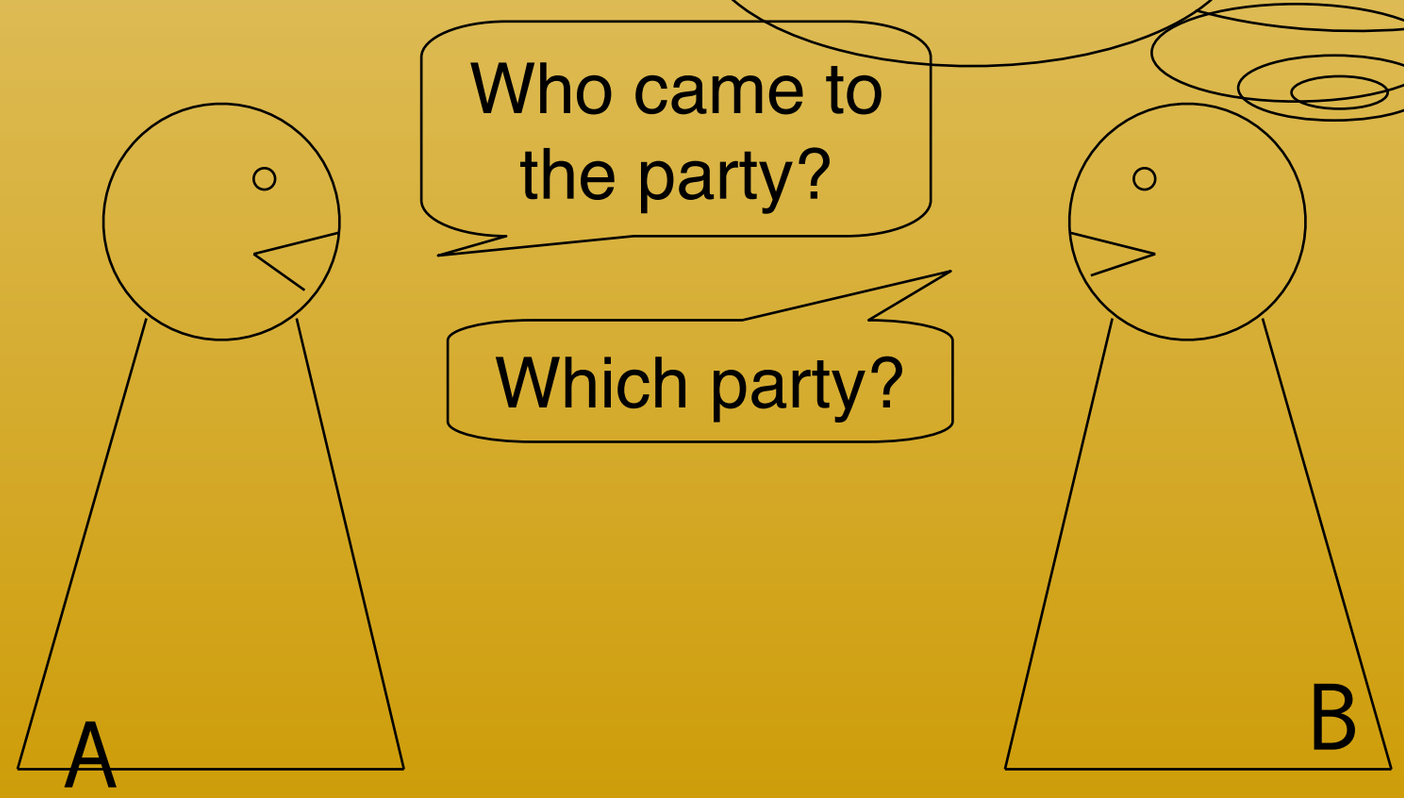
- A: I saw a tiger.
  B: Ok [, you saw a tiger.]
  A: Ok [, you understood that I saw a tiger.]
  B: Ok [, you understood that I understood that you saw a tiger.]
  A: Ok [, you understood that I understood that you understood that [illegible]
  B: Ok [, you understood that I [illegible] understood that I understood that you saw a tiger.]
  A: Ok [, you understood that I understood that you understood that I understood [illegible] my god is this every going to stop I am trapped in a recursion someone send help]
  B: Ok [, lorem ipsum solor sit amet or something like this I'm just typing words now]

Well enough for current purposes!

Grounding - Clarification Requests

- … or signal non-understanding, and request repair:

Level	*Speaker -- Hearer*	
4	proposal & consideration	X
3	meaning & understanding	X
2	presentation & identification	√
1	execution & attention	√



Grounding - Clarification Requests

- frequent: around 5% of utterances in task-oriented dialogues

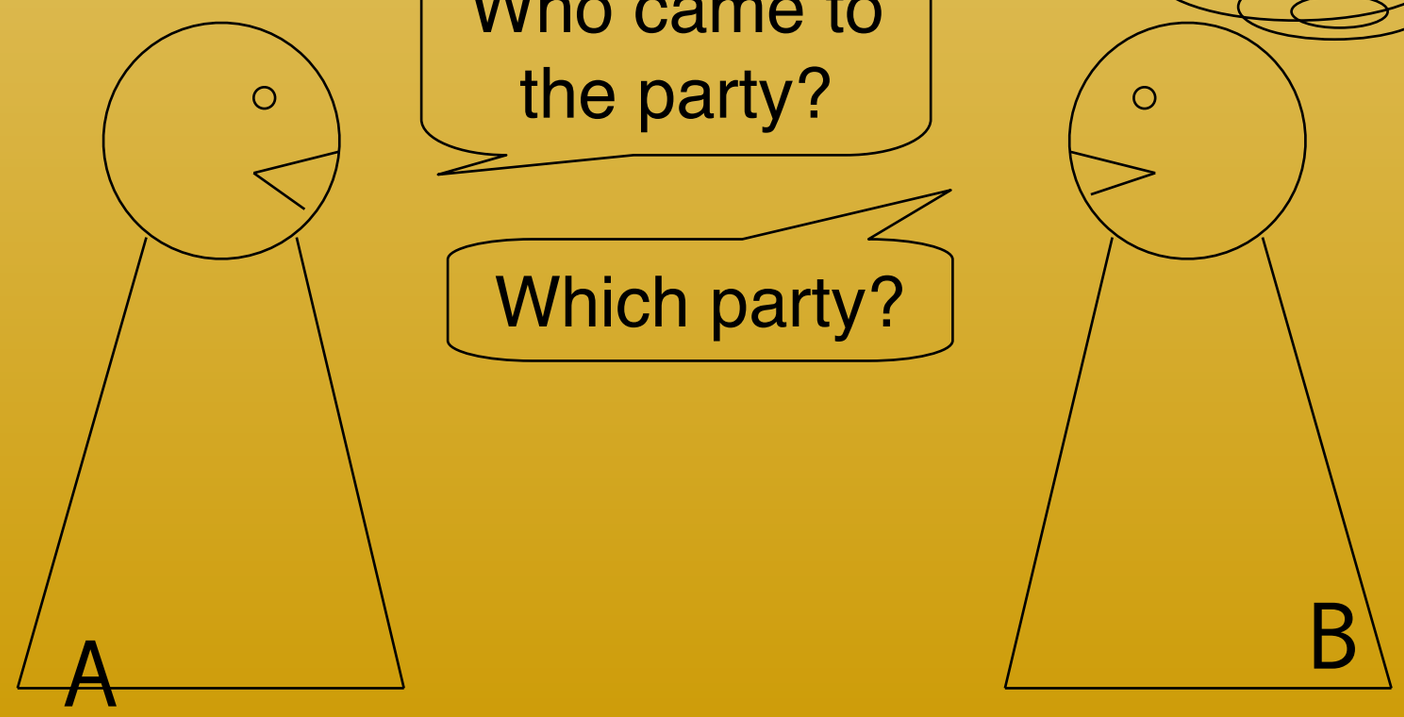
  (Purver et al. 2001, Rodríguez & Schlangen 2004)

- multi-dimensional classification in (Schlangen 2004):
  - Level of problem
  - Extent
  - Severity

Clarification Requests

Dimension 1: Level of problem

Level	*Speaker -- Hearer*	
4	proposal & consideration	X
3	meaning & understanding	X
2	presentation & identification	√
1	execution & attention	√



H. Clark's Grounding Model

"Principle of joint closure: The participants in a joint action try to establish the mutual belief that they have succeeded well enough for current purposes."

Principle of opportunistic closure: Agents consider an action complete <u>just as soon</u> as they have evidence sufficient for current purposes that it is complete.

Principle of repair: When agents detect a problem serious enough to warrant a repair, they try to initiate and repair the problem <u>at the first opportunity</u> after detecting it.

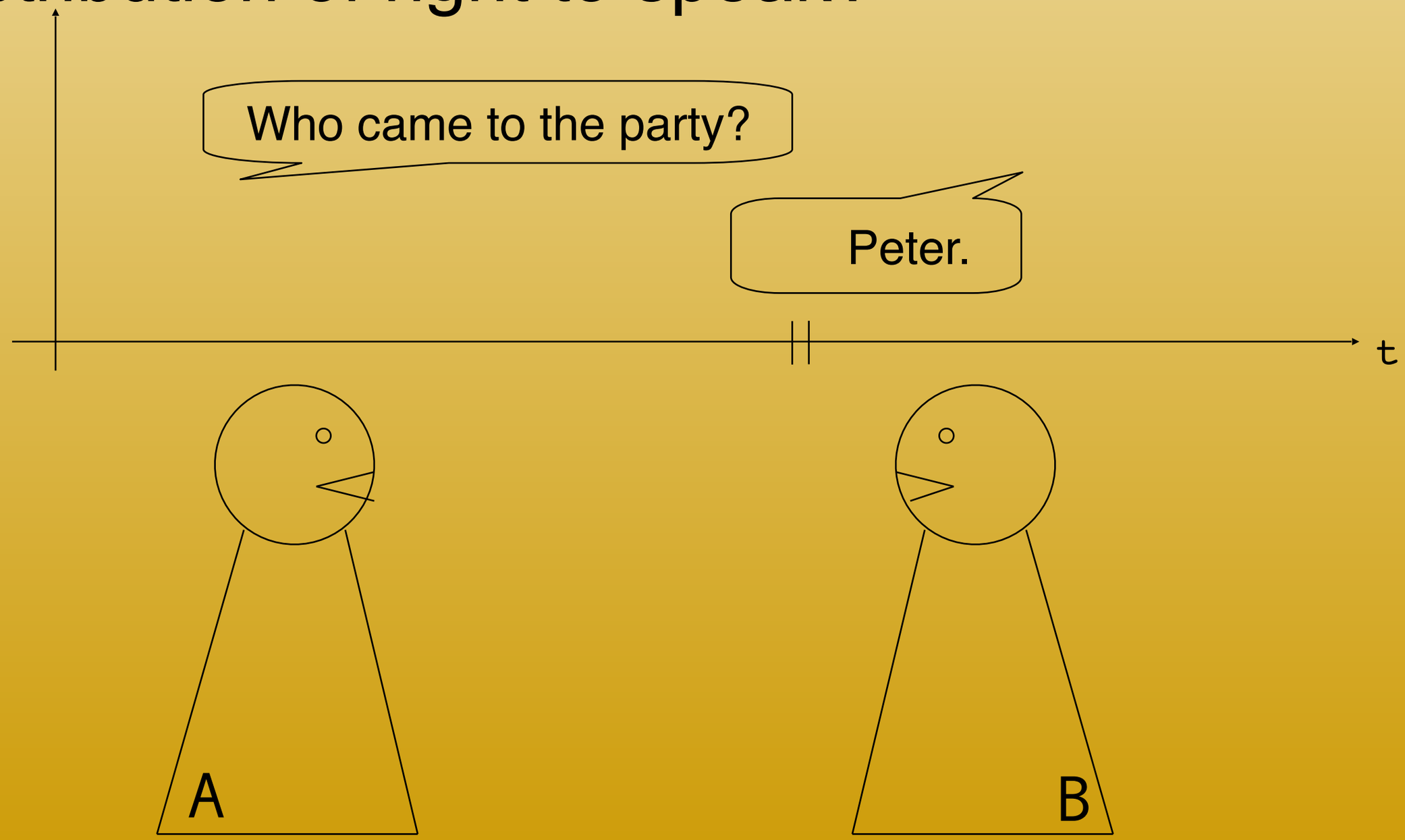
Principle of repair: When agents detect a problem serious enough to warrant a repair, they try to initiate and repair the problem <u>at the first opportunity</u> after detecting it.

```
P   (and then it + the top of the T) fits (into: + next to)
the first piece

P   where the L is   the backwards L
```

Turn-taking

- how do participants in a dialogue organise distribution of right to speak?



Turn-taking

- Observations to account for:
  - overlaps are fairly rare in dialogue (less than 5%)
  - pauses between turns are very short (around 200ms) --- shorter than motor-planning of new utterance!

Turn-taking

- Sacks et al. model (1974):
  - At each transition-relevant-point (TRP) of each turn, the following holds:
    1. If during this turn the current speaker has <u>selected</u> A as the next speaker, then A must speak next.
    2. If the current speaker does not select the next speaker, any other speaker may take the next turn.
    3. If no one else takes the next turn, the current speaker may take the next turn.

Turn-taking

- Selection, how?
  - By asking a question, making a suggestion, etc...
    --> adjacency pairs

    A: Who came to the party?
    B: <silence>
    A: What's up? Did I say something wrong?

Turn-taking

- Model
  - is projective, i.e. utterance itself indicates whether TRP is coming up, and whether other speaker is selected, not "signal-reaction" model
  - can explain "significant silence"
- Although turn-taking works exactly the same way in non-visual modalities (on phone), if visual info is there, then gaze etc. give additional indications.

Turn-taking

- holds only for "track-1" contributions: backchannels systematically overlap!

- rules can be broken: competition for getting floor, upgrading, shouting matches...

H. Clark's Grounding Model & turn taking

speaker	hearer
propose j project	consider proposal
signal *p*	recognize *p*
present signal	identify signal
execute behaviour	attend to behaviour

Principle of opportunistic closure: Agents consider an action complete <u>just as soon</u> as they have evidence sufficient for current purposes that it is complete.

Principle of repair: When agents detect a problem serious enough to warrant a repair, they try to initiate and repair the problem at the first opportunity after detecting it.

* Only one primary presentation at a time
* If it's your turn, start ASAP.

Our takeaways

- Dialogue participants
  - try to reach mutual understanding; need evidence that they have
  - continuously monitor whether they have reached it
  - and, if necessary, repair ASAP;
  - so if you don’t react, you risk repair.

Our takeaways

- Why ASAP?
  - Life’s too short!
- Responsiveness is built into the fabric of dialogue.
- Reducing it makes dialogue *harder*. (Cf. eg. (Brannigon *et al.* 2011)

Responsive Agents

- working definition:
  are responsive to the needs of the dialogue partner(s), *at all times*
  minimize <u>time</u> between *event* and *response*
  respond to many more types of events than “end of turn”
  because they optimize mutual understanding

- Qs:
  - why?
  - how?
  - what type of events?
  - which types of responses?
  - who / what creates these events?
  - does an event have to have occurred to respond to it?
  - what are the optimization criteria?

- presentation events
- understanding events
- feedback responses
- repair responses

sociology philosophy CL / AI

anthrop. psychology linguistics speech eng.

J.L. Austin, 1955: *How to do things with words*

T. Schelling, 1960
The Strategy of Conflict

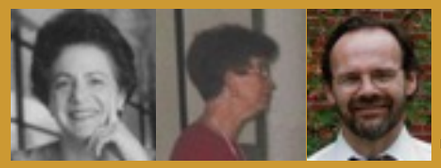
J. Searle, 1969: *Speech Acts*

P. Grice, 1957, '69, '75
Logic and Conversation

D. Lewis, 1969
Convention

H. Sacks,
E. Schegloff,
G. Jefferson
1960ff.
Conversation Analysis

H. Clark, 1978ff.
Joint Action Theory



...

B. Grosz, C. Sidner, J. Allen,
et al. *Communication & Planning*

gestures,
cultural
(in)variants

eye tracking,
visual world paradigm;
mechanistic theories of d.

mid '80s: Discourse Structure
DRT, RST, SDRT, D-TAG, ...

mid '90s: Formal Semantics /
Pragmx of Dial.: SDRT, KOS, ...

< 1960

1960s

1970s

1980s

1990s

2000s

Overview of Day 1

- What does responsiveness mean here?
- What do people do in dialogues?
- Dialogue as coordinated, joint action / as process.
  - Grounding, Turn-Taking, etc.
- State of the art in responsive conversational agents

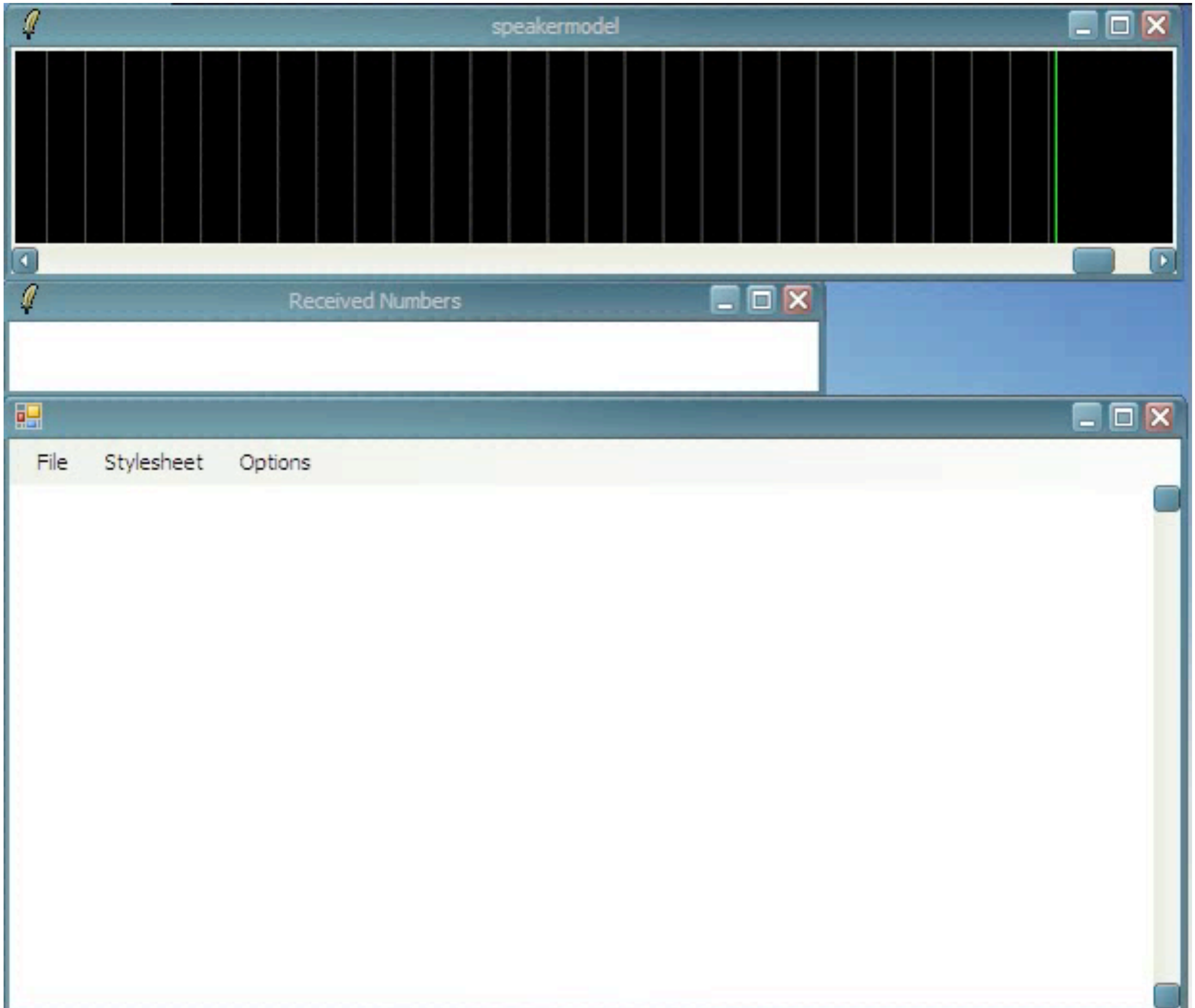
The NUMBERS systems
fast turn-taking

joint work with Gabriel Skantze
(Skantze & Schlangen, EACL 2009)

The NUMBERS systems
fast turn-taking

- user dictates a string of digits to system
- system tries to ground its understanding, as quickly as possible
- processing based on IU-model:
  - minimal units trigger updates
  - processors implement update functions

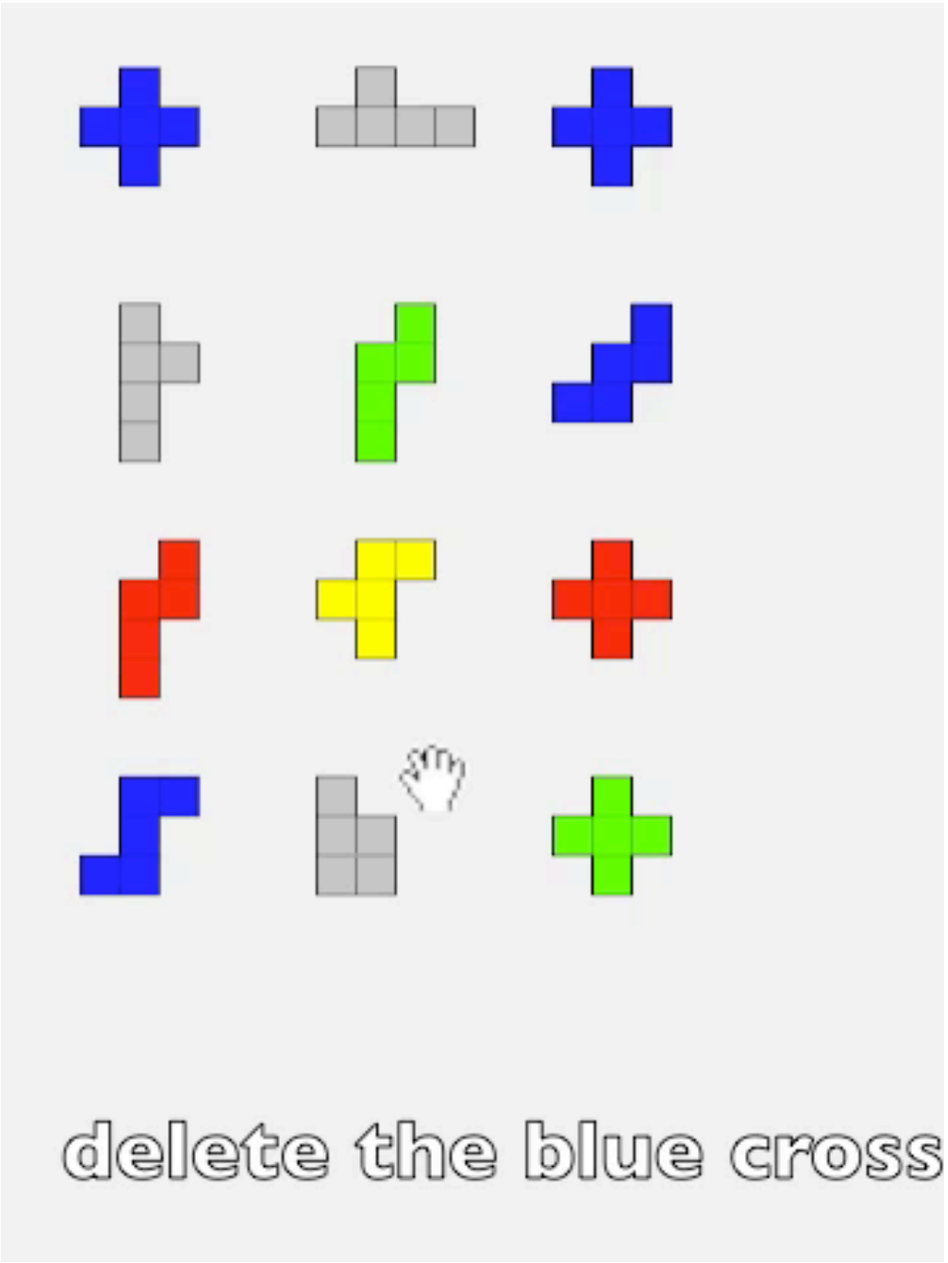
the *numbers* system



The PENTO-10 system
fast turn-taking, immediate exec

joint work with Okko Buß
(Buß *et al.*, SIGdial 2010, semdial 2010, 2011)

Pentomino System



U: *delete the blue cross*
S: which piece?
U: *top right.*
S: ok?
U: *right, now take the yellow [one]...*
S: yes?
U: *... and turn it...*
S: yes?
U: *... to the left*
S: ok.
U: *now flip the stairs...*
S: ok
U: *horizontally*
U: *that's right*
U: *erm now delete the red [one]*
S: *wh-*
U: *bottom right*
U: *correct.*

Evaluation

- Faster task completion compared to non-incremental versions of the systems
- Higher subjective ratings („would use again“, „behaves as expected“, „natural”)
- *Not* higher task success rate
- (Skantze & Schlangen 2009; Buß et al. 2011)

Embodied Conversational Agents

„*Computer interfaces that hold up their end of conversational, have bodies and know how to use it for conversational behaviors as a function of the demands of dialogue and of emotion, personality, and social convention*“ (Cassell 2000)

Required features:

- Recognize and interpret verbal and nonverbal input behavior
- Generate verbal and nonverbal output behavior
- Process multiple functions of conversational behavior
- Take an active role in dialogue (mixed-initiative)

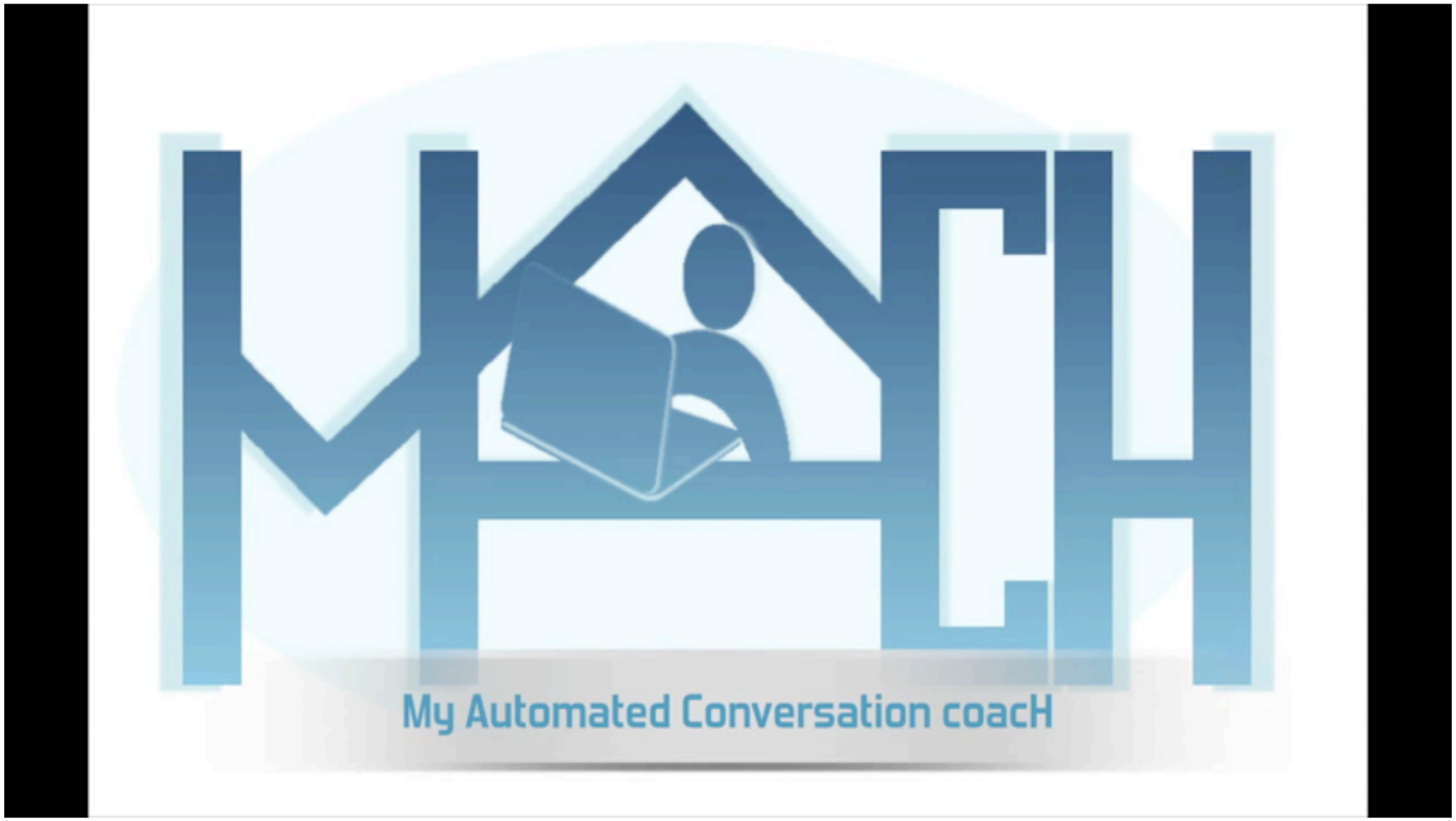
Embodied Conversational Agents

1994 1997 1999 2002 2005 2008

Virtual Real Estate Agent (Rea)

MIT Media Lab (J. Cassell et al.)

Tutoring: Communication training



Conversation Coach by MIT (R. Picard et al.)

Information kiosk

Ada & Gras @ Boston Science Museum (ICT)

Personal assistant

Elder Companion „Billie“ (CITEC, U. Bielefeld)

Overview of Day 1

- What does responsiveness mean here?
- What do people do in dialogues?
- Dialogue as coordinated, joint action / as process.
  - Grounding, Turn-Taking, etc.
- State of the art in responsive conversational agents

Questions?

End of Day 1

Tomorrow: Technical Challenges, Background