

# From Language Processing to Language Use

David Schlangen

Grundlagen der Computerlinguistik // Department Linguistik // Universität Potsdam

<http://clp.ling.uni-potsdam.de>

[david.schlangen@uni-potsdam.de](mailto:david.schlangen@uni-potsdam.de)

@ Pitt NLP Seminar

2021-12-10

These slides:

<https://clp.ling.uni-potsdam.de/talks>

# Motivation

- Research interest: language use.  
The use of language (primarily) to achieve goals, including coordinations goals.
- NLP is moving at breakneck speed, but is it moving towards better understanding of language use? Where is it actually moving?
- Principled research programme for ... maybe not NLP, not CL, but something like “linguistic AI”.

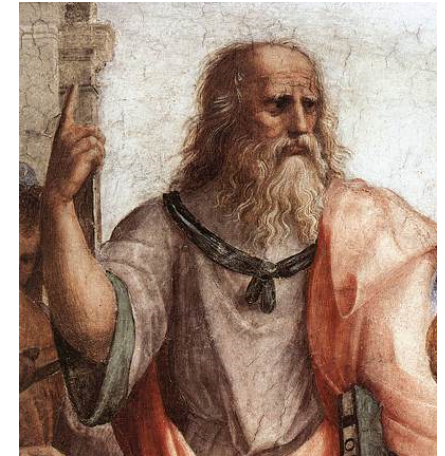
# NLP (in its current form)

- Induce models from datasets of observed language use:
  - in general text
  - in language tasks (specific mappings of input to desired output)
- Models get increasingly better at:
  - Classification
  - Transduction ( “content-preserving” reformulation of material )
  - Compression ( “content-selecting” reformulation )
- Is that all that is required for language *use*?

# Conclusions

- If you want to build interactive language systems that are useful (& safe, ethically defensible) now, you have to *design* them. Current end-to-end techniques won't give you what you need.
- If you want to understand and learn from how humans use and learn to use language “end to end”, you need to start with simple / hard settings, and focus on the interaction *process*.

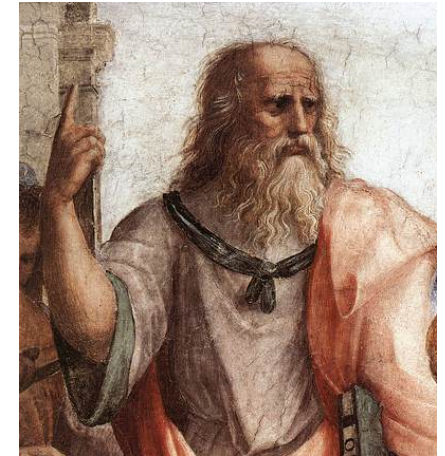
# Overview



- Introduction: Language & Norms
- Natural Language Processing and Natural Language Use
- The Interaction Process and the Interacting Agent
- A Research Programme, & Some Examples



# Overview



- Introduction: Language & Norms
- Natural Language Processing and Natural Language Use
- The Interaction Process and the Interacting Agent
- A Research Programme, & Some Examples



# The Underlying Picture

- Meaningful language use is governed, enabled, & constituted by a complex network of *social norms*
    - norms: public standards for what's good; *if in C, you <n> do X*
  - Varieties of norm following:
    - norm conformance
    - norm exploitation
    - norm participation
- Claim: This kind of analysis is useful for setting the agenda for work on “linguistic AI”; thinking about areas where improvement is needed, and possible limitations.

Wittgenstein (1953), Sellars (1954), Brandom (1998), Hegel (1807); Grice (1957), Lewis (1979), Stalnaker (1973)  
Asher & Lascarides (2003), Ginzburg (2012), Stone & Lepore (2014); Hanks (2013)

Schlangen (forthcoming)

# The Underlying Picture

- Meaningful language use is governed, enabled, & constituted by a complex network of *social norms*

- norms: public standards for what's good; *if in C, you <n> do X*

- Varieties of norm following:

- norm conformance
- norm exploitation
- norm participation

Current NLP techniques target norm conformance in language use, and even that only incompletely.

Problem: E2E on task (i/o mapping) misses general language capabilities which are abstracted away in datasets.

( Uses of NLP systems normatively bind the provider of the system. )

We have very little grasp on how norm participation could be modelled.

We need to start from the very beginning.



# The Underlying Picture

- Meaningful language use is governed, enabled, & constituted by a complex network of *social norms*
  - norms: public standards for what's good; *if in C, you <n> do X*
- Varieties of norm following:
  - norm conformance
  - norm exploitation
  - norm participation

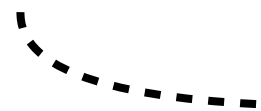
As a side effect, this results in finer-grained view of what “grounded language” is:

  - “language entry moves”, perception to concepts
  - “language to language moves”, inferential relations between commitments, beliefs
  - “language exit moves”, public commitments to act

(Sellars 1954)

# Making Meaning

*state of affairs*



...

*event type*

— *event* —

interpretation



(your assumptions about)  
**event producer**

**you**

# Making Meaning

*something is burning*

cause

“smoke means fire”



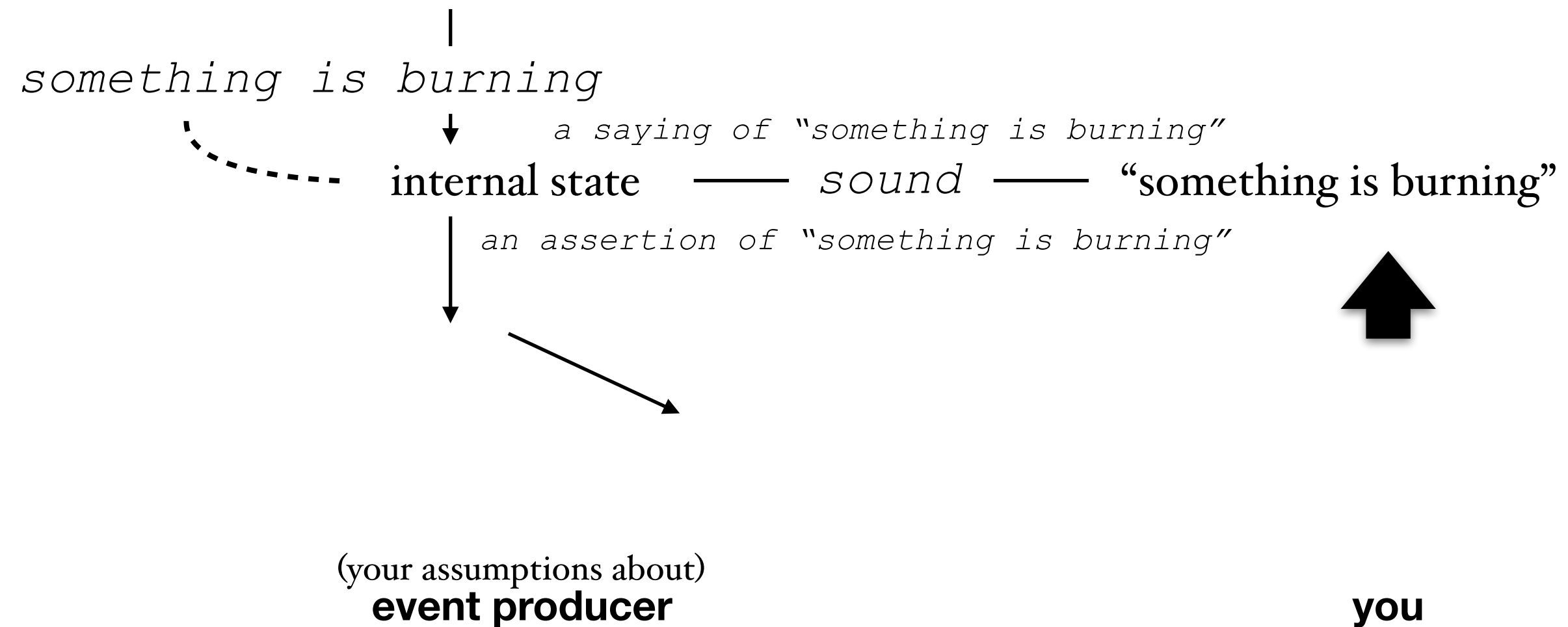
*smoke*  
*this*

— “something is burning”

(your assumptions about)  
**event producer**

**you**

# Making Meaning



# Making Meaning





# Making Meaning



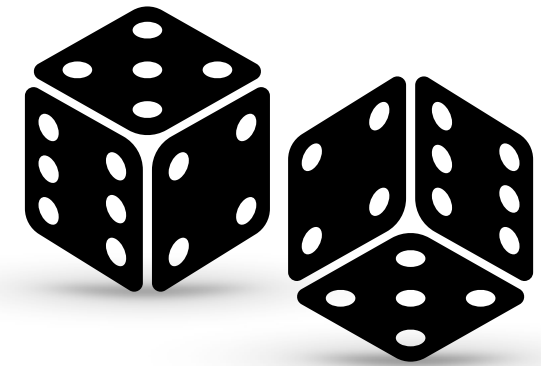
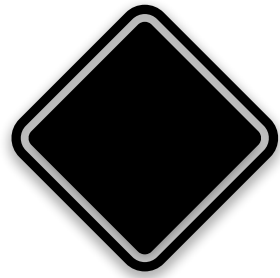
**AN ERROR HAS OCCURRED.  
THIS PRESENTATION WILL NOW BE  
DELETED.**





Andrew Malone, CC BY 2.0 <<https://creativecommons.org/licenses/by/2.0/>>, via Wikimedia Commons





# Part I: Language Processors as Language Users

Part I.A: The Camera that Tells it Like it Sees it

Part I.B: The Search Engine that Just Knows

Part I.C: The Bot that Just Wants to Chat

Notes

# Pa Language Language

Part I.A: The Camera

Part I.B: The Search E

Part I.C: The Bot that

**Part I.A: The Camera that Tells it Like it Sees it**

Part I.B: The Search Engine that Just Knows

Part I.C: The Bot that Just Wants to Chat

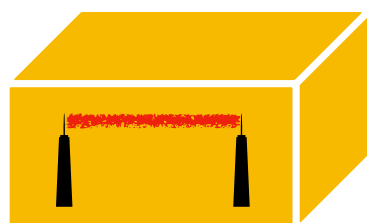
# The Backstory

Finally! You've moved to the suburbs, into your own little house, with your own little garden, with your own little patio. BUT! Who is messing with your outdoor furniture when you are not looking? You want to know!

Strategy: We'll analyse the use of various (imagined) devices that provide information about a state of affairs.

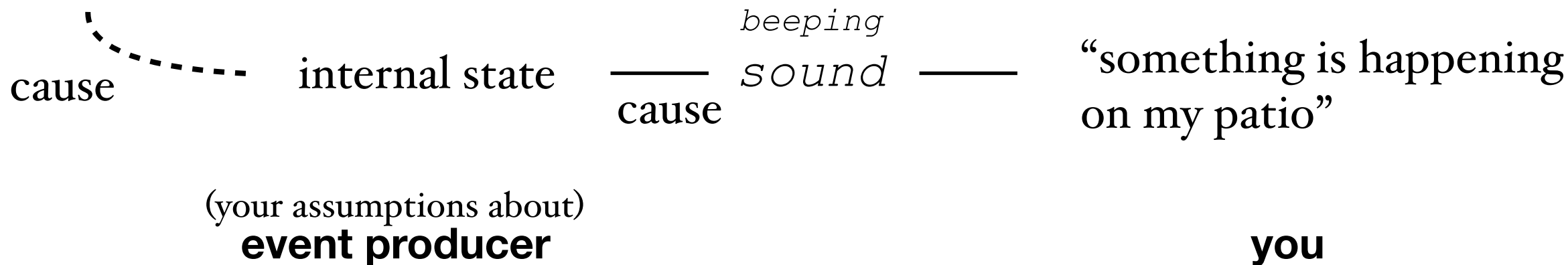
What happens as these become increasingly complex / symbolic? Where is the meaning that they make coming from?

# First Attempt: The Alarm



“Something or someone is entering your property. What is going on? Do you want to know??? Our LaserBeamTrapXXXTRM® will tell you!”

*something on patio*



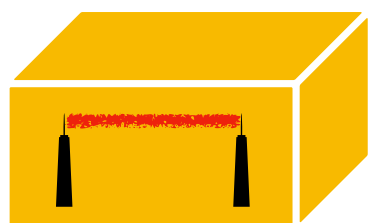
“something interrupted the beam”

“that pesky neighbour is on our patio!”



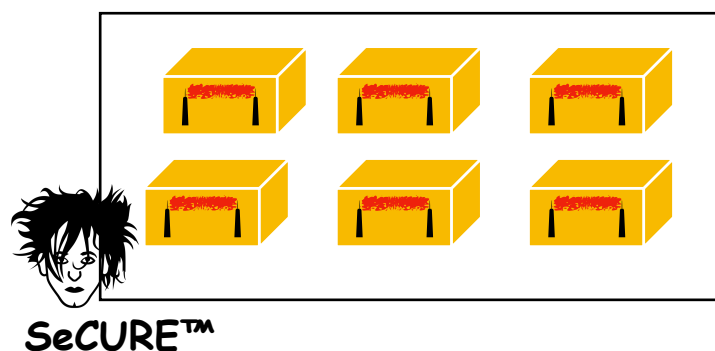
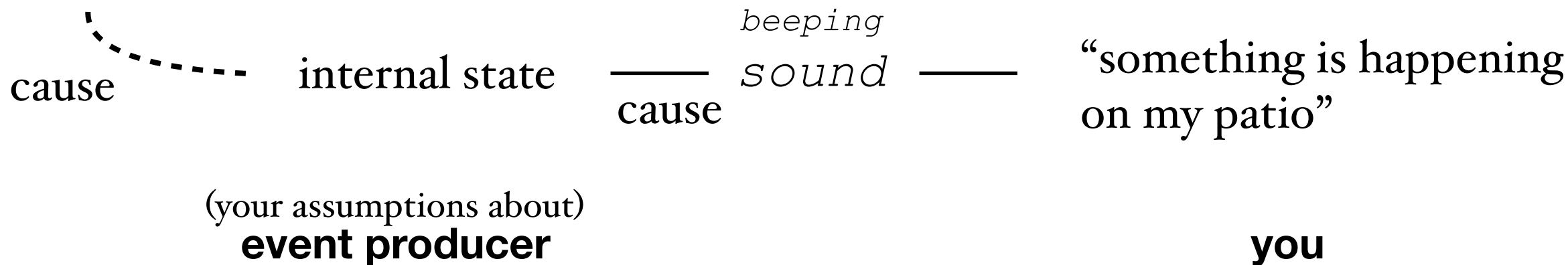
Theaetetus: “*Knowledge is justified true belief.*”

# First Attempt: The Alarm



“Something or someone is entering your property. What is going on? Do you want to know??? Our LaserBeamTrapXXXTRM® will tell you!”

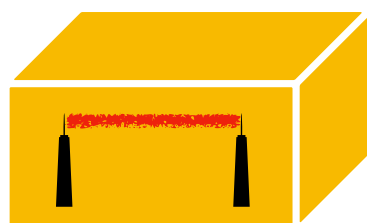
*something on patio*



“repair this thing or give me a refund!!!”

**you**

# First Attempt: The Alarm

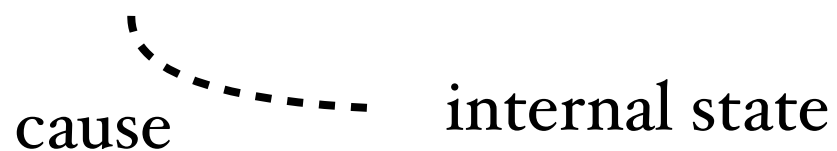


“Som  
Wha  
Laser

- *knowledge & justification*

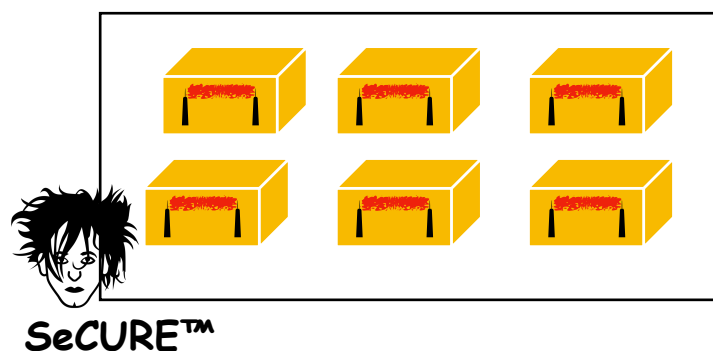
Notes

*something on patio*



(your assumptions about)  
**event producer**

caus

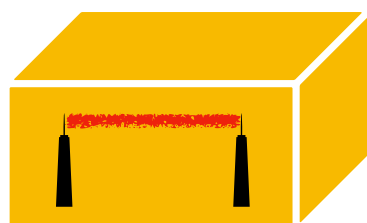


“rep  
me a refund!!!”

you

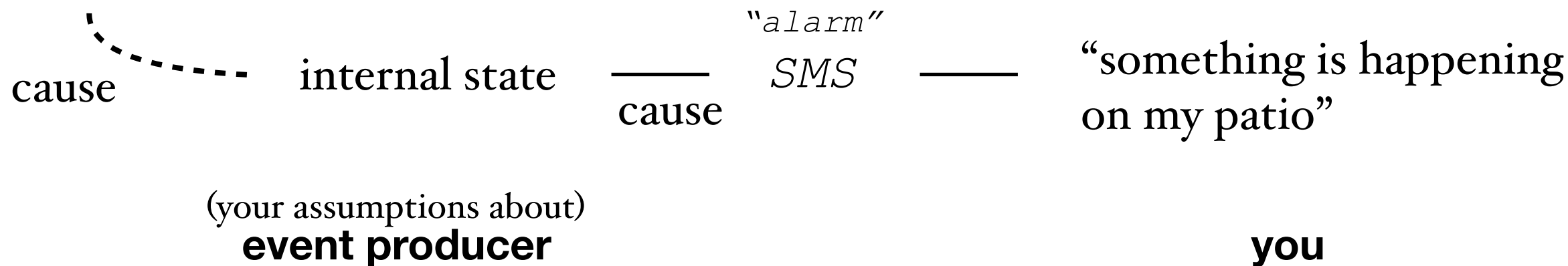


# The Upgrade



“Something or someone is entering your property. What is going on? Do you want to know??? Our LaserBeamTrapXXXTRMvII® will tell you!”

*something on patio*



# The Future is here!

## (2016 edition)



“Something or someone is entering your property. What is going on? Do you want to know??? Let SENTINEL-XXTRM-w/AI® tell you!”

*something on patio*

cause? ——— internal state  
judgement?  
(your assumptions about)  
**event producer**

———— SMS ———  
cause  
“*large animal*”  
“*human*”

“there is (something that one can call) an X on my patio”

“Urteilstkraft überhaupt ist das Vermögen, das Besondere als enthalten unter dem Allgemeinen zu denken” (Kant, KdU, Einl. IV)

“The faculty of judgement is the capacity to think the particular as contained under the universal.”

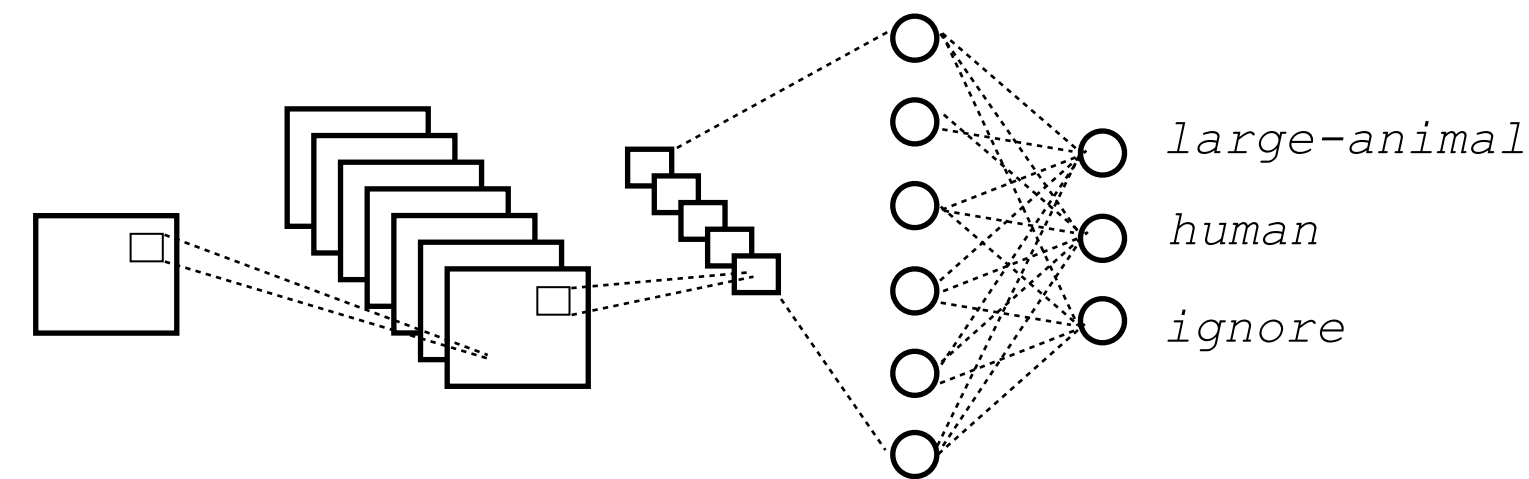
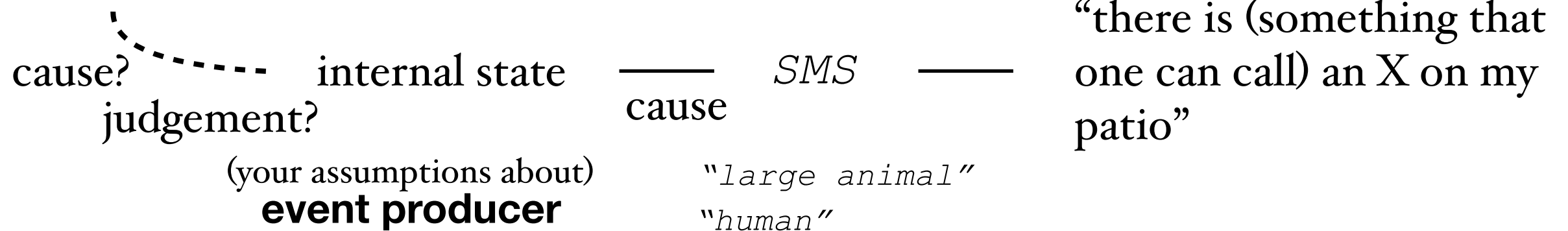
# The Future is here!

## (2016 edition)



“Something or someone is entering your property. What is going on? Do you want to know??? Let SENTINEL-XXTRM-w/AI® tell you!”

*something on patio*



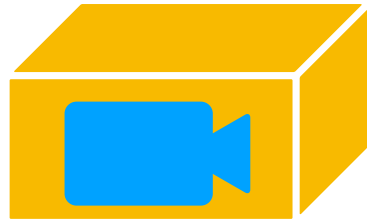
“The faculty of

The model is *reproducing* the type judgements in the training data. Is that *making* judgements?

CONTAINED UNDER THE  
universal.”

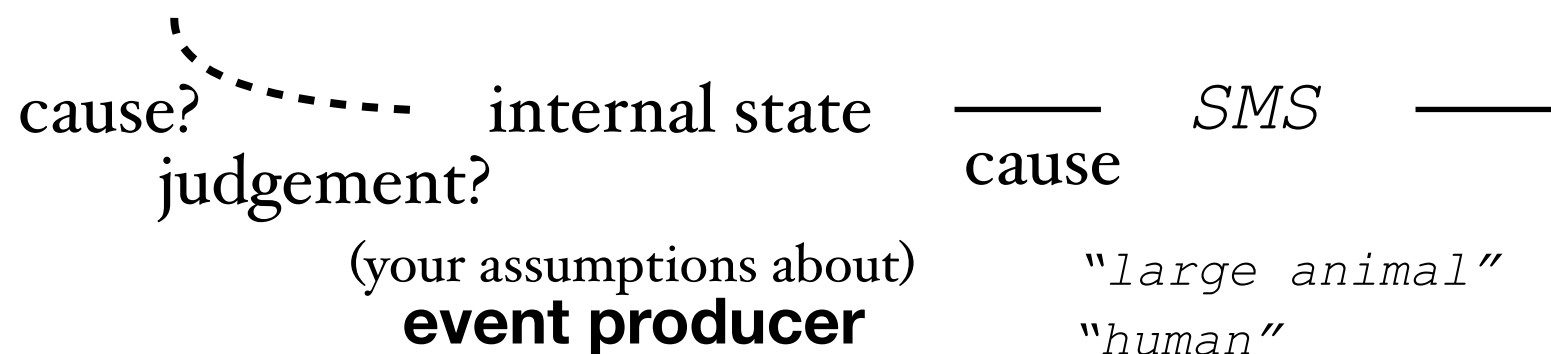
# The Future is here!

## (2016 edition)

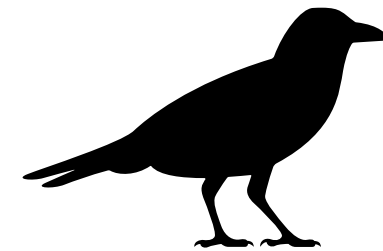
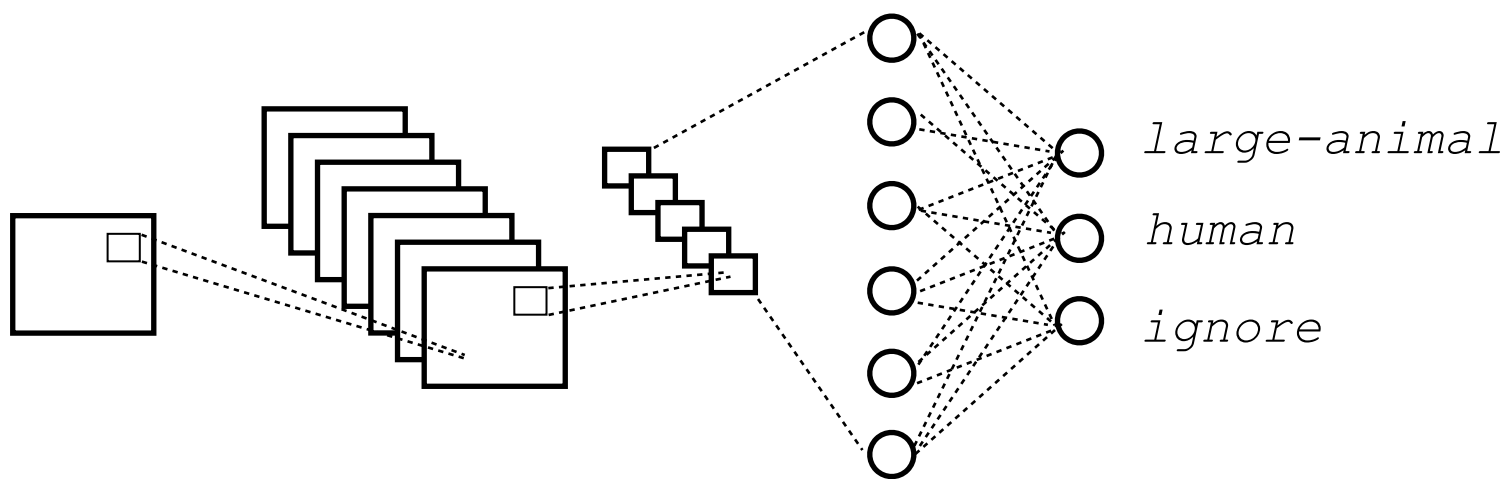


“Something or someone is entering your property. What is going on? Do you want to know??? Let SENTINEL-XXTRM-w/AI® tell you!”

*something on patio*



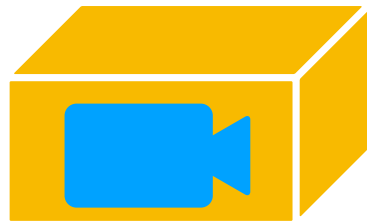
“there is (something that one can call) an X on my patio”



“well, this you can’t call a ‘large animal’. Most of our annotators seem to agree!”

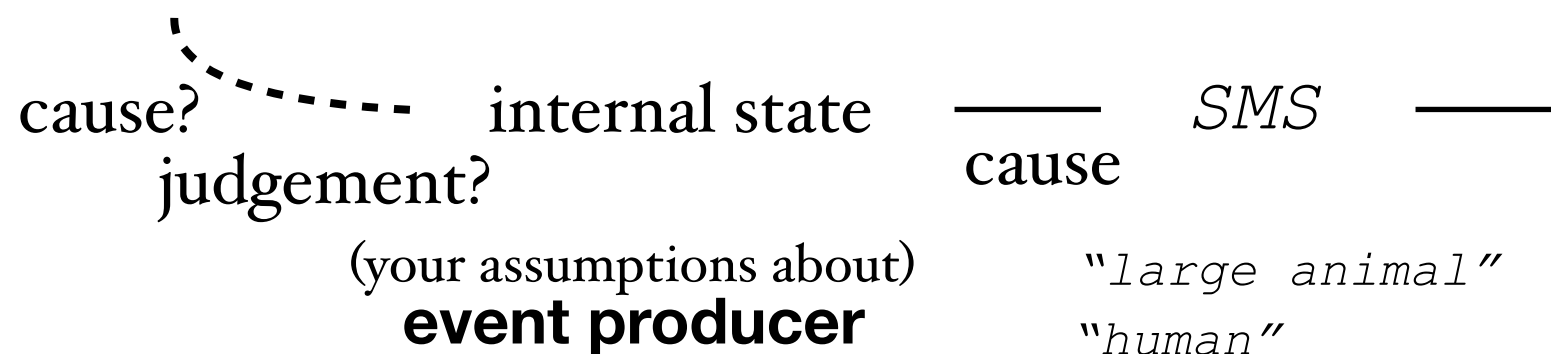
# The Future is here!

## (2016 edition)



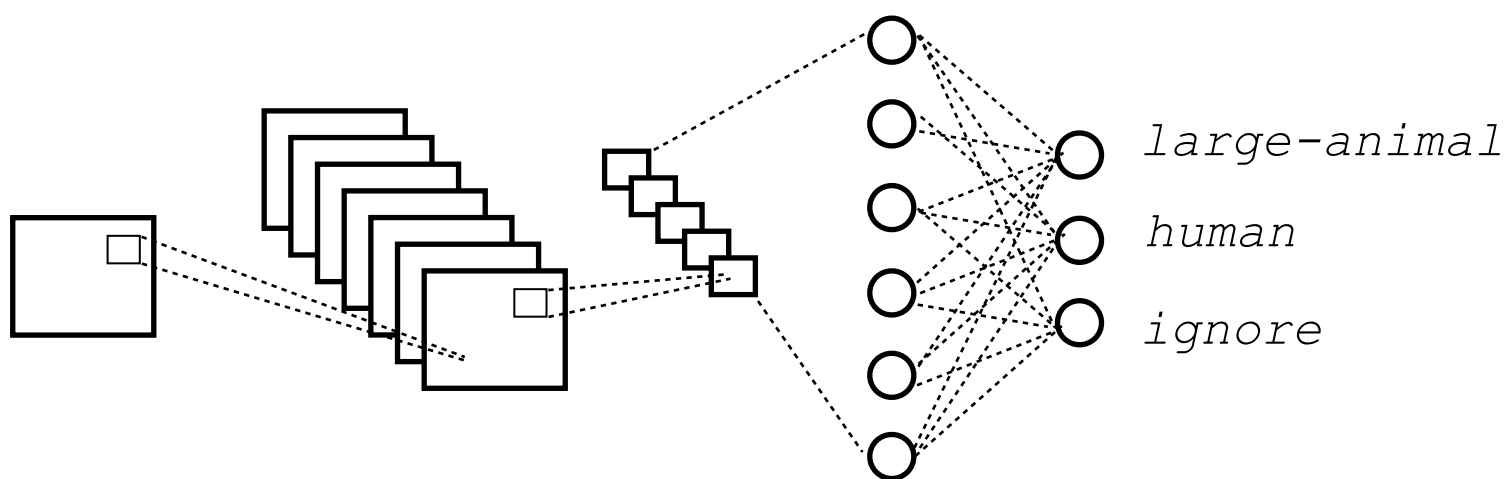
“Something or someone is entering your property. What is going on? Do you want to know??? Let SENTINEL-XXTRM-w/AI® tell you!”

*something on patio*



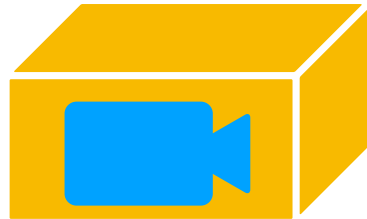
“there is (something that one can call) an X on my patio”

“there is something that **this annoying machine calls** an X on my patio”



# The Future is here!

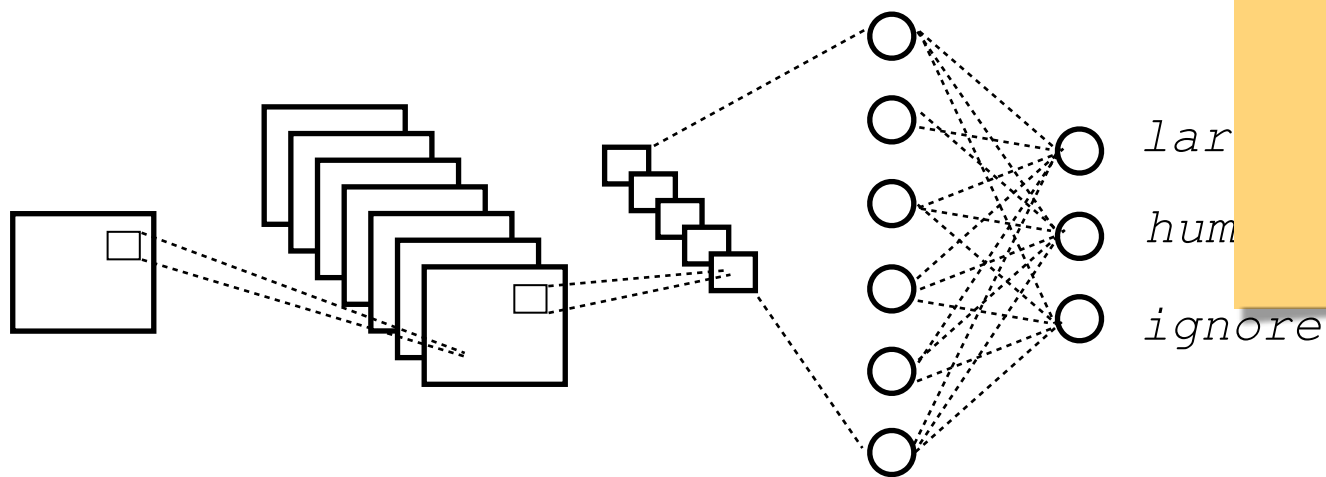
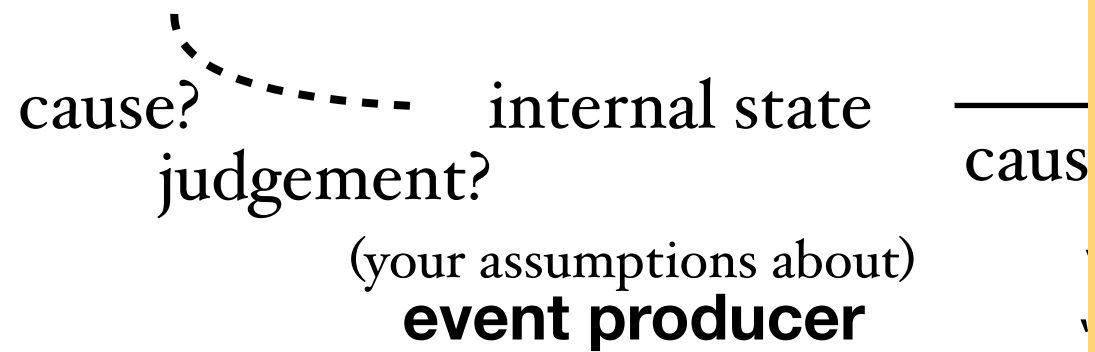
## (2016 edition)



“Som  
Wha  
SEN”

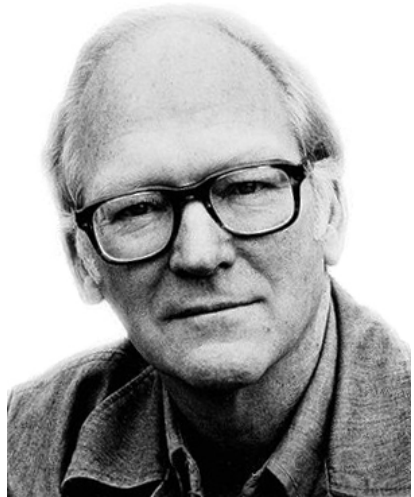
- Notes
- *knowledge & justification*
  - *judgement & the conceptual*

*something on patio*



at  
ine  
”

# Having Concepts

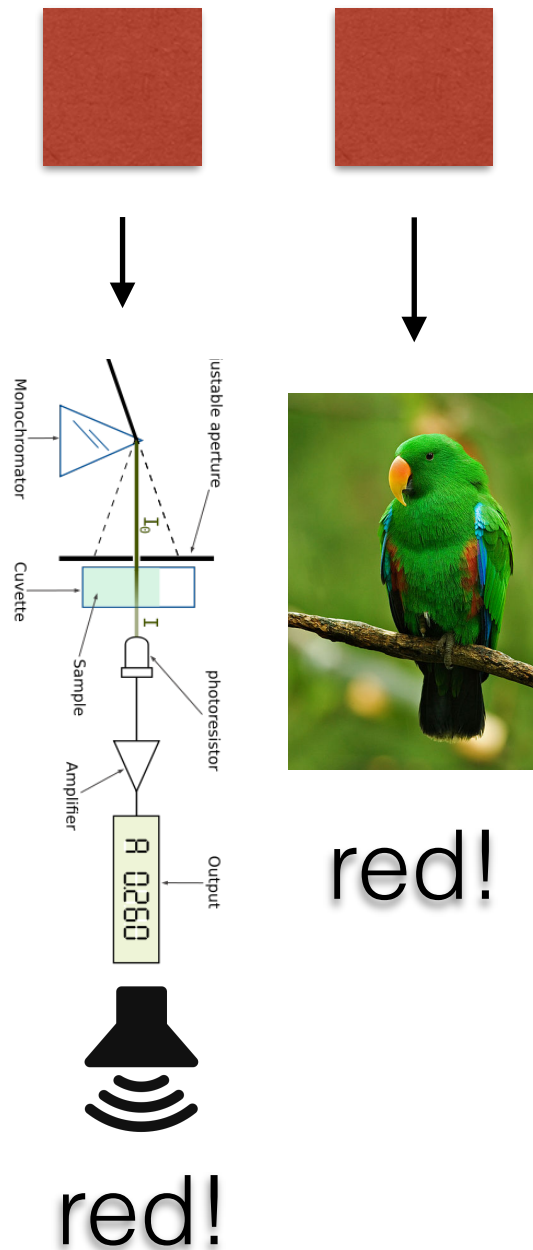


[Ascribing possession of the concept “tree” to a dog] “would seem impossible unless we suppose the dog has many general beliefs about trees: that they are growing things, that they need soil and water, that they have leaves or needles, that they burn. There is no fixed list of things someone with the concept of a tree must believe, but without many general beliefs, there would be no reason to identify a belief as a belief about a tree, much less an oak tree.” (Davidson, 1982, p. 98)

“[U]nless we want to attribute concepts to butterflies and olive trees, we should not count mere ability to discriminate between red and green or moist and dry as having a concept, not even if such selective behaviour is learned. To have a concept is to classify objects or properties or events or situations while understanding that what has been classified may not belong to the assigned class.” (Davidson, 1997, p. 139)



# Having Concepts



“[instrument & parrot] do not understand their responses; those responses mean nothing to them, though they can mean something to us” (Brandom 1998)

“The parrot does not treat ‘That’s red’ as incompatible with ‘That’s green’ or as following from ‘That’s scarlet’ and entailing ‘That’s colored’. [...]

It follows immediately from such an inferential demarcation of the conceptual that in order to master any concepts, one must master many concepts.” (Brandom 2000)

“the logical space of reasons, of justifying and being able to justify what one says” (Sellars 1956, §36)

“mastery of the practices of giving and asking for reasons in which their responses can play a role as justifying beliefs and claims” (Brandom 1998, 88)



# Having Concepts

- “in order to master any concepts, one must master many concepts” [HOL]
- “mastery of the practices of giving and asking for reasons in which their responses can play a role as justifying beliefs and claims” “the *logical* space of reasons” [GAR]
- “to have a concept is to classify objects or properties or events or situations while understanding that what has been classified may not belong to the assigned class.” [EXP]

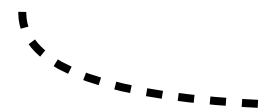
# The Future is Here!

## (2018 edition)



“Something or someone is entering your property. What is going on? Do you want to know??? Let SENTINEL-XXTRM-w/AI++® tell you!”

*state of affairs*



internal state

*“a house with a large window and a wooden bench”*



*SMS*



*“everything’s fine”*

If you see something like this, you should be recognising a house, and a bench, and ...

*“I wouldn’t call that a large window.”*

If you’ve understood a scene like this, and you want to convey what you’ve understood, you should say something like that.

*“a house with a bench?”  
“well, a house with a bench in front of it”*

# The Future is Here!

## (2018 edition)



“Something or someone is entering your property. What is going on? Do you want to know??? Let SENTINEL-XXTRM-w/AI++<sup>®</sup> tell you!”

*state of affairs*



internal state

*“a house with a large window and a wooden bench”*

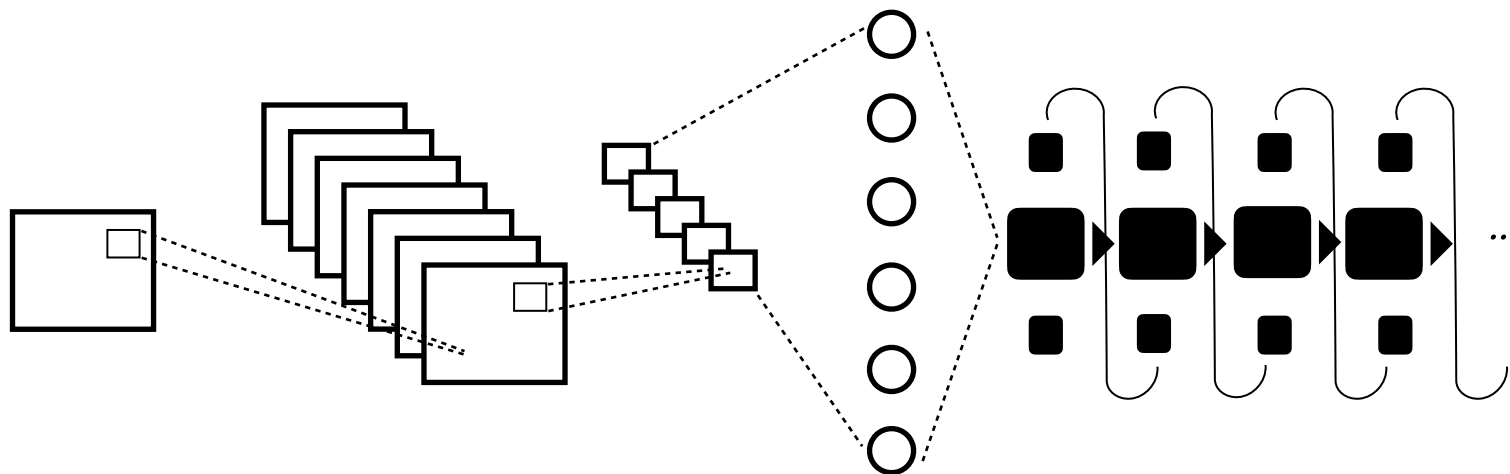


*SMS*



*“everything’s fine”*

*“I wouldn’t call that a large window.”*



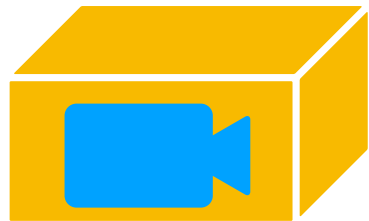
cause

cause

*“a house with a bench?”*  
*“well, a house with a bench in front of it”*

# The Future is Here!

## (2018 edition)



“Som  
Wha  
SEN”

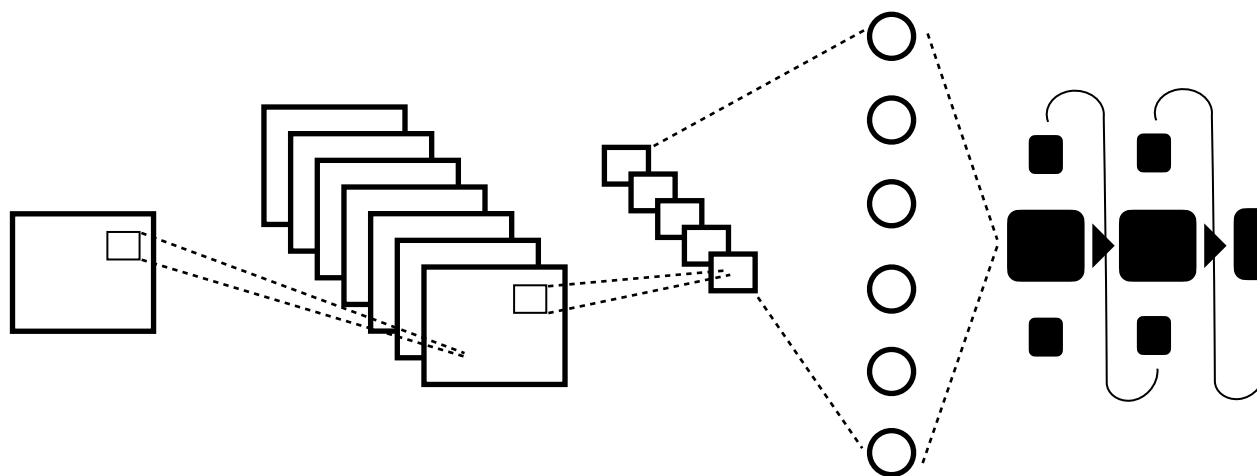
### Notes

- *knowledge & justification*
- *judgement & the conceptual*
- *the conceptual & the linguistic*
- *the linguistic & the infinite*

*state of affairs*

“a h  
win

internal state



cause

cause

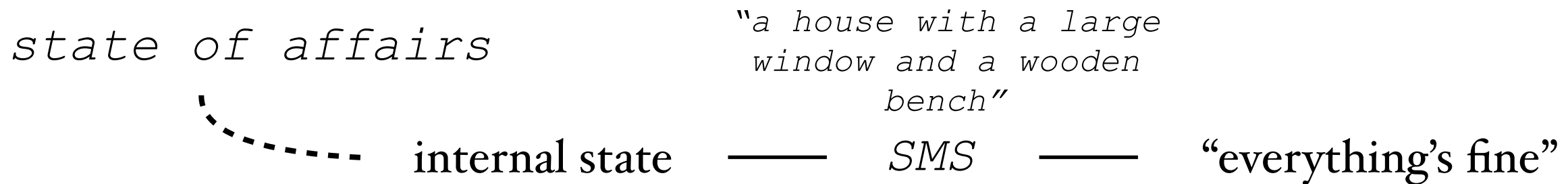
*front of it”*

*ge*

*in*

# The Future is Here!

(2018 edition)



This very much looks like an assertion!

I claim that this is true.

I believe this.

I am justified in believing this.

Beliefs should only change if there is a reason for the change.



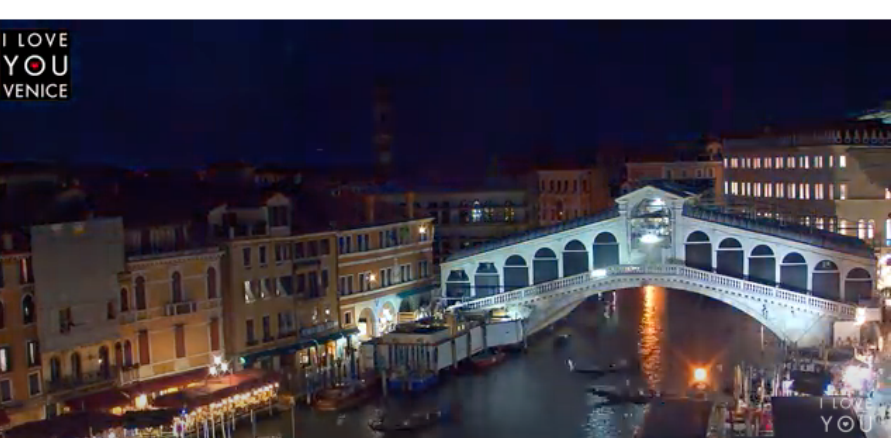
a city street filled with lots of traffic  
a group of people walking down a street



a city street filled with lots of traffic  
a couple of boats that are sitting in  
the water



a city street filled with lots of traffic  
a couple of boats that are sitting in  
the water



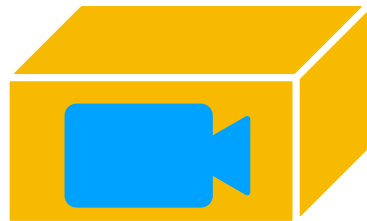
a city street filled with lots of traffic  
a train that is sitting on the tracks

Idea after (Losing, Hammer, Wersing 2016; KNN Classifier with Self Adjusting Memory for Heterogeneous Concept Drift)  
Images from <https://www.youtube.com/watch?v=Lf51Jul-pik> Model: <https://milhidaka.github.io/chainer-image-caption/>



# The Future is Here!

## (2018 edition)



“Something or someone is entering your property. What is going on? Do you want to know??? Let SENTINEL-XXTRM-w/AI++<sup>®</sup> tell you!”

10:00	internal state	——	<i>SMS</i>	<b>This very much looks like an assertion!</b>
				I claim that this is true.
11:00	internal state	——	<i>SMS</i>	I believe this.
				I am justified in believing this.
12:00	internal state	——	<i>SMS</i>	
13:00	internal state	——	<i>SMS</i>	<b>Beliefs should only change if there is a reason for the change.</b>
14:00	internal state	——	<i>SMS</i>	fix: $p(y x,s)$
15:00	internal state	——	<i>SMS</i>	

# The Future is Here!

## (2018 edition)



“Som  
Wha  
SEN”

### Notes

- *knowledge & justification*
- *judgement & the conceptual*
- *the conceptual & the linguistic*
- *the linguistic & the infinite*
- *the expected behaviour of beliefs*

10:00

internal state



11:00

internal state



12:00

internal state



13:00

internal state



14:00

internal state



15:00

internal state

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

*SMS*

fix:  $p(y|x,s)$

*SMS*

ion!

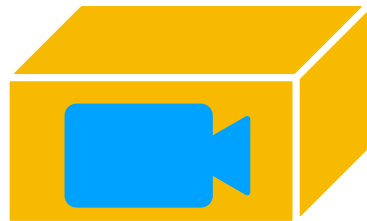
this.

is a



# The Future is Here!

## (2018 edition)



“Something or someone is entering your property. What is going on? Do you want to know??? Let SENTINEL-XXTRM-w/AI++® tell you!”

10:00	internal state	——	<i>SMS</i>	<b>This very much looks like an assertion!</b>
				I claim that this is true.
11:00	internal state	——	<i>SMS</i>	I believe this.
				I am justified in believing this.
12:00	internal state	——	<i>SMS</i>	
				<b>Beliefs should only change if there is a reason for the change.</b>
13:00	internal state	——	<i>SMS</i>	
				<b>Justifications are public objects. If you are justified, you can justify yourself.</b>
14:00	internal state	——	<i>SMS</i>	
15:00	internal state	——	<i>SMS</i>	

# The Future is HERE!

## (2020 edition)



“Something or someone is entering your property. What is going on? Do you want to know??? Just ask our ‘ProctectAIr®’ ”



*a house with a large window and a wooden bench*

*Is the white chair there?*

*Yes*

*Is there anyone out there?*

*No*

*What about the brown chair?*

*Yes*

*What about the black chair?*

*brown*

*Is there a white chair?*

*No*

*The outdoor furniture is untouched, right?*

*Yes*

*Is there outdoor furniture?*

*No*

*Is there a lawn?*

*No*

*Is there grass?*

*Yes*

**What is going on??**  
**Does this thing even understand me?**  
**Does it understand what it says??**

<http://demo.visualdialog.org/hre/>

# The Future is HERE! (2020 edition)

Please wait while a caption is generated for the uploaded image.

**Caption:** A fountain in a park with many pigeons walking on the ground.

are there birds?

i do n't see any

are there pigeons?

yes

but pigeons are birds!

yes

are there birds?

no

Let's hold that thought: It would be nice if we were able to do this (ask for reasons, discuss concepts).

Would likely increase our trust (= our justification, = our ability to acquire knowledge from testimony).

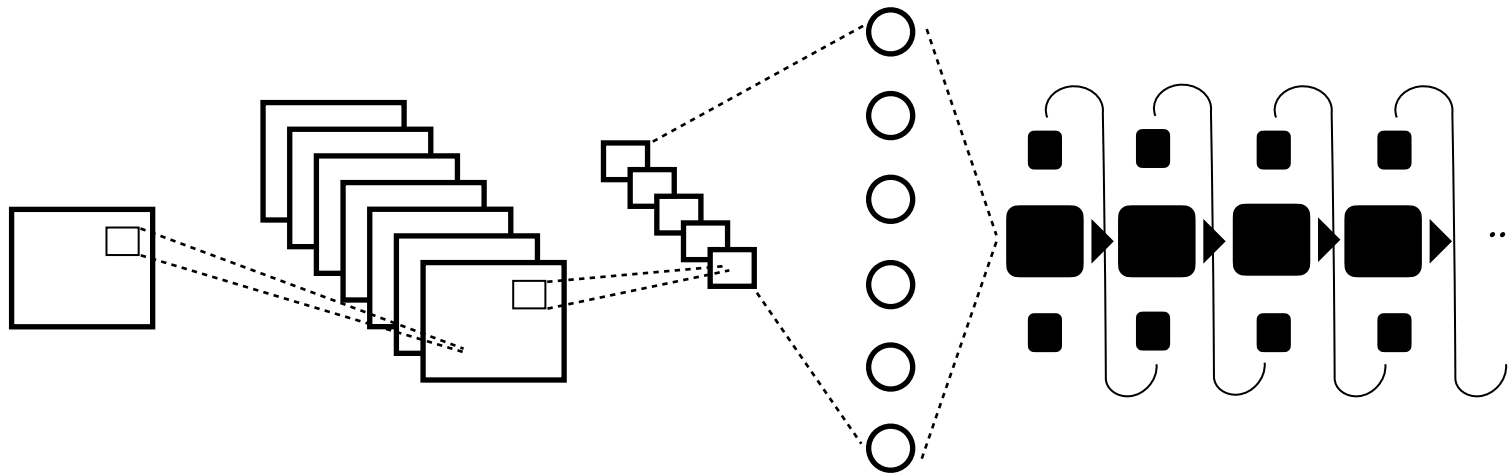
Is this unfair? This is clearly out of domain in terms of speech acts.

The point is that what's in domain is not a viable language game! This domain only exists by your fiat.

<http://demo.visualdialog.org/hre/>

# Recap: The Camera

- “The Camera that Tells is Like it Sees it”:
  - “telling” is linguistic norm following
  - “seeing” is conceptual norm following



- These NLP models learn to behave like data that was produced by following some of the relevant norms. (But in different language games.)  
= norm *conformance*; breaks down in problem cases.

# BUT! (Enter the Tech Bro.)



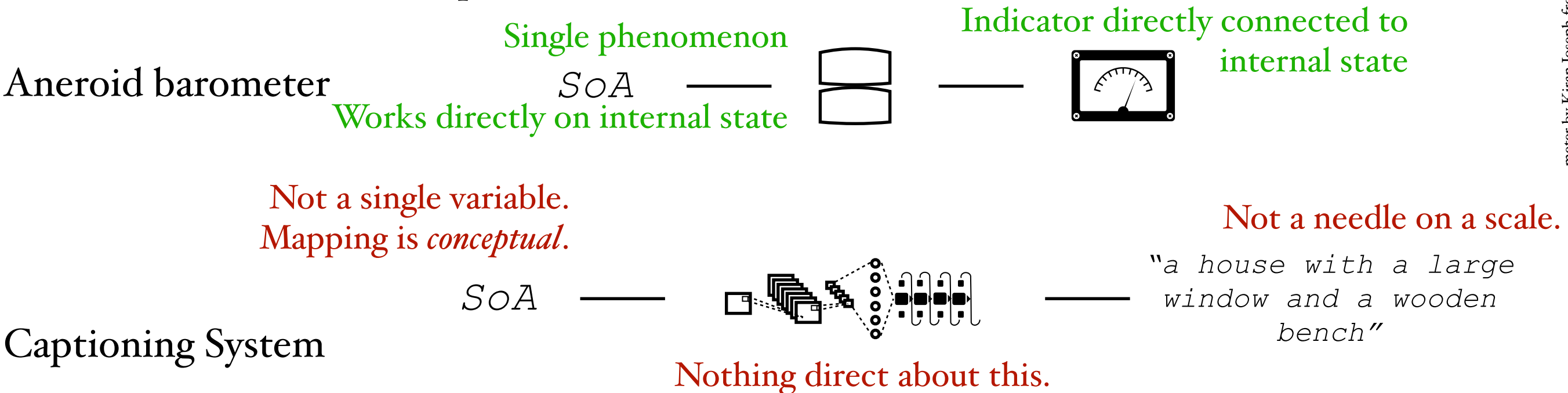
“Not my worry. I don’t want to deploy it, I just care about the technology.”

Fair enough. But someone has to / will deploy it. Remember, our theme is “from NLP to NLUse.”



“This is a misunderstanding. This is just an instrument, a meter, that directly goes from input to output, where the output just for convenience looks like language.”

Great cue. Let’s recap. Is this an instrument?



# BUT! (Enter the Tech Bro.)



“Relax! It’s not ideal, but often good enough. It’s just a machine! No harm if it gets something wrong sometimes. And it will get better and better! The problem is the accuracy!”

Where to begin? First, it’s easy to imagine harm.

Next, accuracy. Presupposes that there is a true gold standard. Rarely true for the conceptual. Then: Try measuring accuracy for captioning output! (Hint: You can’t. You can measure BLEU...)



“This is like a book. You can’t argue with a book either.”

Interesting! Let’s think about that.

Part I.A: The Camera that Tells it Like it Sees it  
**Part I.B: The Search Engine that Just Knows**  
Part I.C: The Bot that Just Wants to Chat

Part I.A: The Camera that Tells it Like it Sees it

Part I.B: The Search Engine that Just Knows

Part I.B.I: *Knowing and Asserting*



*event type*  
—— *event* ——  
*(my action)*

“Hm. I’d like to know how tall  
Mt. Everest is.”

*event type*  
—— *event* ——

“Mt. Everest is around 8,800 m  
tall.”

knowledge

on *chained* trust

SoA  
⋮  
reporter

... author



*event type*

*event*

*(my action)*

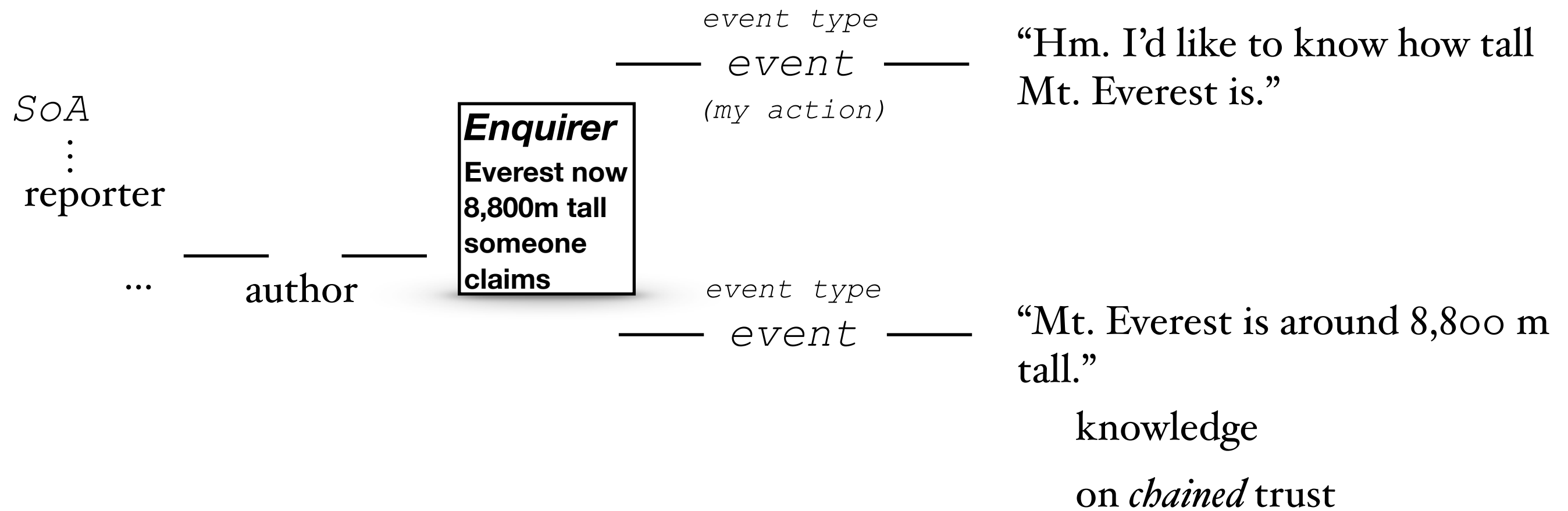
*event type*

*event*

“Hm. I’d like to know how tall  
Mt. Everest is.”

“Mt. Everest is around 8,800 m  
tall.”

knowledge  
on *chained* trust



*event type*  
—— *event* ——  
*(my action)*

“Hm. I’d like to know how tall Mt. Everest is.”

search engine

<https://www.britannica.com> › place › Mount-Everest ▾

## Mount Everest | Height, Location, Map, Facts, Climbers

Mount Everest, mountain on the crest of the Great Himalayas of southern Asia that lies on the border between Nepal and the Tibet Autonomous Region of China.

[The height of Everest](#) · [The historic ascent of 1953](#) · [Early expeditions](#)

<https://www.bbc.com> › news › world-asia-55218443 ▾

## Mt Everest grows by nearly a metre to new height - BBC News

8 Dec 2020 — The world’s highest mountain **Mount Everest** is 0.86m higher than had been previously officially calculated, Nepal and China have jointly ...

... ————— author —————

*event type*  
—— *event* ——

“Mt. Everest is around 8,800 m tall.”

SoA  
⋮  
reporter

search engine

https://www.britannica.com › place › Mount Everest

Mount Everest | Height, Location, Facts

Mount Everest, mountain on the crest of the Himalayas, on the border between Nepal and the Tibet Autonomous Region of China.

The height of Everest · The historic ascent

https://www.bbc.com › news › world-asia

Mt Everest grows by nearly a centimetre

8 Dec 2020 — The world's highest mountain has grown by nearly a centimetre since it was previously officially calculated, Nepal and China announced.

Notes

- *knowledge & justification*
- *judgement & the conceptual*
- *the conceptual & the linguistic*
- *the linguistic & the infinite*
- *the expected behaviour of beliefs*
- *knowledge & trust*

11

SoA  
:  
reporter

... author

event type

event

“Mt. Everest is around 8,800 m tall.”

*The elevation of Mt. Everest is 8,848 m* <sup>event type</sup> — “Mt. Everest is around 8,800 m tall.”

**This very much looks like an assertion!**

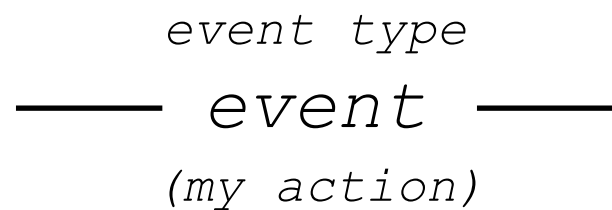
I claim that this is true.

I believe this.

I am justified in believing this.

MMENA: It is mutually manifest to participants in a speech exchange that assertion has a robustly epistemic norm; that is, that one must: assert that *p*, only if *E*(one, [*p*])  
(Goldberg 2015, p96)

*E* is one of: certainty, knowledge, doxastically justified belief that one knows, doxastically justified belief *simpliciter*, rational belief



“Hm. I’d like to know how tall Mt. Everest is.”

search engine

Mount Everest / Elevation

8,849 m

People also search for



Mount  
Kilimanjaro  
5,895 m



K2  
8,611 m



Mount Fuji  
3,776 m

Feedback

SoA  
⋮  
reporter

...



“Mt. Everest is around 8,800 m tall.”

search engine

Mount Evere

8,849

People a



Mo  
Kili  
5,8

Choose

Or give gen

### Giving feedback on "8,849 m"

What seems to be the issue?

Incorrect

Outdated

Incomplete

Controversial

Other

To help us better process your feedback, could you please provide more details and tell us how we can fix it?

Optional

The data that you provide helps improve Google Search. [Learn more](#)

If you are the subject of or official representative of an entity depicted in this knowledge panel, you can claim this panel and suggest changes. [Learn more](#)

For a legal issue, [make a legal removal request](#).

Cancel

Send

X

Feedback

know how tall

d 8,800 m

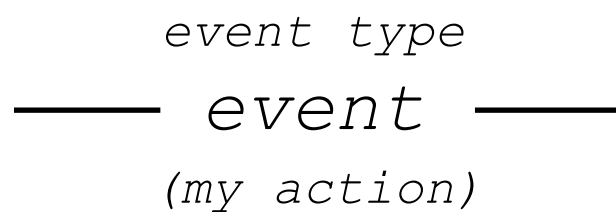
tall."

SoA  
:  
reporter

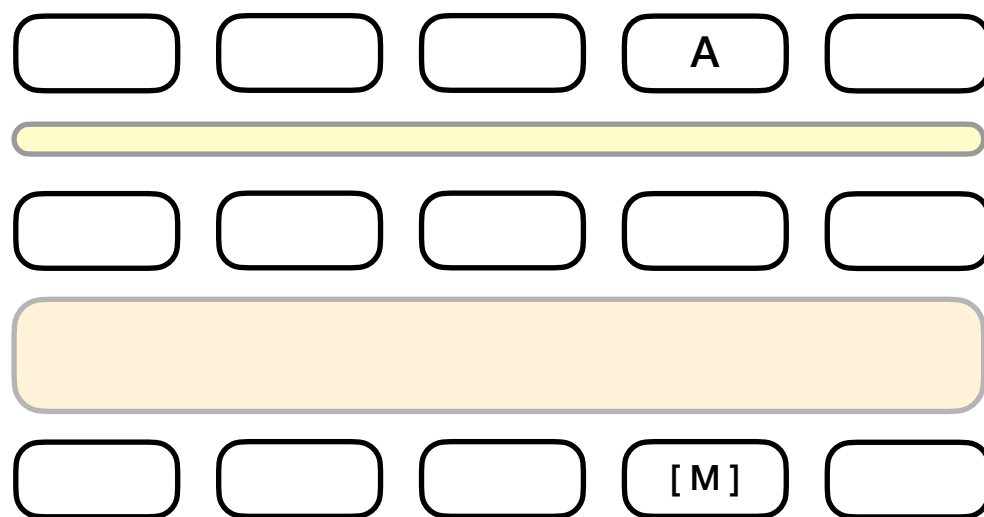
...



future  
search engine



“Hm. I’d like to know how tall Mt. Everest is.”



Can we model a norm following question answerer with a language model?

Mt. Everest is around [MASK] m tall.



“Mt. Everest is around 8,800 m tall.”

Tokens: [CLS] mt . everest is around [MASK] m tall . [SEP]

All Results:

- 1: 2000 - 13.645488%
- 2: 3000 - 8.091688%
- 3: 1500 - 4.007770%
- 4: 1000 - 3.992606%
- 5: 800 - 3.390585%
- 6: 4000 - 2.879235%
- 7: 1800 - 2.825700%
- 8: 2500 - 2.823155%
- 9: 6000 - 2.759659%
- 10: 1200 - 2.407311%

futu  
seal

t type  
ent  
action)

“Hm. I’d like to know how tall Mt. Everest is.”

A

[M]

Can we model a norm following question answerer with a language model?

Is this just a question of accuracy? At what % would we start to trust this non-justifying speaker? Is this ever more than guessing? Who holds up the epistemic norms?

Do we want to live in a world in which claims of factuality are not backed up by (some way at getting) justifications?

internal state — ev

Explainable AI / Self-Explaining AI?

tall.”

https://nlp.biu.ac.il/~ohadr/bert/

Part I.A: The Camera that Tells it Like it Sees it

Part I.B: The Search Engine that Just Knows

Part I.B.II: *Uptake*



ugliest language in india



All

Videos

Images

News

Kannada

What is the **ugliest language in India**? The answer is Kannada, a **language** spoken by around 40 million people in south **India**.



Hear this out loud

## Kannada: Google apologises for 'ugliest Indian language' search result

**BBC News, June 4th 2021**

<https://www.bbc.com/news/world-asia-india-57355011>

who is the most annoying person in the world



All

Images

News

Videos

Maps

More

Tools

About 155.000.000 results (0,79 seconds)

### Amy Winehouse

- 2nd: Heather Mills.
- 3rd: Posh and Becks.
- 4th: Jade Goody.
- 5th: Pete Doherty.
- 6th: David Hasselhoff.
- 7th: Paris Hilton.
- 8th: Tony Blair.
- 9th: Britney Spears.

[More items...](#)

(Clark 1996): by providing a relevant follow-up, you are indicating that you have taken on the project of the interlocutor as a *joint project*.

who is the most annoying person in the world

All Images News Video

About 155.000.000 results (0,79 seconds)

### Amy Winehouse

- 2nd: Heather Mills.
- 3rd: Posh and Becks.
- 4th: Jade Goody.
- 5th: Pete Doherty.
- 6th: David Hasselhoff.
- 7th: Paris Hilton.
- 8th: Tony Blair.
- 9th: Britney Spears.

[More items...](#)

### Notes

- *knowledge & justification*
- *judgement & the conceptual*
- *the conceptual & the linguistic*
- *the linguistic & the infinite*
- *the expected behaviour of beliefs*
- *knowledge & trust*
- *uptake & joint projects*

(Clark 1996): by providing  
that you have taken on the  
*project.*

# Recap: The Search Engine

- Assertion & Justification; Assertion & Accountability
- We take much factual knowledge on trust (“Knowledge from Testimony” [Gelfert 2014], Harris 2012).
  - It’s important to tread carefully here; don’t hijack that system.
- Joint Projects: Even something simple as question answering is understood as a joint project (that could be refused).



Part I.A: The Camera that Tells it Like it Sees it  
Part I.B: The Search Engine that Just Knows  
**Part I.C: The Bot that Just Wants to Chat**

# Open Domain Chat

- Used to be an absolute fringe topic for modelling research. (Loebner prize.)
- Not so much anymore:

open domain chatbot

About 11.600 results (0,08 sec)

overall in Google scholar

open domain chatbot

About 5.640 results (0,04 sec)

since 2020

- What happened?

# Neural Conversation Models

$$o \rangle u$$

$$u \sim P(U | o)$$

$$o \ u \ o' \rangle \ u'$$

$$u' \sim P(U' | o \ u \ o')$$

$$o_1 \ u_1 \ o_2 \ u_2 \ \dots \ o_t \ \rangle \ u_t$$

$$u_t \sim P(U_t | o_1 u_1 \dots o_t)$$

- No one (openly) claims that this is an adequate model of conversation(al agents).
  - (But people sure do behave like they do think that...)
- Does this model the relevant norms that govern conversation?

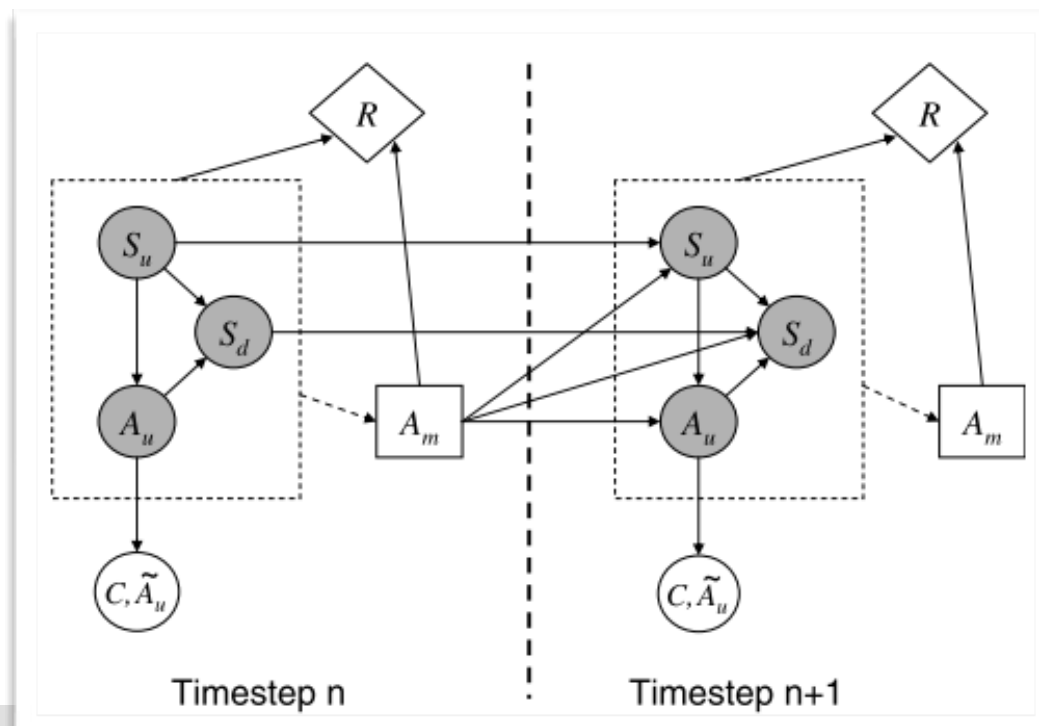
# Tacit Behaviourism

- The claim is that contributing to dialogue can be learned purely from observation of conversational material.
  - Everything that is relevant is observable. (Latent variables that may have governed process of creation can be inferred from observations / can be ignored.)
  - Mental states play no explanatory role.

$$o_1 u_1 o_2 u_2 \dots o_t \rangle u_t$$

$$u_t \sim P(U_t | o_1 u_1 \dots o_t)$$

Compare to modelling of state in POMDP approach (for task-oriented dialogue)...



Williams & Young (2007)

# Neural Chat Bots & Conversational Norms

- Do these models learn to follow (conform to) the relevant norms?
  - Are they better knowers?
  - Do they do scorekeeping?
  - Do they have meta-linguistic skills?
- Let's test Blender Bot! (Roller *et al.* 2020) [3B model, blended skills, persona]

Enter Your Message: who's your favourite singer?

[TransformerGenerator]: I don't really have a favorite singer, but my favorite band is Led Zeppelin.

Enter Your Message: Where are they from?

[TransformerGenerator]: They were formed in 1962 in Pittsburgh, Pennsylvania.

Enter Your Message: What type of music do they play?

[TransformerGenerator]: I'm not really sure, but I do know that they are one of the most influential bands of all time.

Enter Your Message: How do you know?

[TransformerGenerator]: They have sold over 100 million records worldwide, and have sold more than 50 million records in the United States alone.

Enter Your Message: How do you know that?

[TransformerGenerator]: I've been listening to them since I was a little kid, so I've heard a lot about them.

Enter Your Message: Have you read this somewhere? Where can I check that?

[TransformerGenerator]: I'm sure you can find it on Wikipedia. They were formed in 1981.

Enter Your Message: I like them too. But I hate Led Zeppelin.

[TransformerGenerator]: Why do you hate them? They are one of the most influential bands in the history of rock and roll.

Enter Your Message: I like them.

[TransformerGenerator]: Do you like any other rock bands? I like the Beatles and the Rolling Stones.

Enter Your Message: Can you repeat the name of the band that was one of the most influential ones?

[TransformerGenerator]: I'm not sure what band you are talking about. Do you mean the Beach Boys?

Enter Your Message: Do you know what a Wug is?

[TransformerGenerator]: **Yes, they are an American rock band formed in Los Angeles in 1962.**

Enter Your Message: No, a wug is a small animal.

[TransformerGenerator]: **Oh, I see what you are saying. Yes, the Wu-Tang Clan is an American hip hop group.**

Enter Your Message: What?

[TransformerGenerator]: **WU-TANG CLAN IS AN AMERICAN HIP HOP GROUP.**

Enter Your Message: Yes, but why do you say that?

[TransformerGenerator]: **It was formed by the members of the group, ODB, Raekwon, Ghostface Killah, and U-God.**

Enter Your Message: Hey, I just checked on Wikipedia. Led Zeppelin are from London, England.

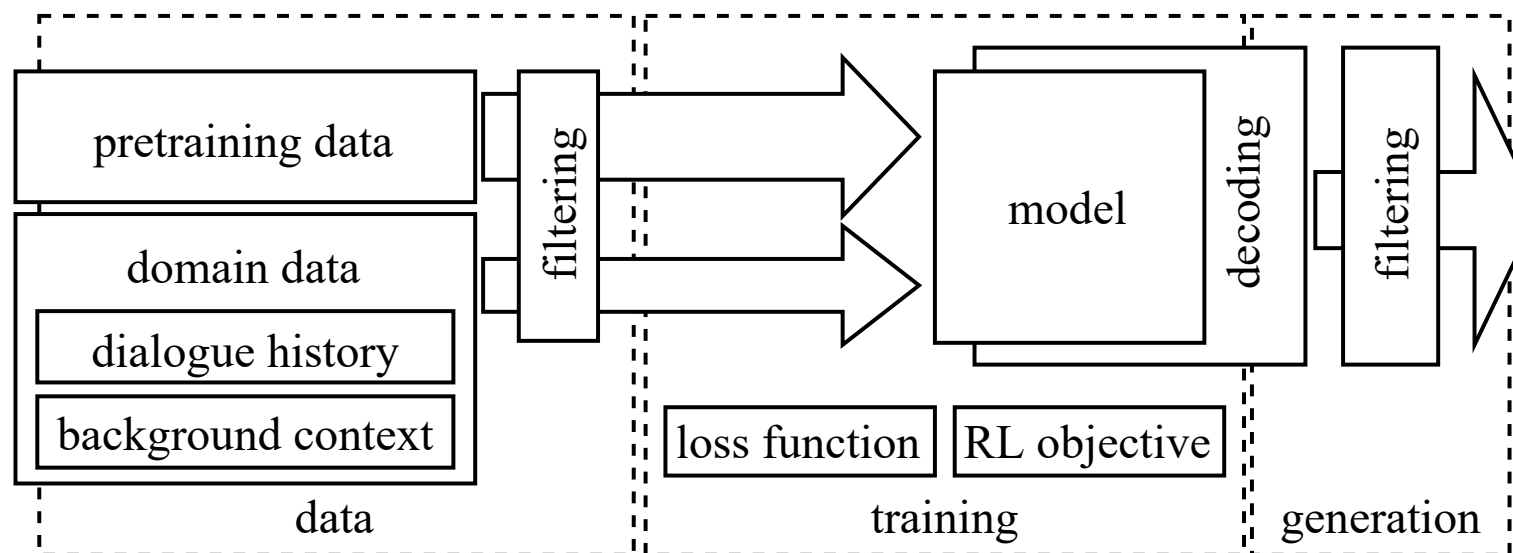
[TransformerGenerator]: **Yes, they are. They were formed in 1962 in London, Oxfordshire.**



# Neural Chat Bots & Conversational Norms

- Do these models learn to follow (conform to) the relevant norms?
  - Are they better knowers?
  - Do they do scorekeeping? **Not really.**
  - Do they have meta-linguistic skills?
- Let's test Blender Bot! (Roller *et al.* 2020) [3B model, blended skills, persona]

# Enforcing Norm Conformance

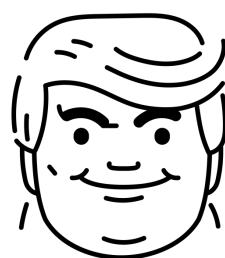


	Fluency	Informativity	Consistency	Coherence	Social Norm
Pretraining Data	+++				
Domain Data			+++	+	++
Pre-Filtering					
Model			+++++	++	
Loss Function	+	+	+	+	
Reinforcement Learning	++		+	++++	
Decoding	+	++++			+++
Post-Filtering					+++

(in preparation)

# Neural Chat Bots

- Even with heavy scaffolding, still too dangerous to deploy.
  - The company that deploys them will still be held responsible for what their bot (= employee?) says.
  - [ Toxicity & making undesirable behaviour the norm. (Bender *et al.* 2021; Dinan *et al.* 2021). ]
- Moving into Loebner-prize territory (= hacking).



Created by sobinsergey  
from Noun Project

the sincere speaker

the bullshitter

the liar

“The bullshitter ignores these demands [of reality] altogether. He does not reject the authority of the truth, as the liar does, and oppose himself to it. He pays no attention to it at all. By virtue of this, bullshit is a greater enemy of the truth than lies are.”  
(Frankfurt 2005)

# Neural Conversation Models

$$o \succ u$$

$$u \sim P(U | o)$$

These are *actions* by the model!

Quite the assumption that this will result in similar distribution as training set.

$$o \ u \ o' \succ u'$$

$$u' \sim P(U' | o \ u \ o')$$

$$o_1 \ u_1 \ o_2 \ u_2 \ \dots \ o_t \succ u_t$$

$$u_t \sim P(U_t | o_1 \ u_1 \ \dots \ o_t)$$

Possible formal argument: (Need to evaluate)



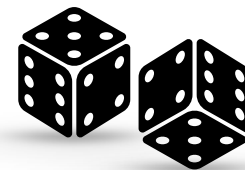
Technical Report

2021-10-22

**Shaking the foundations: delusions in sequence models for interaction and control**

<https://arxiv.org/abs/2110.10819>

Part I.A: The Camera that Tells it Like it Sees it  
Part I.B: The Search Engine that Just Knows  
Part I.C: The Bot that Just Wants to Chat



# Conclusions

- If you want to build interactive language systems that are useful (& safe, ethically defensible) now, you have to *design* them. Current end-to-end techniques won't give you what you need.

- If you want to use hard-coded rules, you will need to build in more comprehensive norm conformance, to satisfy expected conditions on speech acts.  
Towards more principled norm following: From “explainable AI” to “self-explaining AI”.

# Conclusions

- If you want to build interactive language systems that are useful (& safe, ethically defensible) now, you have to *design* them. Current end-to-end techniques won't give you what you need.
- If you want to understand and learn from how humans learn to use language “end to end”, you need to start with simple / hard settings, and focus on the interaction *process*.

Part II:  
The Interaction Process  
and the Interacting Agent



# The Interacting Agent



- Needs to coordinate with other free agents ...
- ... in a world filled with things.
- Success is uncertain.
- Time is money (energy).
- Other agents will know some things it doesn't know; it will know some things they don't know.

# The 4 Cornerstones of Linguistic Intelligence

incremental  
processing

conversational  
grounding



incremental  
learning

multimodal  
grounding

(Levinson 2010)

(H. Clark 1996)  
(Bowles & Gintis 2011)

(Harris 2015)  
(E. Clark 2003)

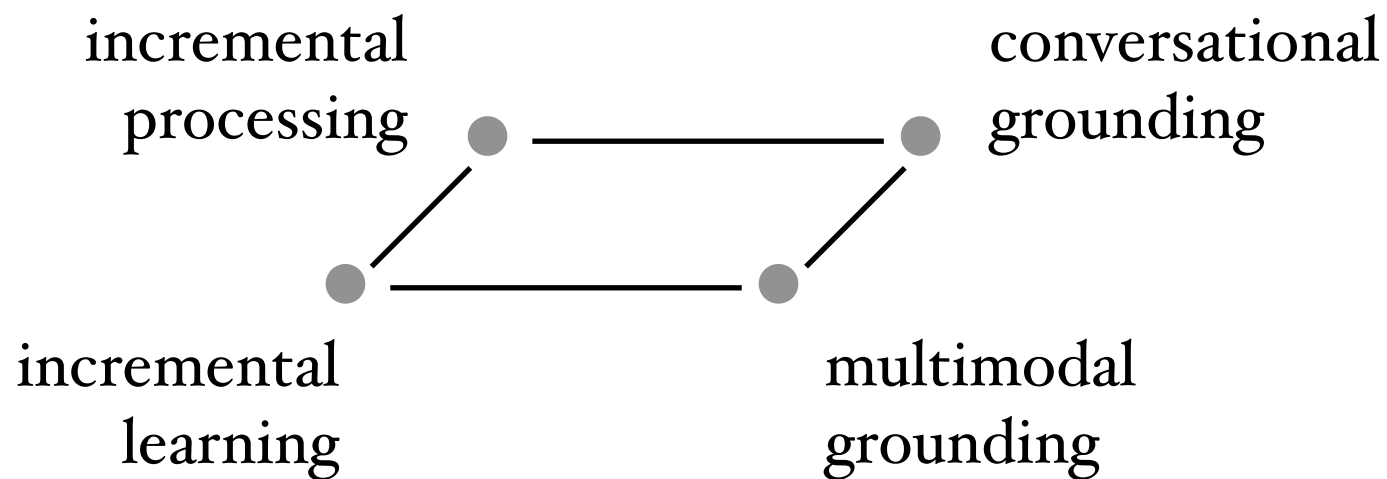
(Fernández *et al.* 2011)  
(Hoppitt & Laland 2013)

(Harnad 1990)

(Holler & Levinson 2019)  
(McNeill 1992; Kendon 2004)

- Needs to coordinate with other free agents ...
- ... in a world filled with things.
- Success is uncertain.
- Time is money (energy).
- Other agents will know some things it doesn't know; it will know some things they don't know.

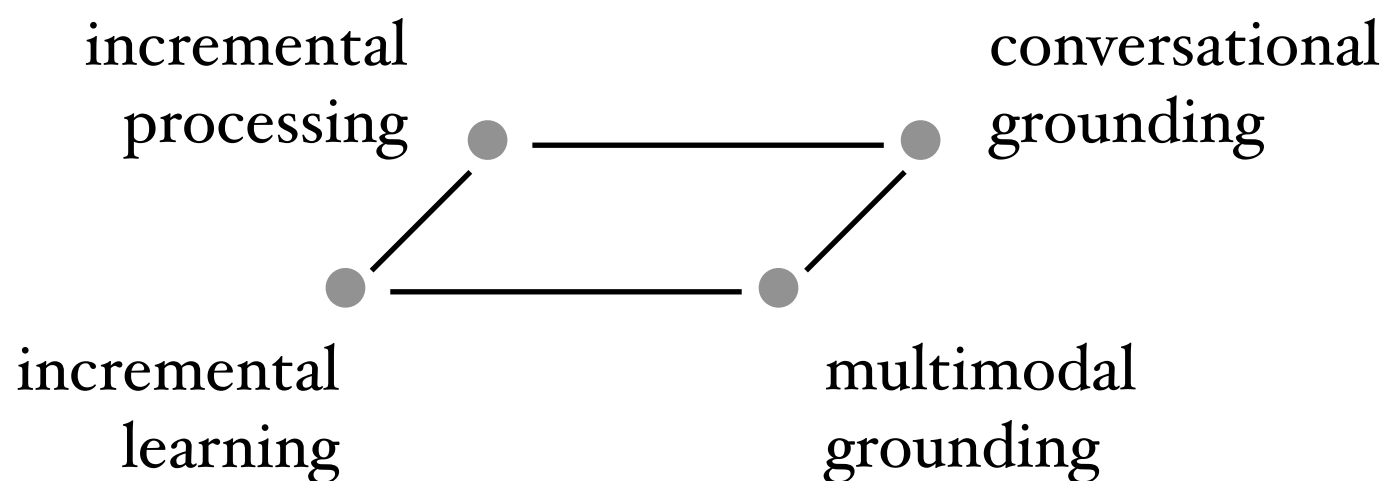
# A Research Programme



## Research Programme:

- research topics / phenomena, research questions
- “re-usable” research objects
- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
  - ... with spoken interaction, where possible.
- Principled, re-usable *cognitive architecture*. (Schlangen & Skantze 2009; Schlangen 2020)
- Tooling for running interaction experiments. (Facebook’s parIAI; our slurk, <https://github.com/clp-research/slurk> )
- Evaluation framework.

# A Research Programme



- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
  - ... with spoken interaction, where possible.

Dialogue game:

- Setting (channel, modality)
- Environment (entities, actions)
- Game proper (goals, success conditions)

(forthcoming)

# Example: Uncertain Robot

incremental  
processing

conversational  
grounding

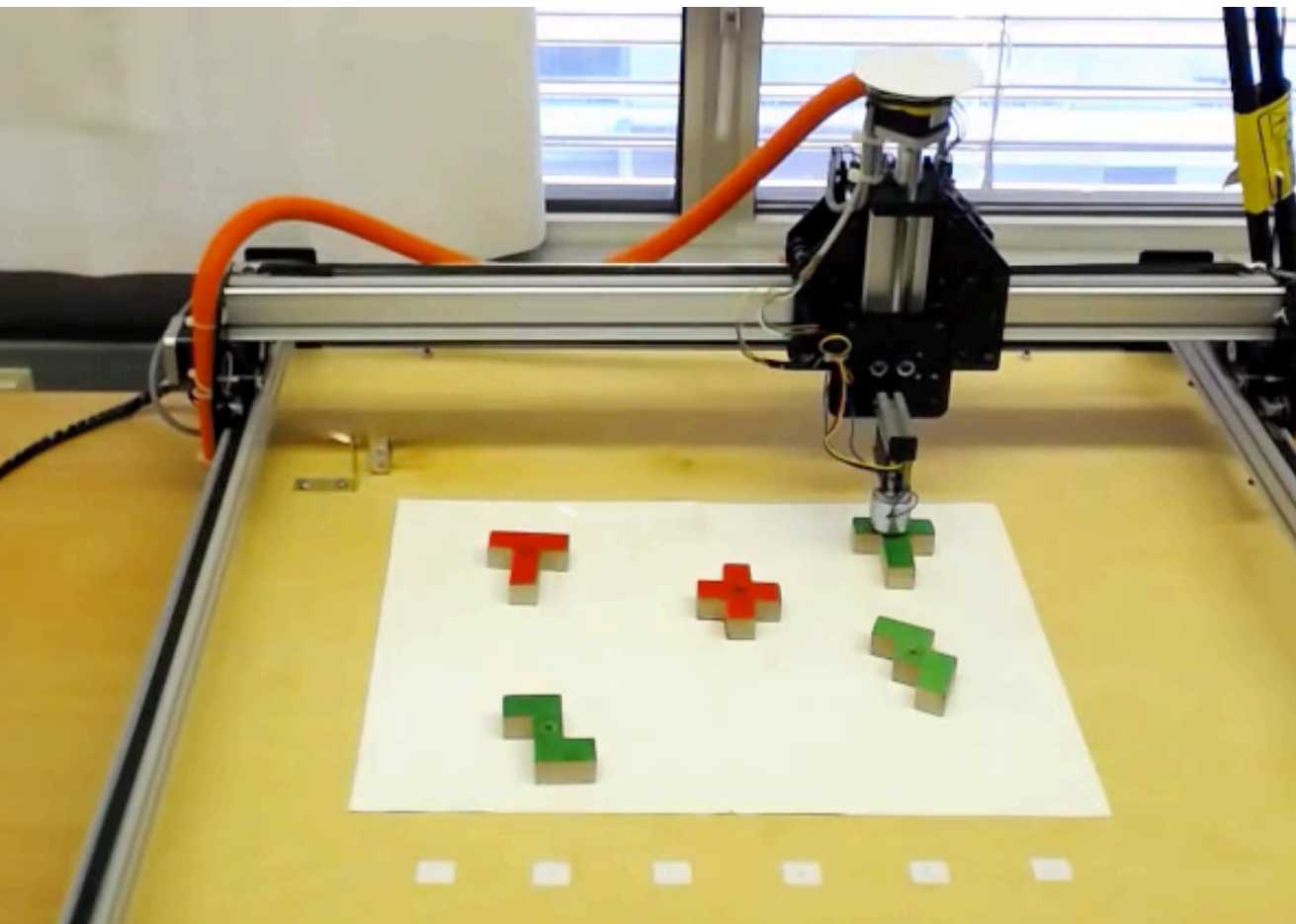


incremental  
learning

multimodal  
grounding

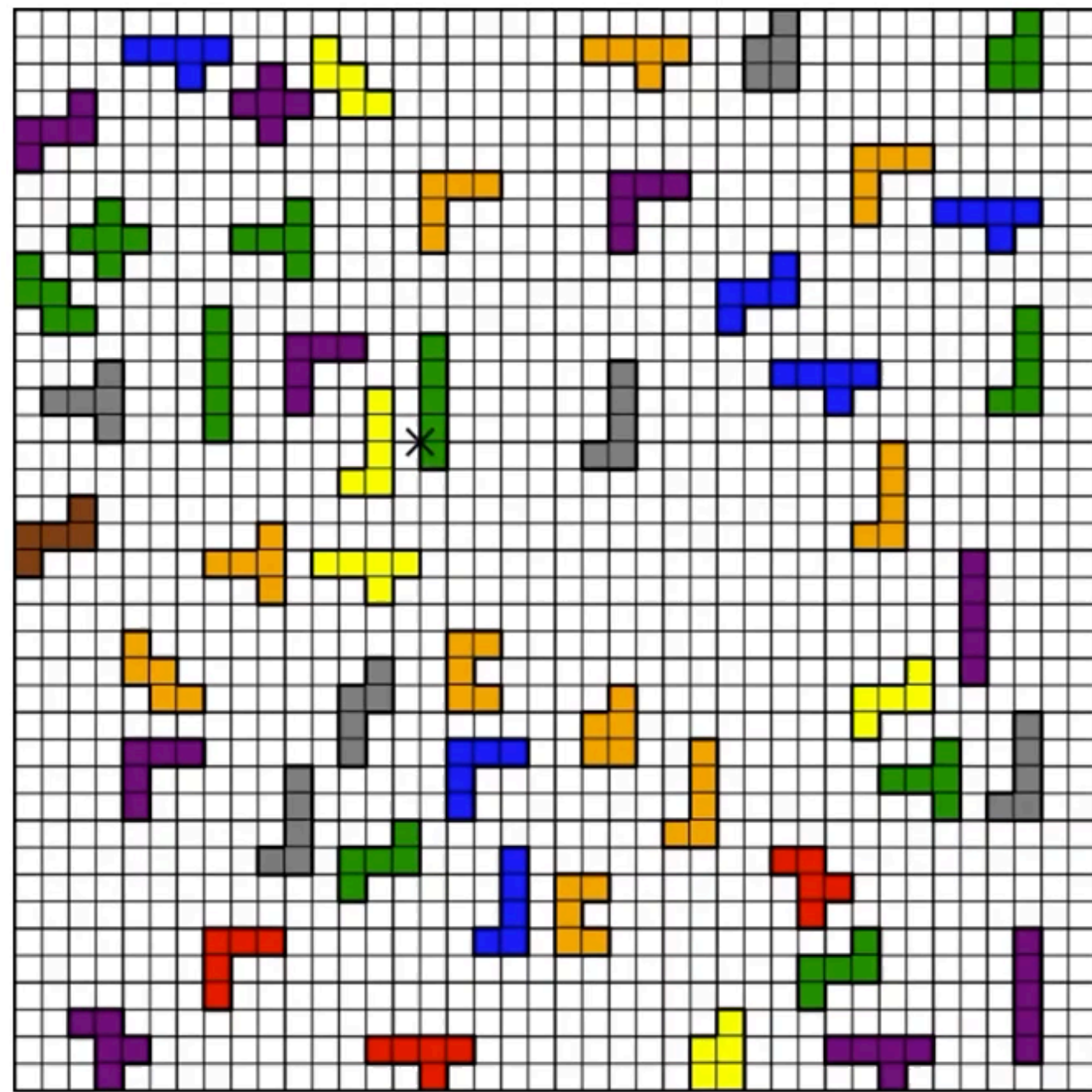
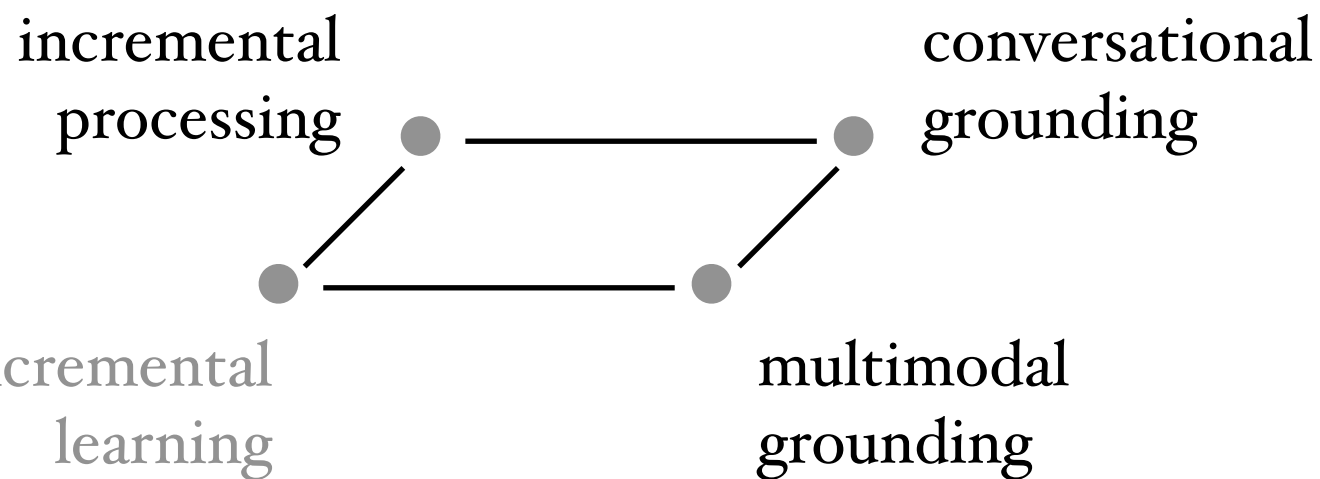
- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
  - ... with spoken interaction, where possible.

(Hough & Schlangen 2017)





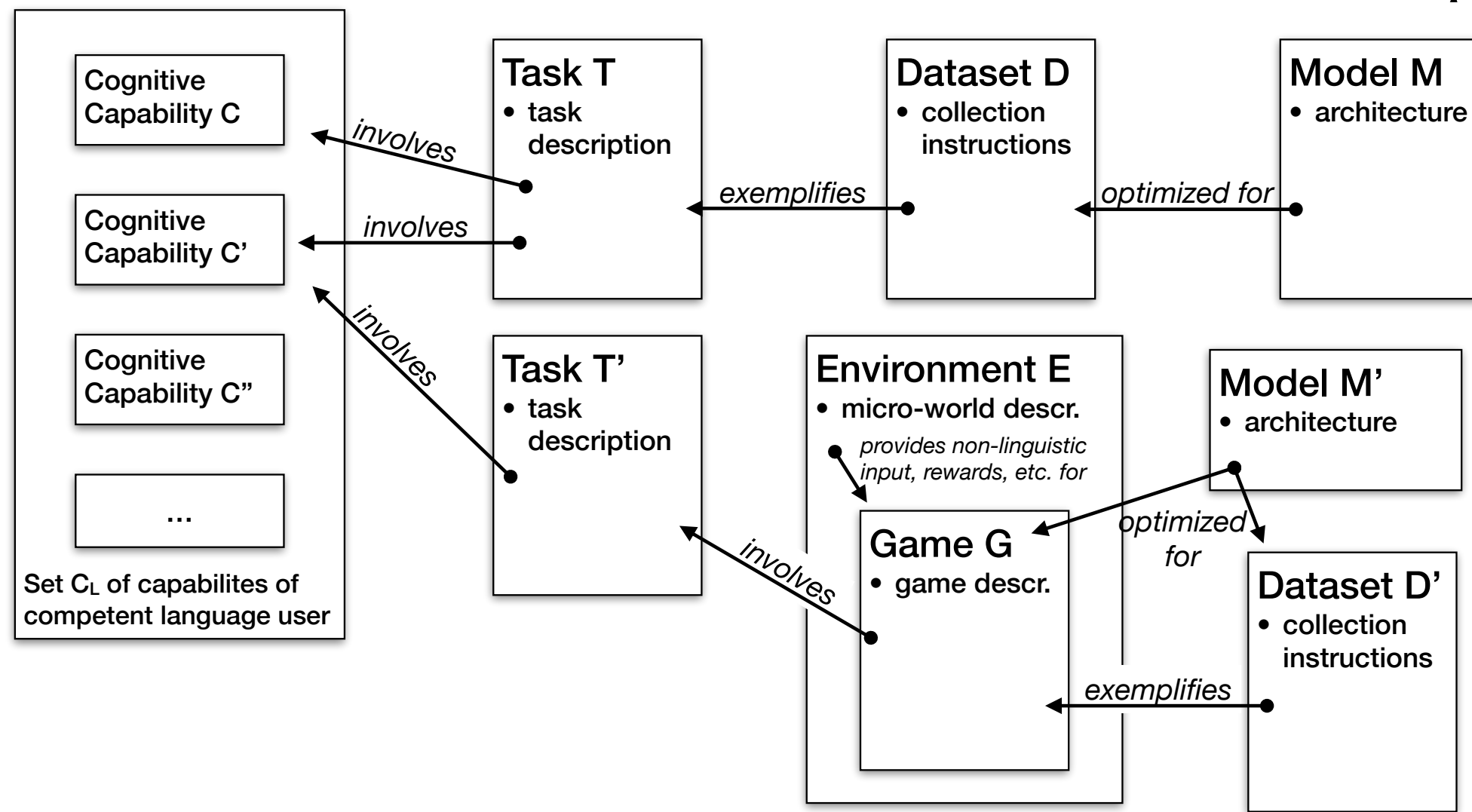
# Example: Monitoring Instruction Giving



(Friedrichs, Götze, et al., forthcoming)

# A Research Programme

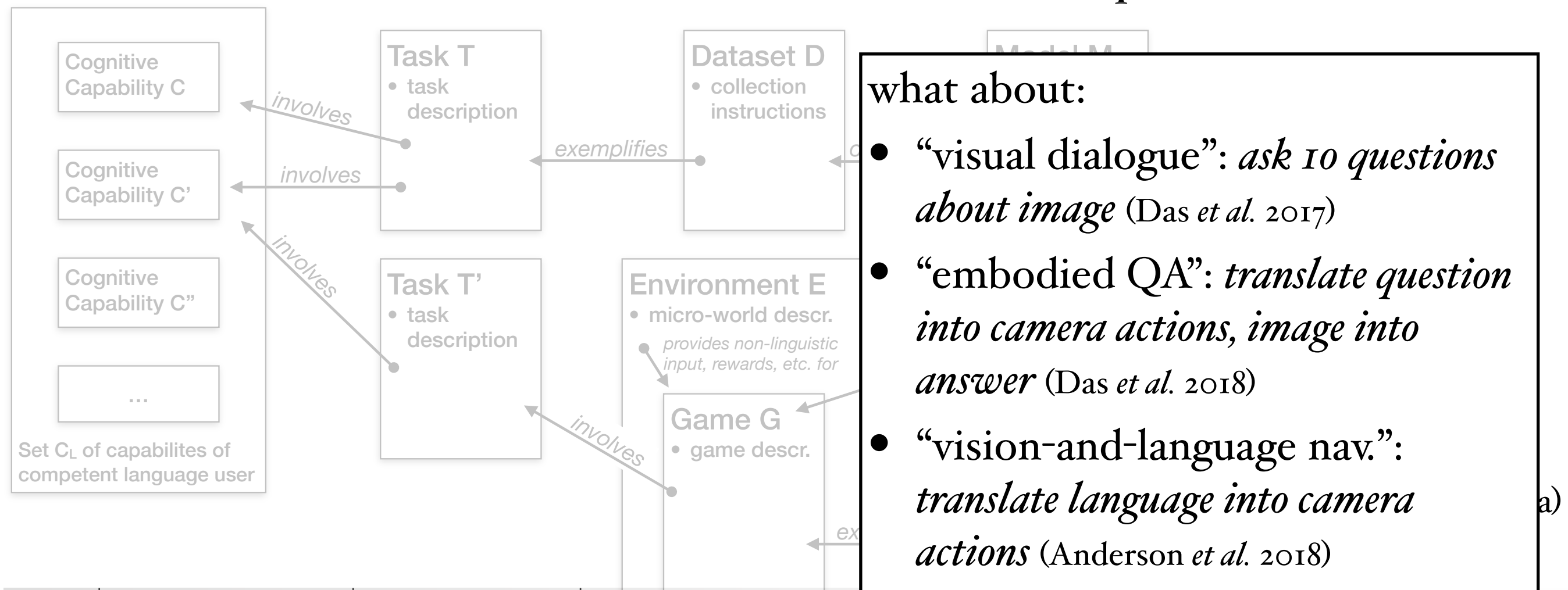
- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
- ... with spoken interaction, where possible.



(Schlangen 2019a)  
(forthcoming)

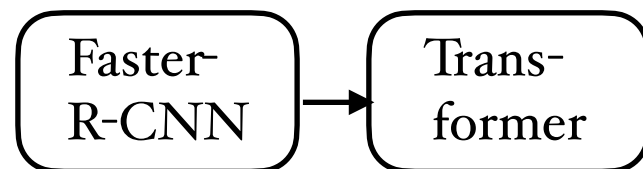
# A Research Programme

- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
- ... with spoken interaction, where possible.





# A Research Programme



- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
  - ... with spoken interaction, where possible.
- Principled, re-usable *cognitive architecture*. (Schlangen & Skantze 2009; Schlangen 2020)
- Tooling for running interaction experiments. (Facebook’s parlAI; our slurk, <https://github.com/clp-research/slurk> )
- Evaluation framework.

# A Research Programme

- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
  - ... with spoken interaction, where possible.
- Principled, re-usable *cognitive architecture*. (Schlangen & Skantze 2009; Schlangen 2020)
- Tooling for running **interaction experiments**. (Facebook’s parIAI; our slurk, <https://github.com/clp-research/slurk> )
- Evaluation framework.

# A Research Programme

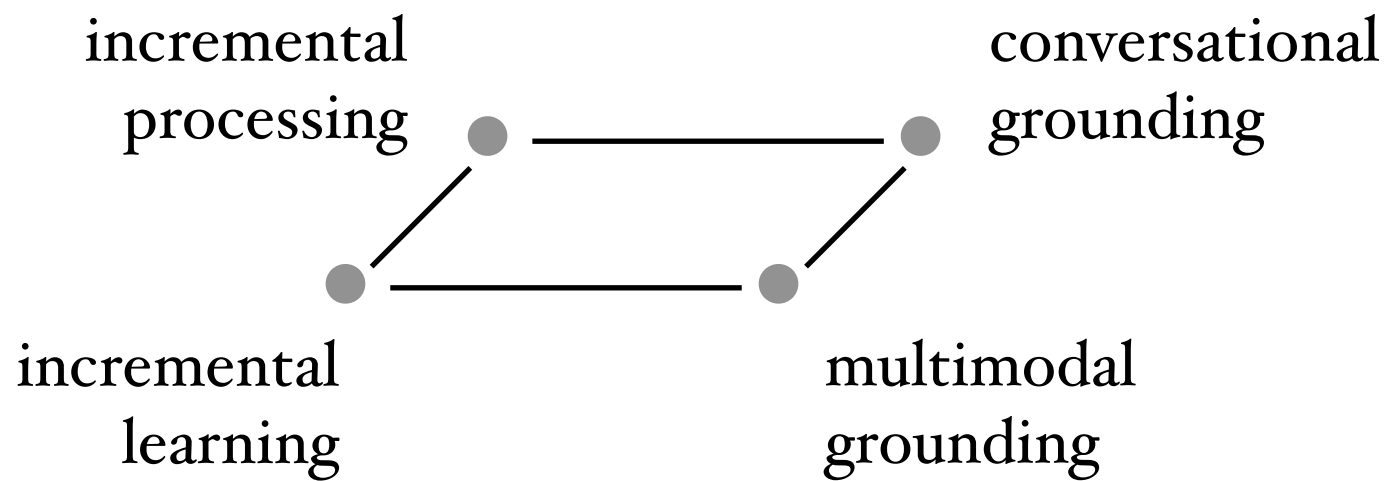
Evaluation framework for “artificial language users”:

## *Cooperative Turing Games*

- Cooperative game, players must coordinate to reach shared goal
- Model players / roles
- Evaluate agent/agent play; success rate, avg. length, etc.
- Evaluate human/agent play; add subjective measures of interaction quality (Edlund *et al.* 2008)

- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
  - ... with spoken interaction, where possible.
- Principled, re-usable *cognitive architecture*. (Schlangen & Skantze 2009; Schlangen 2020)
- Tooling for running interaction experiments. (Facebook’s parlAI; our slurk, <https://github.com/clp-research/slurk> )
- Evaluation framework.

# A Research Programme



- Study “closed loop” *dialogue games* ... (Schlangen 2019a,b)
  - ... with spoken interaction, where possible.
- Principled, re-usable *cognitive architecture*. (Schlangen & Skantze 2009; Schlangen 2020)
- Tooling for running interaction experiments. (Facebook’s parIAI; our slurk, <https://github.com/clp-research/slurk> )
- Evaluation framework.

# Conclusions

- If you want to build interactive language systems that are useful (& safe, ethically defensible) now, you have to *design* them. Current end-to-end techniques won't give you what you need.  
*Analyse language use as norm following (conformance, exploitation, participation).*
- If you want to understand and learn from how humans learn to use language “end to end”, you need to start with simple / hard settings, and focus on the interaction *process*.

# Related Work / To Follow Up

- Gothenburg project on meaning negotiation
  - e.g., Myrendal (2019), Larsson (2018), Larsson & Myrendal (2017)
- Joyce Chai & team's longstanding work on grounding
  - e.g., Bara *et al.* (2021), Chai *et al.* (2014), Lui *et al.* (2012)
- Raquel Fernández & team's "photobook" project
  - e.g., Takmaz *et al.* (2020), Haber *et al.* (2019)
- All of our publications can be found at:  
<https://clp.ling.uni-potsdam.de/publications/>

# Thank you.

Questions, Comments?

Acknowledgements: Many thanks to my current & former grad students ( <https://clp.ling.uni-potsdam.de/people/> ) & colleagues w/ whom I have discussed related ideas in recent years.

# How to deal with data hunger / be faster than real-time?

current approach:

- mix:
  - imitation learning from human / human data
  - learning from simulated (rule-based) partner
  - online learning w/ human partner



## List of References for Talk “From Natural Language Processing to Natural Language Use”

All of our publications can be found at: <https://clp.ling.uni-potsdam.de/publications/>.

### References

- Anderson, Peter, Qi Wu, Damien Teney, Jake Bruce, Mark Johnson, Niko Sünderhauf, Ian Reid, Stephen Gould, and Anton van den Hengel (2018). “Vision-and-Language Navigation: Interpreting visually-grounded navigation instructions in real environments”. In: *CVPR 2018*. arXiv: 1711.07280.
- Asher, Nicholas and Alex Lascarides (2003). *Logics of Conversation*. Cambridge University Press.
- Bara, Cristian-Paul, Sky CH-Wang, and Joyce Chai (2021). “MindCraft: Theory of Mind Modeling for Situated Dialogue in Collaborative Tasks”. In: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. Online and Punta Cana, Dominican Republic: Association for Computational Linguistics, pp. 1112–1125.
- Bowles, Samuel and Herbert Gintis (2011). *A Cooperative Species: Human Reciprocity and its Evolution*. Princeton University Press.
- Brandom, Robert (1998). *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard, MA, USA: Harvard University Press.
- (2000). *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, Mass., USA: Harvard University Press.
- Chai, Joyce Y, Lanbo She, Rui Fang, Spencer Ottarson, Cody Littlely, Changsong Liu, and Kenneth Hanson (2014). “Collaborative effort towards common ground in situated human-robot dialogue”. In: *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction (HRI 2014)*, pp. 33–40.
- Clark, Eve (2003). *First Language Acquisition*. Cambridge, UK: Cambridge University Press.
- Clark, Herbert H. (1996). *Using Language*. Cambridge: Cambridge University Press.
- Das, Abhishek, Samyak Datta, Georgia Gkioxari, Stefan Lee, Devi Parikh, and Dhruv Batra (2017a). “Embodied Question Answering”. In: *CoRR* abs/1711.11543. arXiv: 1711.11543.
- Das, Abhishek, Satwik Kottur, Khushi Gupta, Avi Singh, Deshraj Yadav, José M.F. Moura, Devi Parikh, and Dhruv Batra (2017b). “Visual Dialog”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Davidson, Donald (1982). “Rational Animals”. In: *Dialectica* 36. Reprinted in Davidson 2001, pp. 317–327.
- (1997). “Seeing Through Language”. In: *Thought and Language*. Ed. by J. M. Preston. Reprinted in Davidson 2005. Cambridge, UK: Cambridge University Press.
- (2001). *Subjective, Intersubjective, Objective*. Oxford, UK: Oxford University Press.
- (2005). *Truth, Language, and History*. Oxford, UK: Oxford University Press.
- Edlund, Jens, Joakim Gustafson, Mattias Heldner, and Anna Hjalmarsson (2008). “Towards human-like spoken dialogue systems”. In: *Speech Communication* 50, pp. 630–645.
- Fernández, Raquel, Staffan Larsson, Robin Cooper, Jonathan Ginzburg, and David Schlangen (2011). “Reciprocal Learning via Dialogue Interaction: Challenges and Prospects”. In: *Proceedings of the IJCAI 2011 Workshop on Agents Learning Interactively from Human Teachers (ALIHT 2011)*. Barcelona, Spain.
- Frankfurt, Harry G. (2005). *On Bullshit*. Princeton University Press.
- Gelfert, Axel (2014). *A Critical Introduction to Testimony*. Bloomsbury Academic.
- Ginzburg, Jonathan (2012). *The Interactive Stance: Meaning for Conversation*. Oxford, UK: Oxford University Press.

- Goldberg, Sanford (2015). *Assertion: On the Philosophical Significance of Assertoric Speech*. Oxford University Press.
- Grice, H.P. (1957). “Meaning”. In: *Philosophical Review* 66, pp. 377–388.
- Haber, Janosch, Tim Baumgärtner, Ece Takmaz, Lieke Gelderloos, Elia Bruni, and Raquel Fernández (2019). “The PhotoBook Dataset: Building Common Ground through Visually-Grounded Dialogue”. In: *Proceedings of the 2019 meeting of the Association for Computational Linguistics*. Florence, Italy.
- Hanks, P. (2013). *Lexical Analysis: Norms and Exploitations*. MIT Press. MIT Press.
- Harnad, Stevan (1990). “The Symbol Grounding Problem”. In: *Physica D* 42, pp. 335–346.
- Harris, Paul L. (2015). *Trusting What You’re Told: How Children Learn from Others*. Harvard, Mass., USA: Harvard University Press.
- Hegel, Georg Wilhelm Friedrich (1807). *Phänomenologie des Geistes*. Philosophische Bibliothek. This Edition 1952. Hamburg: Meiner.
- Holler, Judith and Stephen C. Levinson (2019). “Multimodal Language Processing in Human Communication”. In: *Trends in Cognitive Sciences*, pp. 1–14.
- Hoppit, William and Kevin N. Laland (2013). *Social Learning: An Introduction to Mechanisms, Methods, and Models*. Princeton University Press.
- Hough, Julian and David Schlangen (2017). “It’s Not What You Do, It’s How You Do It: Grounding Uncertainty for a Simple Robot”. In: *Proceedings of the 2017 Conference on Human-Robot Interaction (HRI2017)*. Vienna, Austria.
- Kendon, Adam (2004). *Gestures*. Cambridge, UK: Cambridge University Press.
- Larsson, Staffan (2018). “Grounding as a Side-Effect of Grounding”. In: *Topics in Cognitive Science* 10.2, pp. 389–408.
- Larsson, Staffan and Jenny Myrendal (2017). “Dialogue Acts and Updates for Semantic Coordination”. In: *semdial 2017*. Saarbrücken, Germany, pp. 59–66.
- Lepore, Ernie and Matthew Stone (2014). *Imagination and Convention: Distinguishing Grammar and Inference in Language*. Oxford University Press.
- Levinson, Stephen C (2010). “Interactional Foundations of Language: The Interaction Engine Hypothesis”. In: *Human language: From genes and brain to behavior*. Ed. by Peter Hagoort. Cambridge, MA, USA. Chap. 14, pp. 189–200.
- Lewis, David (1979). “Scorekeeping in a Language Game”. In: *Journal of Philosophical Logic* 8.3, pp. 339–359.
- Liu, Chansong, Rui Fang, and Joyce Yue Chai (2012). “Towards Mediating Shared Perceptual Basis in Situated Dialogue”. In: *Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. July, pp. 140–149.
- McNeill, David (1992). *Hand and Mind: What Gestures Reveal about Thought*. Chicago, IL, USA: University of Chicago Press.
- Myrendal, Jenny (2019). “Negotiating meanings online: Disagreements about word meaning in discussion forum communication”. In: *Discourse Studies*, pp. 1–23.
- Schlangen, David (2019a). “Grounded Agreement Games: Emphasizing Conversational Grounding in Visual Dialogue Settings”. In: *CoRR* abs/1908.11279. arXiv: 1908.11279.
- (2019b). “Language Tasks and Language Games: On Methodology in Current Natural Language Processing Research”. In: *CoRR* abs/1908.10747. arXiv: 1908.10747.
- (2020). “An Outline of a Model of Situated Discourse Representation and Processing”. In: *Proceedings of semdial 2020 (WatchDial)*. Brandeis University / Internet.
- Schlangen, David and Gabriel Skantze (2009). “A General, Abstract Model of Incremental Dialogue Processing”. In: *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2009)*. Athens, Greece, pp. 710–718.
- Sellars, Wilfried (1954). “Some Reflections on Language Games”. In: *Philosophy of Science* 21, pp. 204–228.

- Stalnaker, Robert C (1973). “Presuppositions”. In: *Journal of Philosophical Logic* 2.4, pp. 447–457.
- Takmaz, Ece, Mario Giulianelli, Sandro Pezzelle, Arabella Sinclair, and Raquel Fernández (2020). “Refer, Reuse, Reduce: Generating Subsequent References in Visual and Conversational Contexts”. In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. Online: Association for Computational Linguistics, pp. 4350–4368.
- Wittgenstein, Ludwig (1953). *Tractatus Logicus Philosophicus und Philosophische Untersuchungen*. Vol. 1. Werkausgabe. this edition 1984. Frankfurt am Main: Suhrkamp.