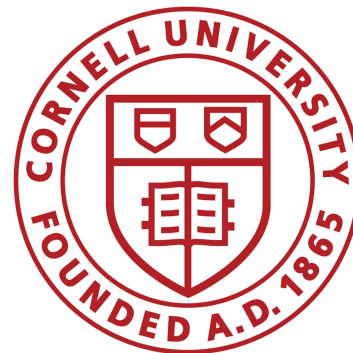


Reference-Centric Models for Grounded Collaborative Dialogue

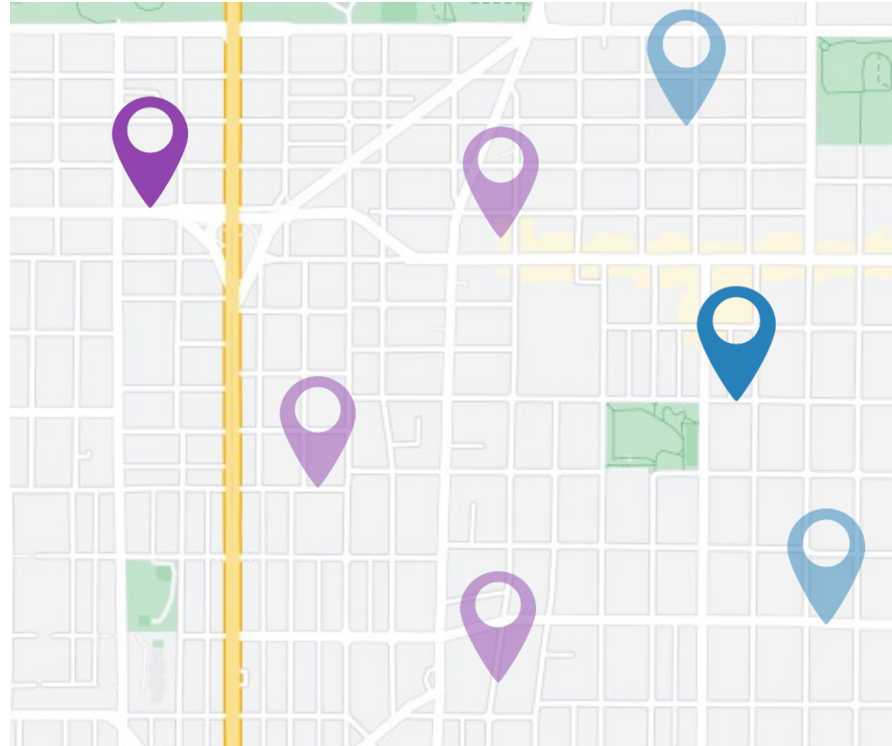
Berkeley



Daniel Fried, Justin Chiu, and Dan Klein



Grounded Collaborative Dialogue



A: Do you know the Thai restaurant west of the highway?

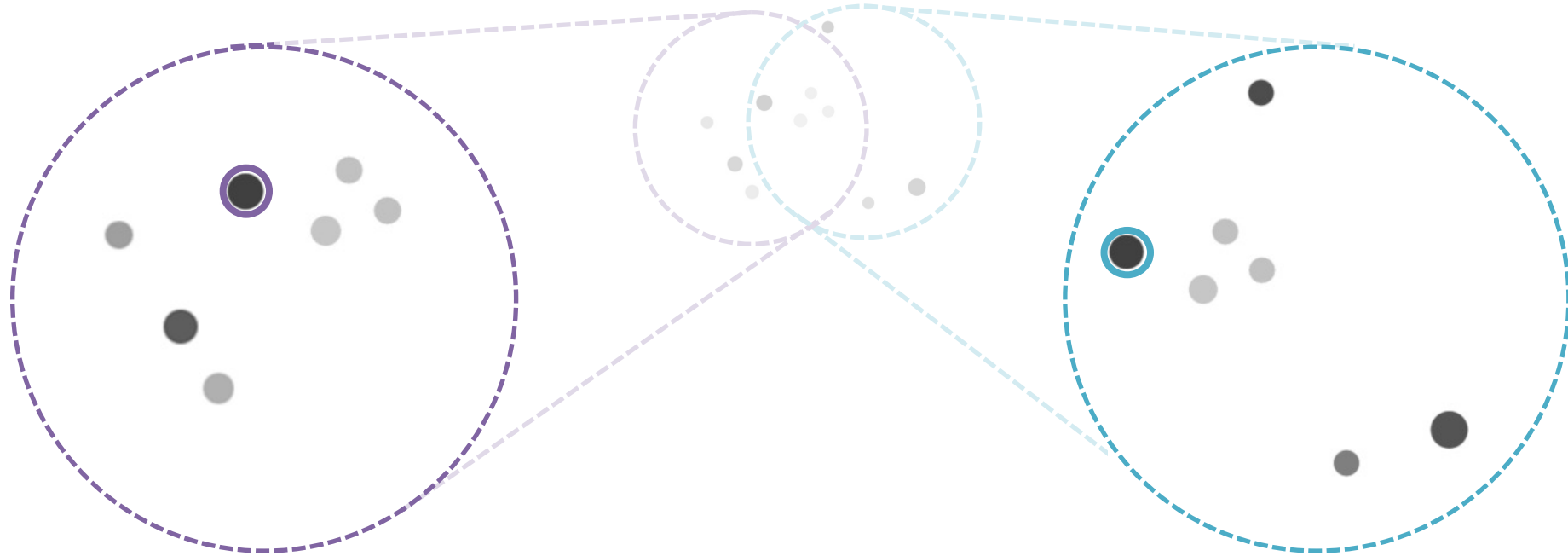
A: I know that one, sounds good!



B: I don't, but how about the coffee place near the park?



Grounded Collaborative Dialogue



A: I have three dots in a line with a dark one in the center.

A: Is there a large black dot to the left of the three grey dots?

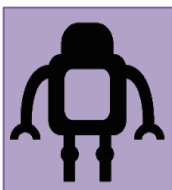
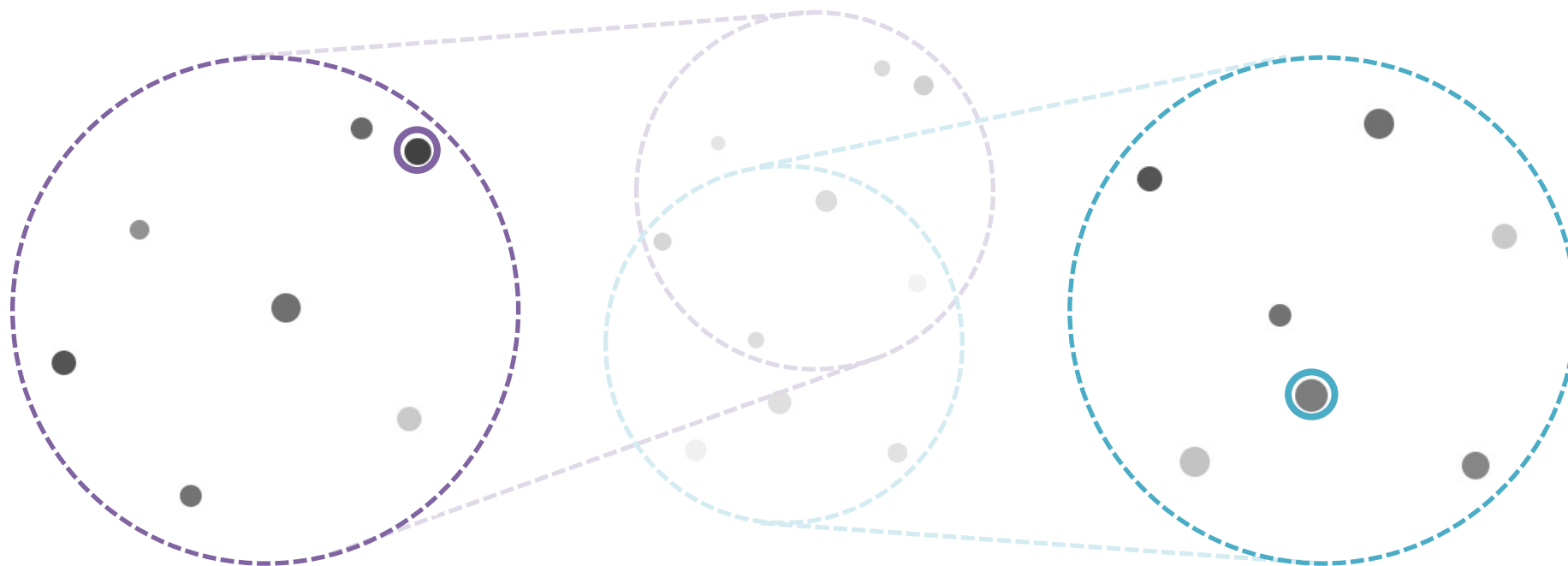


B: I don't have that. Do you have a cluster of three grey dots in a triangle?

B: Yes, let's select the black one.



Previous State-of-the-Art



A: I have a large black dot with a smaller and lighter dot above it and to the right.

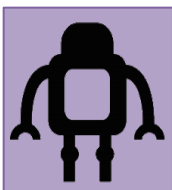
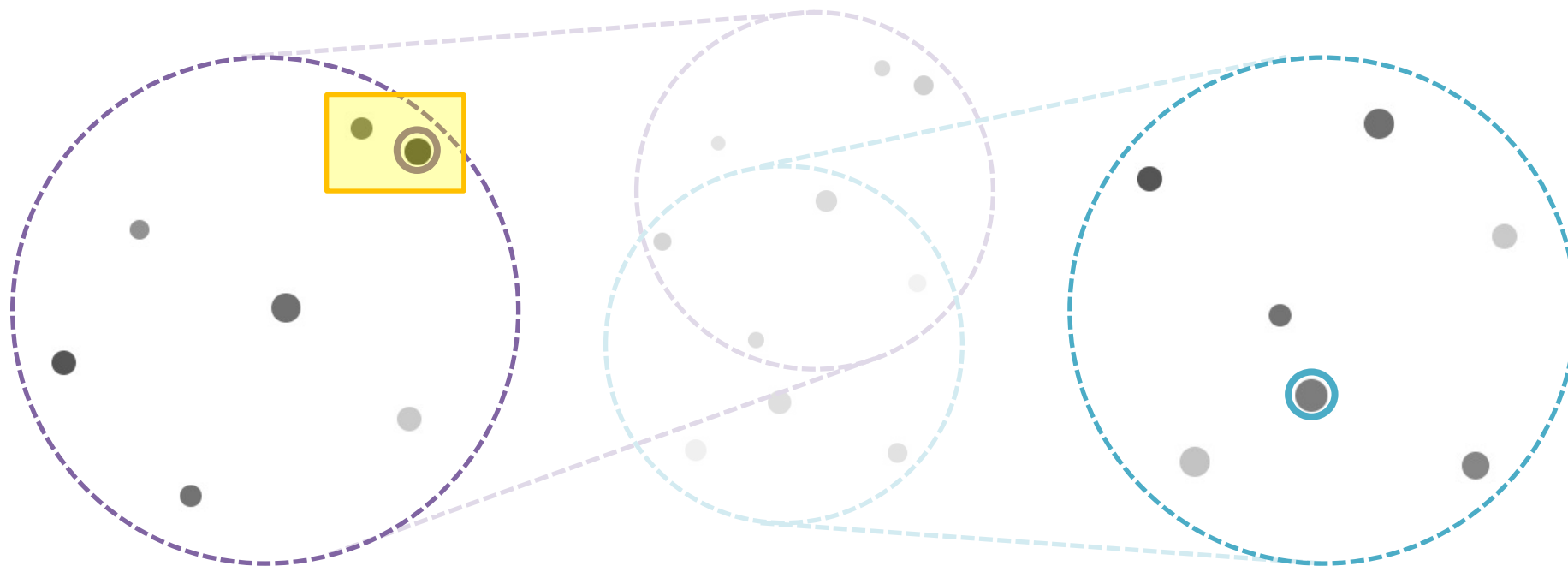
A: Yes, I have that. Let's pick the large black one.



B: I don't have that. How about a dark grey very large dot with a slightly darker dot above and slightly to its left?



Previous State-of-the-Art



A: I have a large black dot with a smaller and lighter dot above it and to the right.

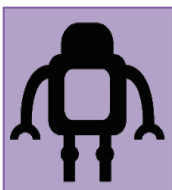
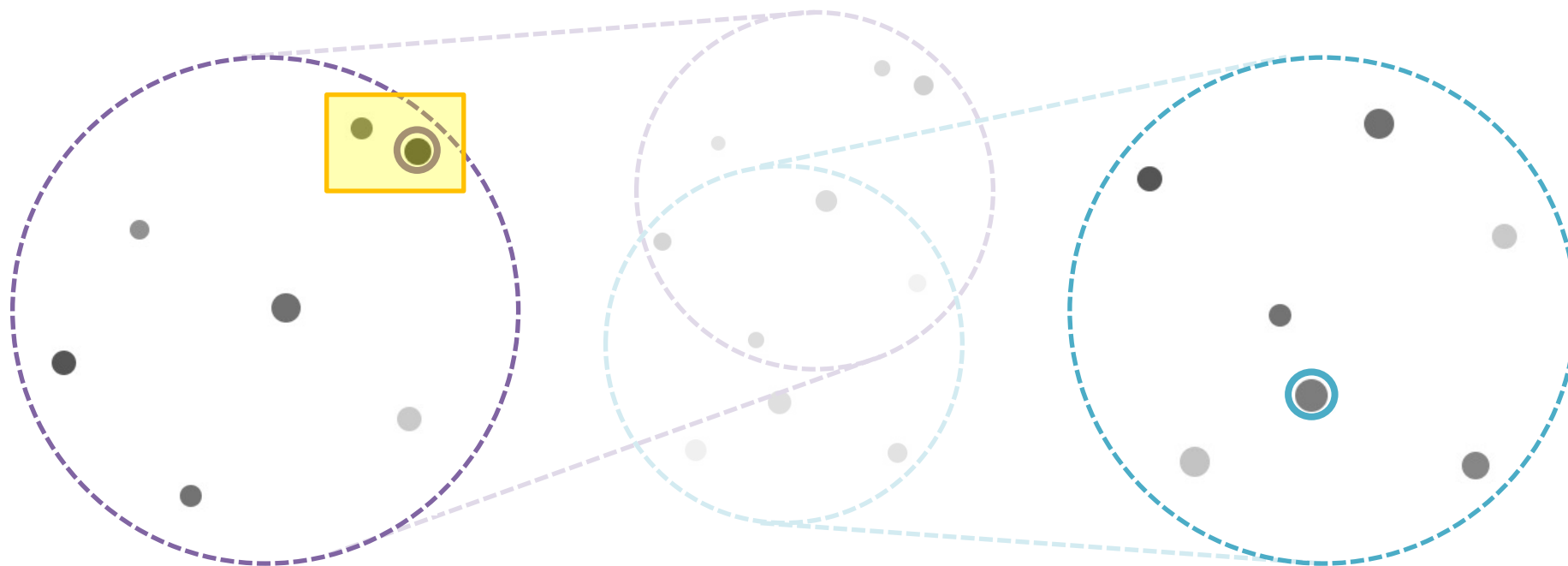
A: Yes, I have that. Let's pick the large black one.



B: I don't have that. How about a dark grey very large dot with *a slightly darker* dot above and slightly to its left?



Previous State-of-the-Art



A: I have a large black dot with a smaller and lighter dot above it and *to the right*.

A: Yes, I have that. Let's pick the large black one.



B: I don't have that. How about a dark grey very large dot with *a slightly darker* dot above and slightly to its left?



Reference-Centric Models

A modular decomposition of the task:

Listener

1) Reference resolution

Controller

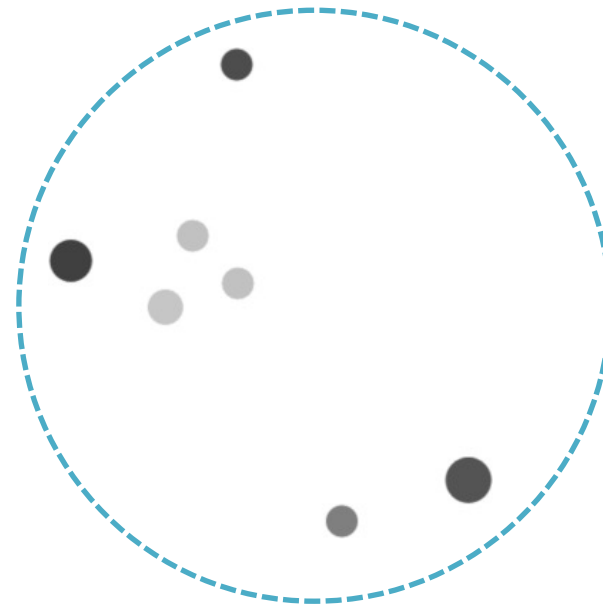
2) Content selection

Speaker

3) Surface realization



Reference-Centric Models



A: I have three dots in a line with a dark one in the center.

B: I don't have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???



Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don't have that. Do you have a group of three grey dots? → **A:** Is there a large black dot to the left of the three grey dots? → **B:???**



Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don't have that. Do you have a group of three grey dots? → **A:** Is there a large black dot to the left of the three grey dots? → **B:???**



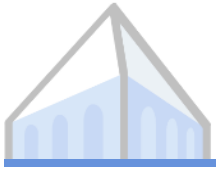
Reference-Centric Models

Reference resolution

Content selection

Surface realization

B: I don't have that. Do you have a group of three grey dots? → A: Is there a large black dot to the left of the three grey dots? → B:???

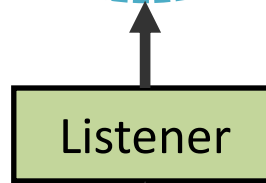
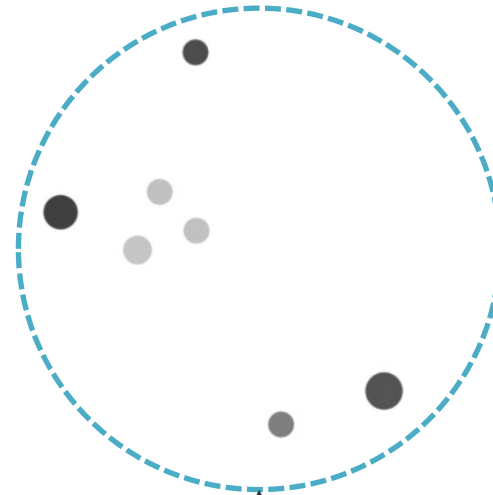


Reference-Centric Models

Reference resolution

Content selection

Surface realization



B: I don't have that. Do you have a group of three grey dots?



A: Is there a large black dot to the left of the three grey dots?



B:???

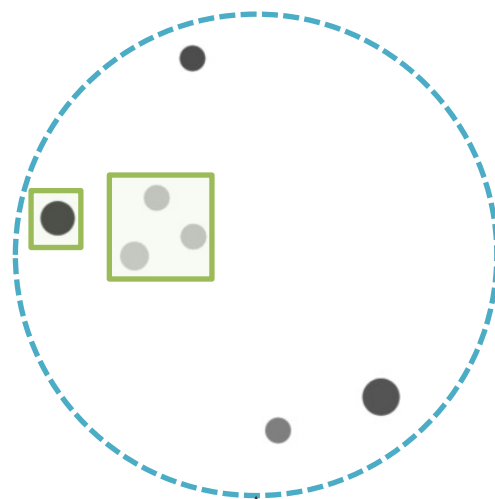


Reference-Centric Models

Reference resolution

Content selection

Surface realization



Listener

B: I don't have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

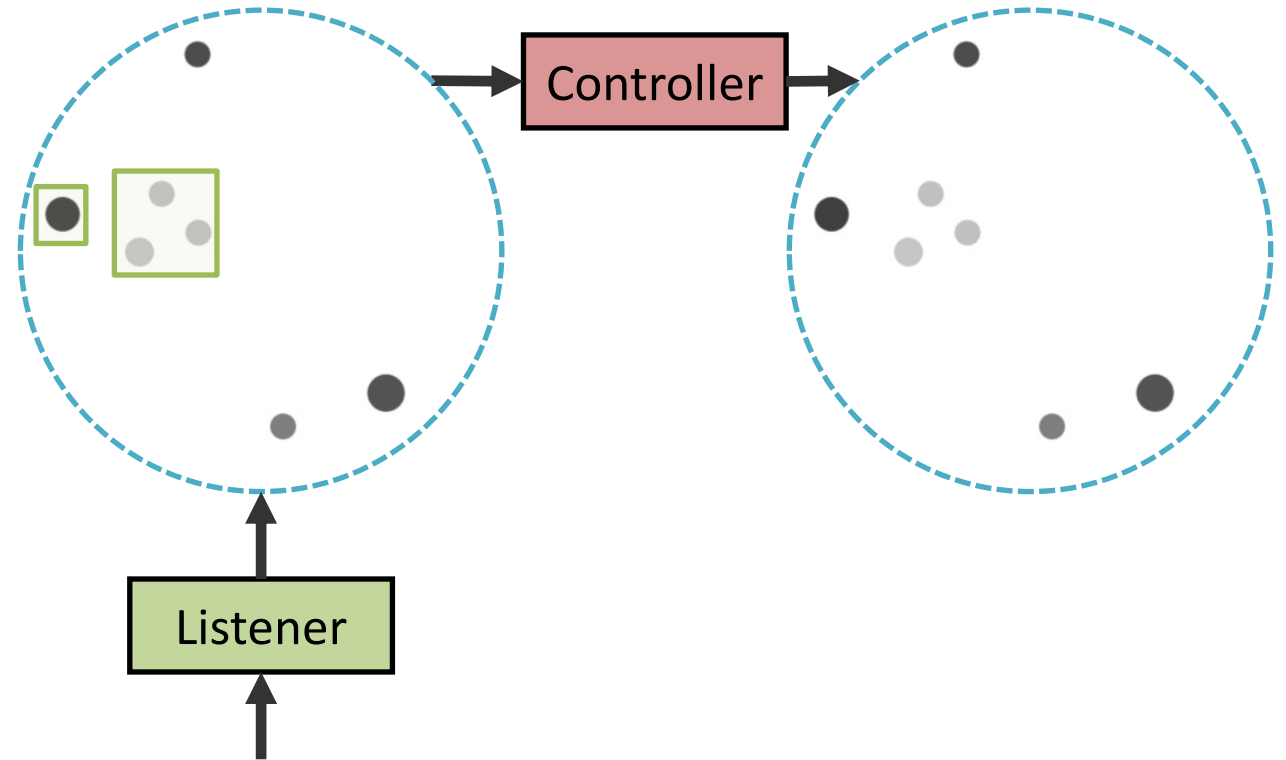


Reference-Centric Models

Reference resolution

Content selection

Surface realization



B: I don't have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

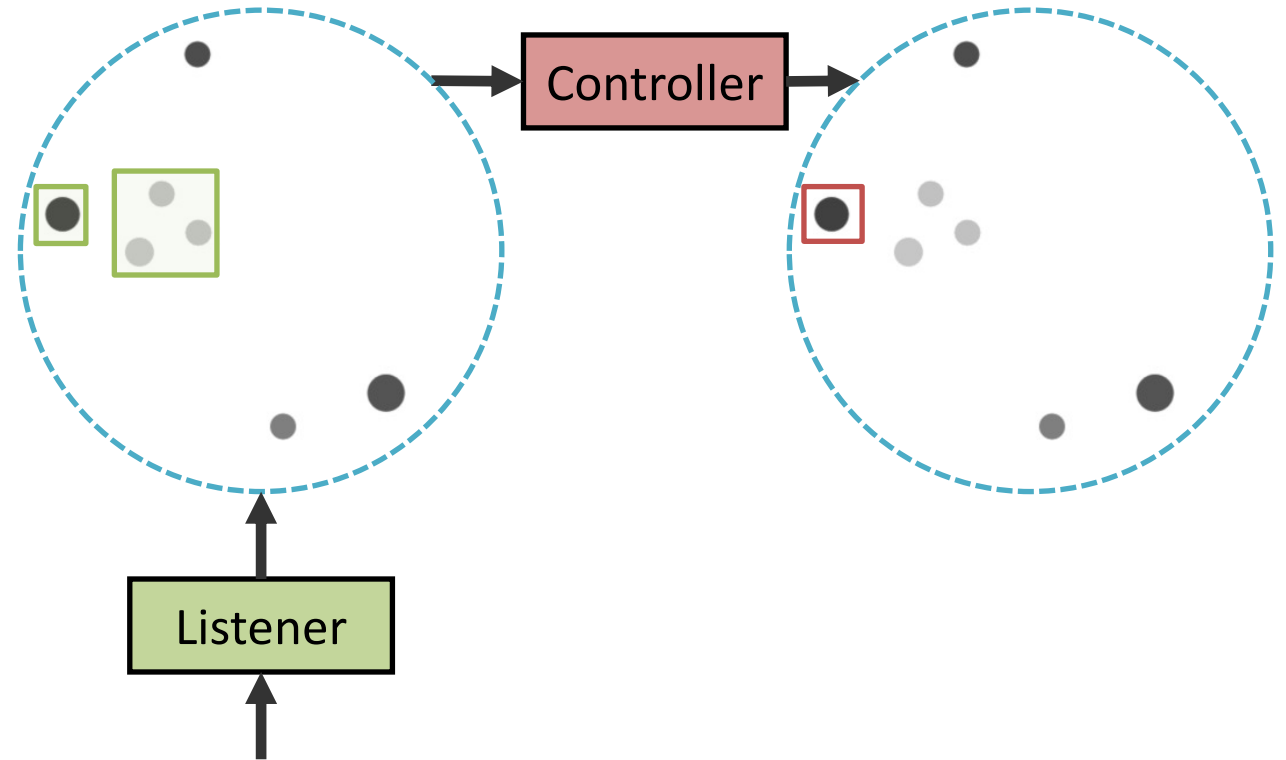


Reference-Centric Models

Reference resolution

Content selection

Surface realization



B: I don't have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

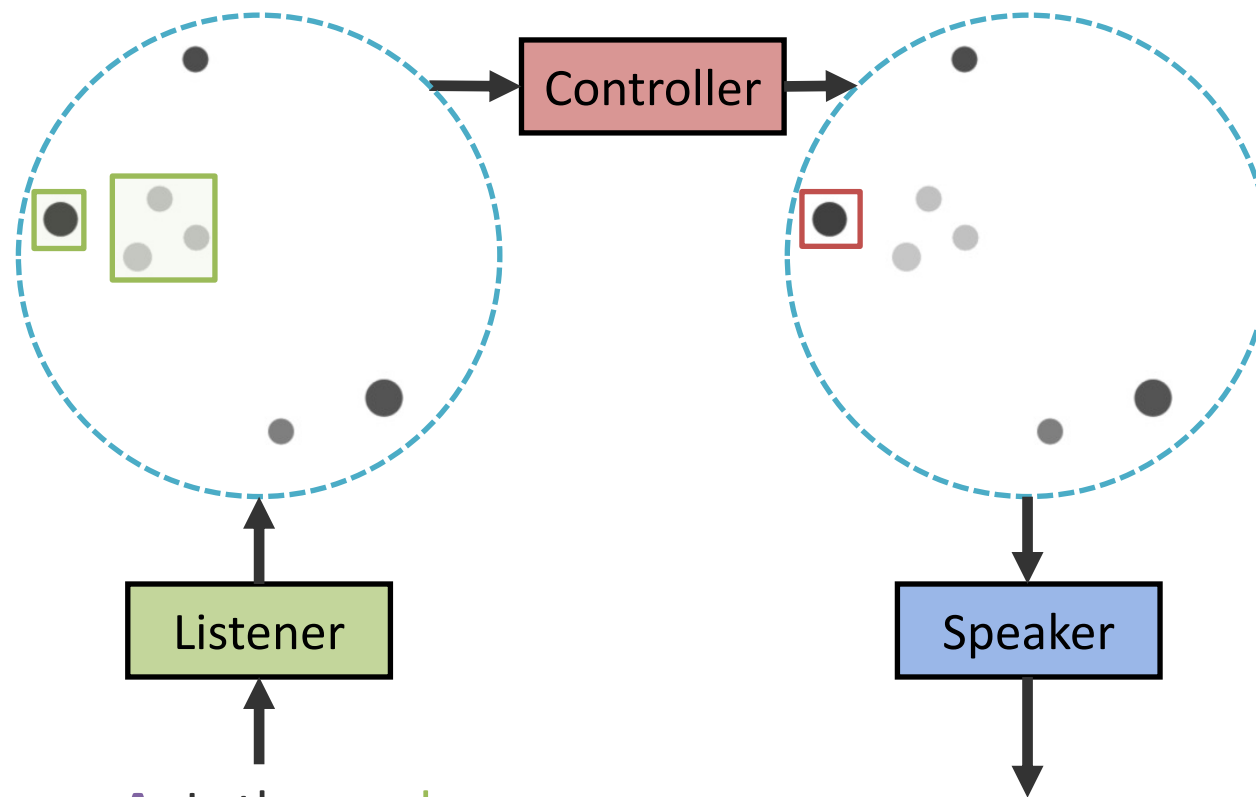


Reference-Centric Models

Reference resolution

Content selection

Surface realization



B: I don't have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B:???

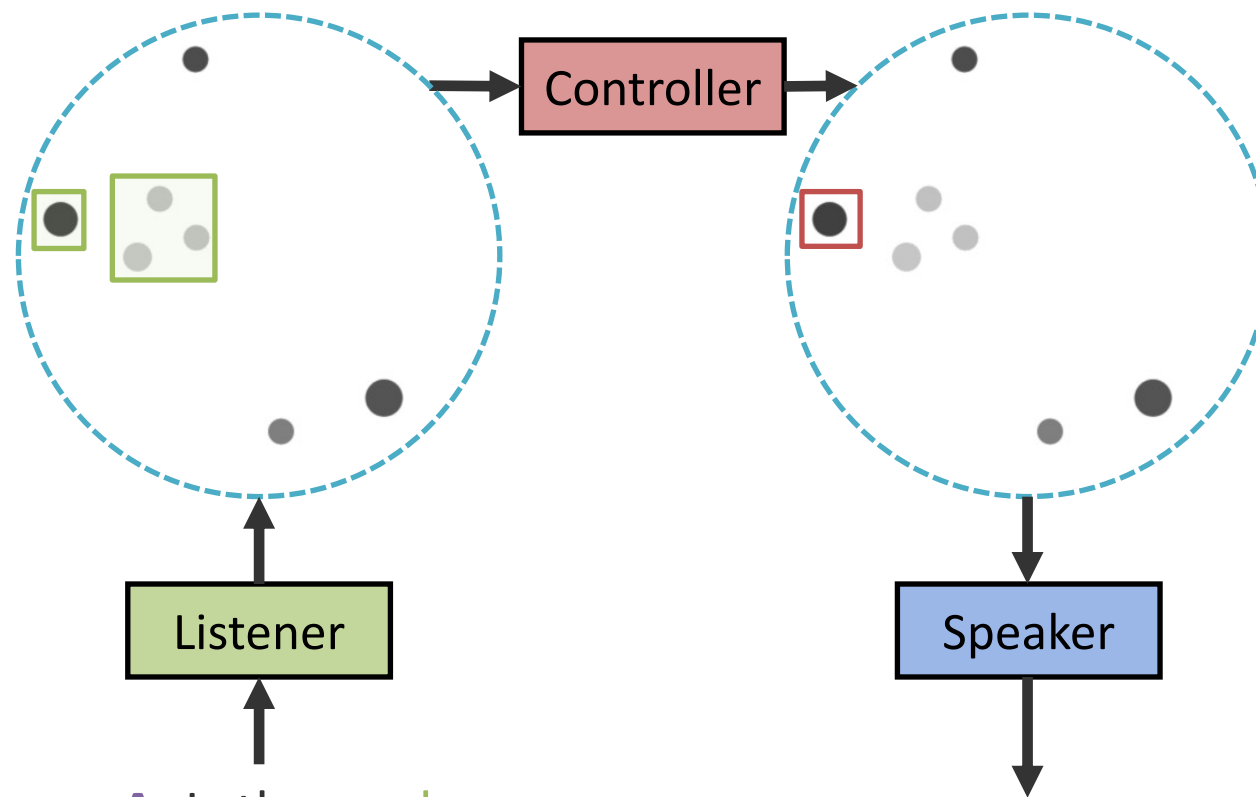


Reference-Centric Models

Reference resolution

Content selection

Surface realization



B: I don't have that. Do you have a group of three grey dots?

A: Is there a large black dot to the left of the three grey dots?

B: Yes, let's select the black one.

Other components (see paper): a structured memory, a confirmation module, and a selection module.



Combining Work on Grounded...

Reference Resolution

Listener

Schlangen et al. 2009; Liu et al. 2013,
Kennington & Schlangen 2015

Reference Generation

Speaker

Dale 1989; Dale & Reiter 1995; Jordan &
Walker 2005; Viethen et al. 2011;

Heeman 1991; Mao et al. 2016; Yu et al. 2017; Takmaz et al. 2020

Goal-Oriented Dialogue

Controller

Traum 1994; Potts 2012; Liu et al.
2013; Das et al. 2017; He et al. 2017;
Thomason et al. 2019; Haber et al. 2019;
Ilinykh et al. 2019; Udagawa et al. 2019

Pragmatics

Listener

Speaker

Golland et al. 2010; Frank & Goodman
2012; Vogel et al. 2013; Monroe et al.
2017; Khani et al. 2018; Cohn-Gordon
et al. 2018; Fried et al. 2018

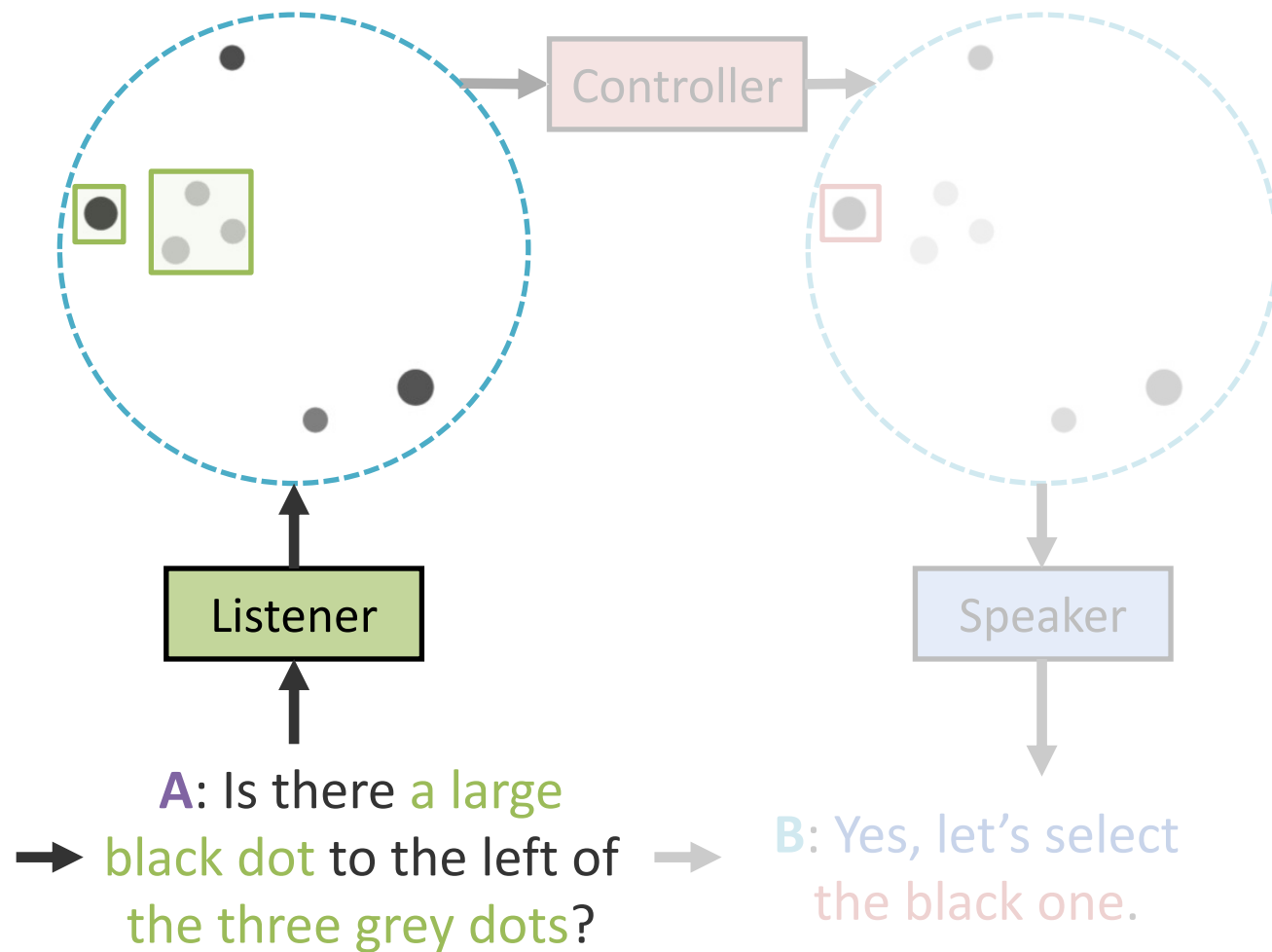


Interpretation

Reference resolution

Content selection

Surface realization

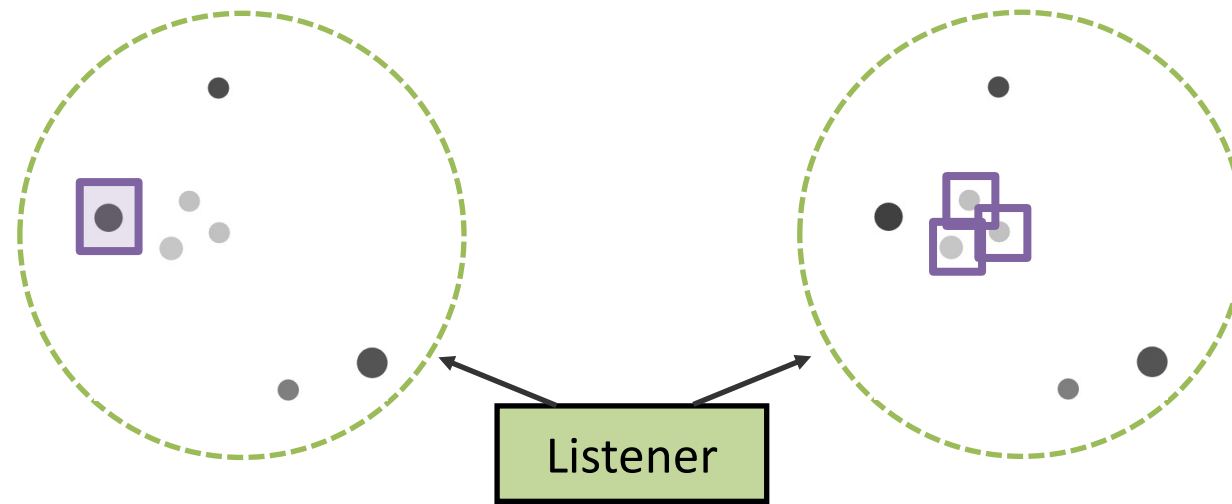




Listener Module

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

Predict individual dots in referents using graph neural net representations



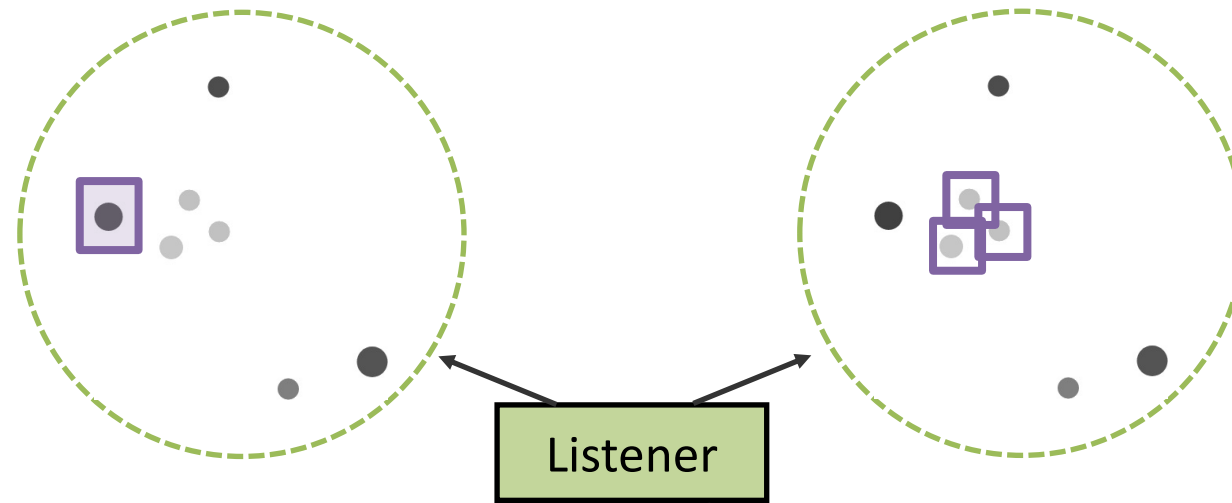
Is there a large black dot to the left of the three grey dots?



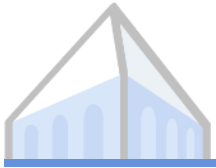
Listener Module

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

Predict **individual dots** in referents using graph neural net representations



Is there a large black dot to the left of the three grey dots?



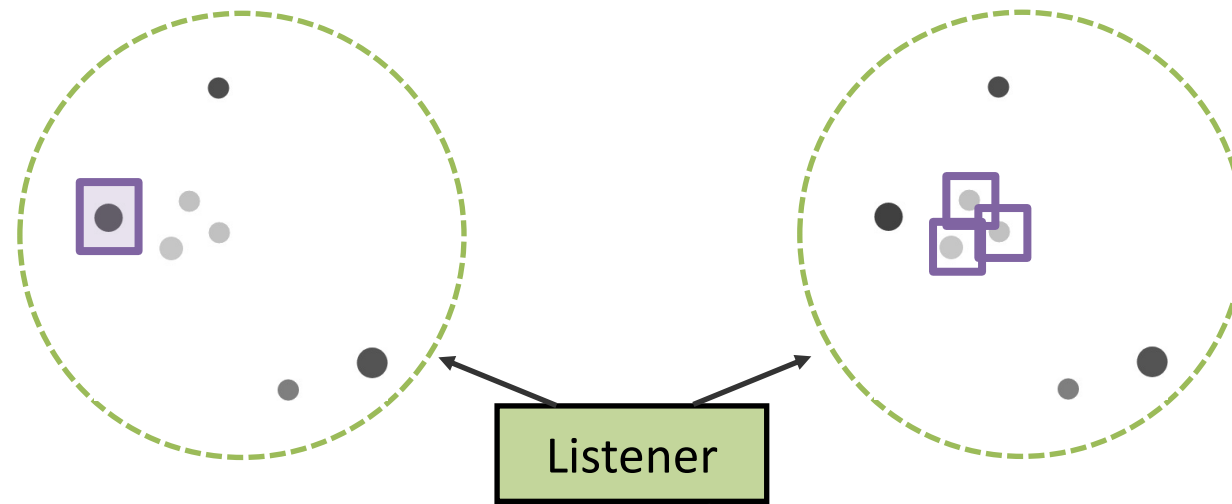
Listener Module

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

Predict **individual dots** in referents using graph neural net representations

+ Structured Neural CRF

Add neural potentials over groups and relations



Is there a large black dot to the left of the three grey dots?



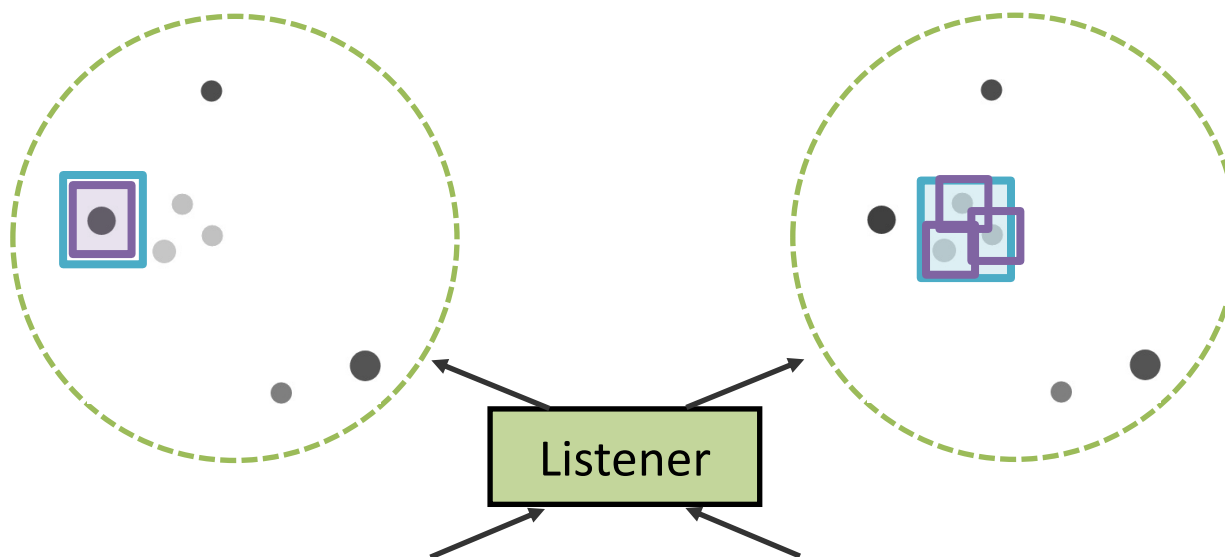
Listener Module

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

Predict **individual dots** in referents using graph neural net representations

+ Structured Neural CRF

Add neural potentials over **groups** and relations



Is there a large black dot to the left of the three grey dots?



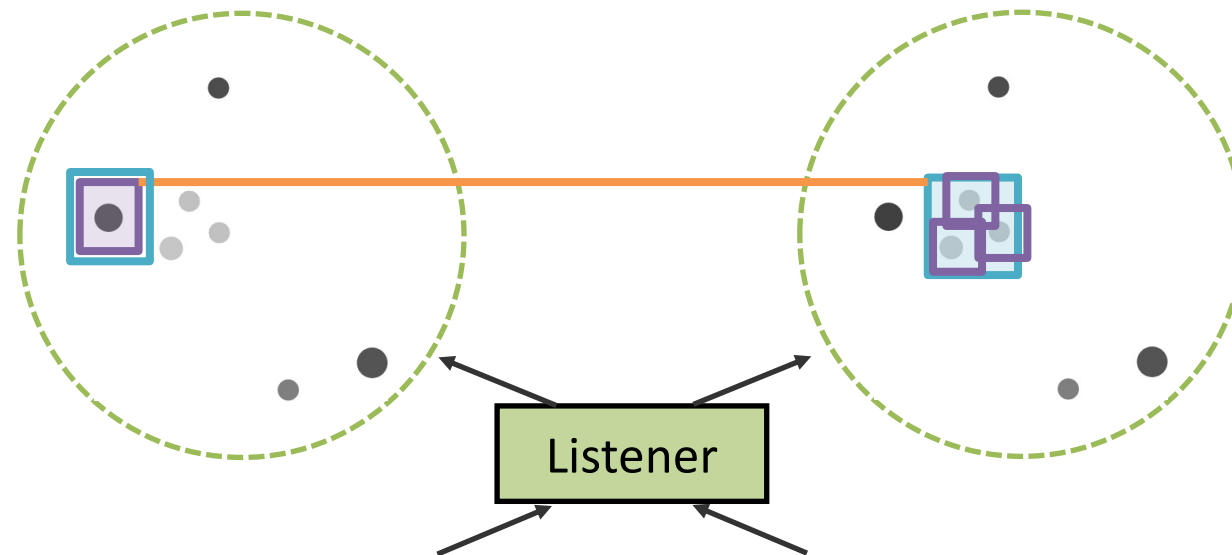
Listener Module

RelationNet [Santoro et al. 2017, Udagawa and Aizawa 2020]

Predict **individual dots** in referents using graph neural net representations

+ Structured Neural CRF

Add neural potentials over **groups** and **relations**



Is there a large black dot to the left of the three grey dots?

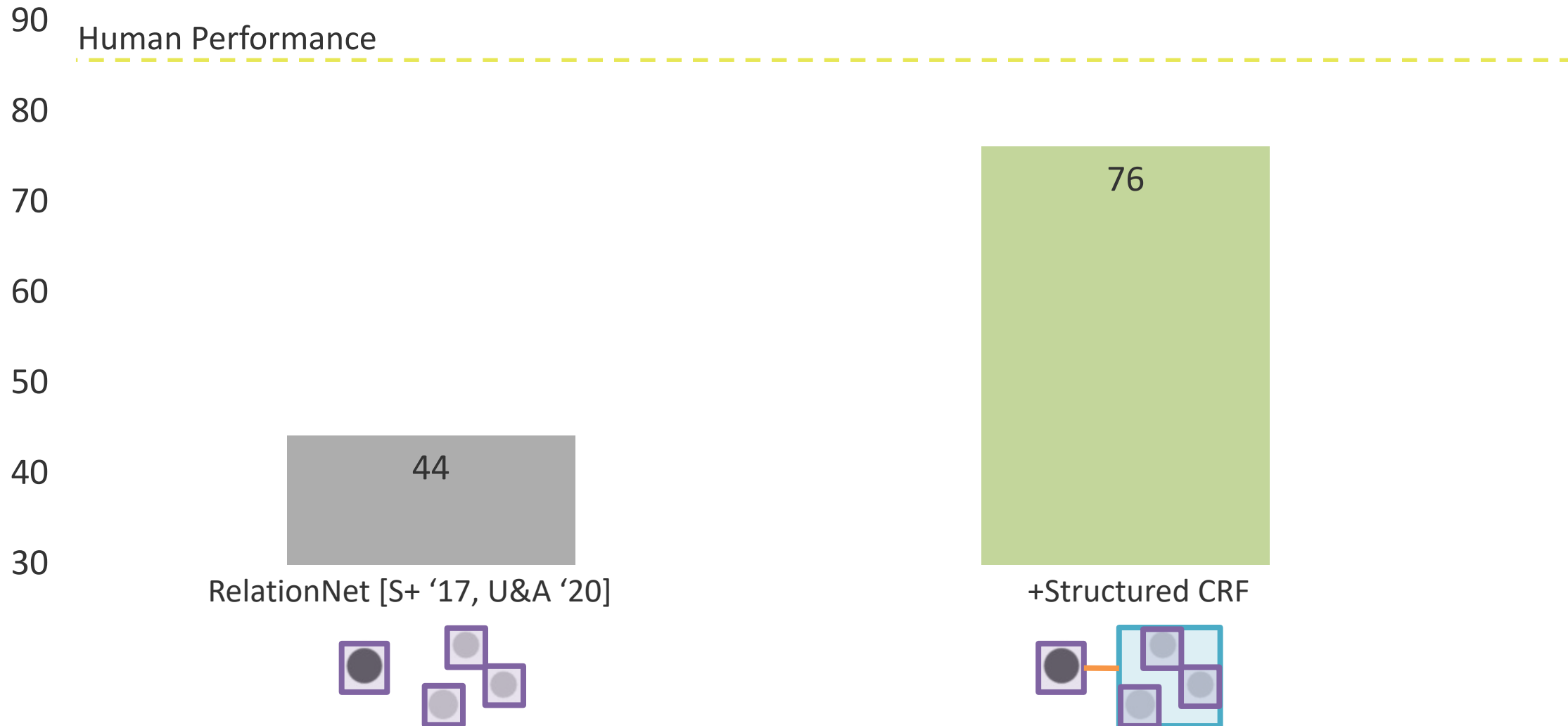
Structured CRF potentials condition on encoded dot attributes and an LSTM representation of the text.

A linear chain dynamic program makes training and inference efficient.



Listener Module Evaluation

Reference Resolution Exact Match





Generation

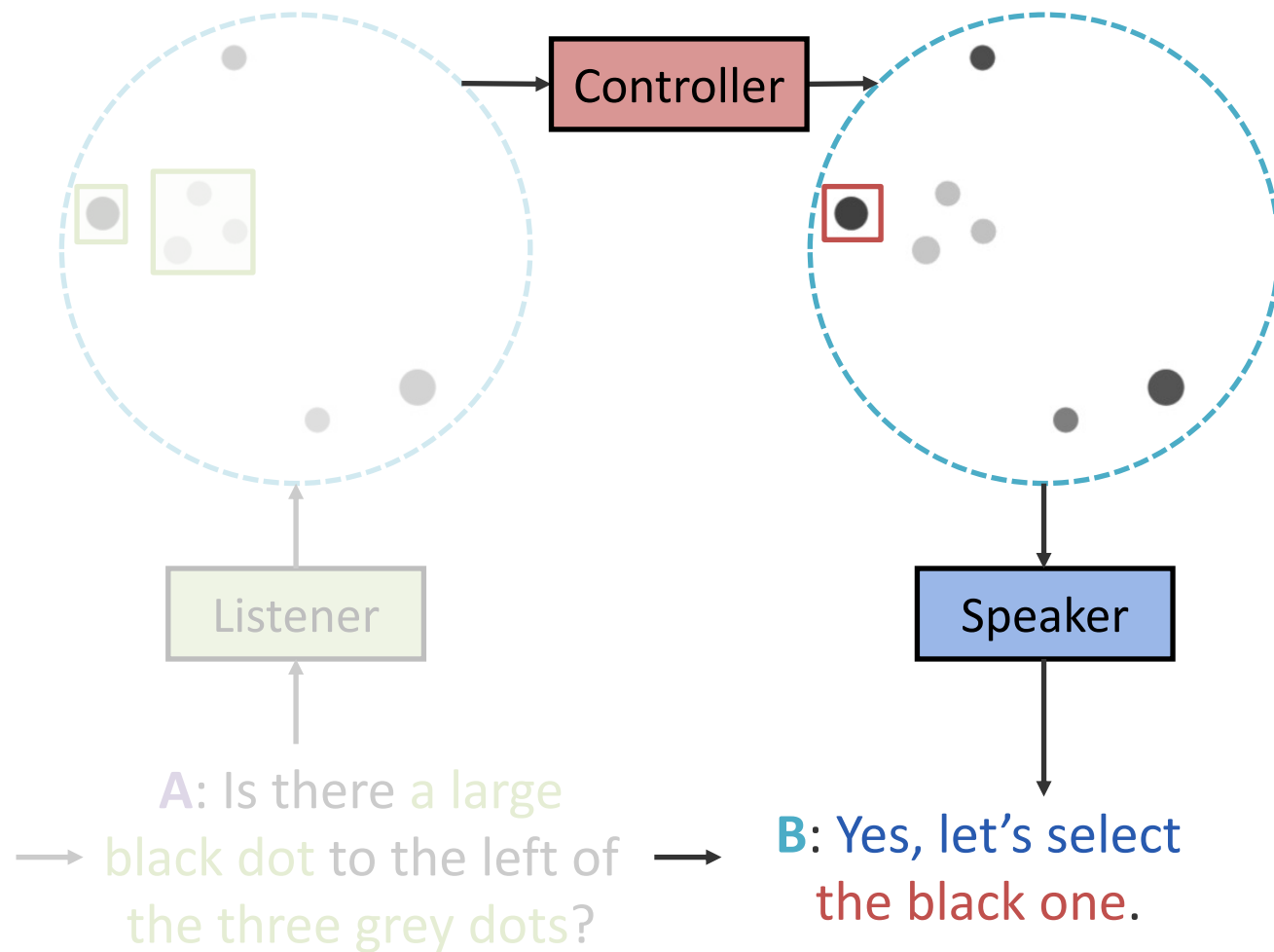
Reference resolution

Content selection

Surface realization

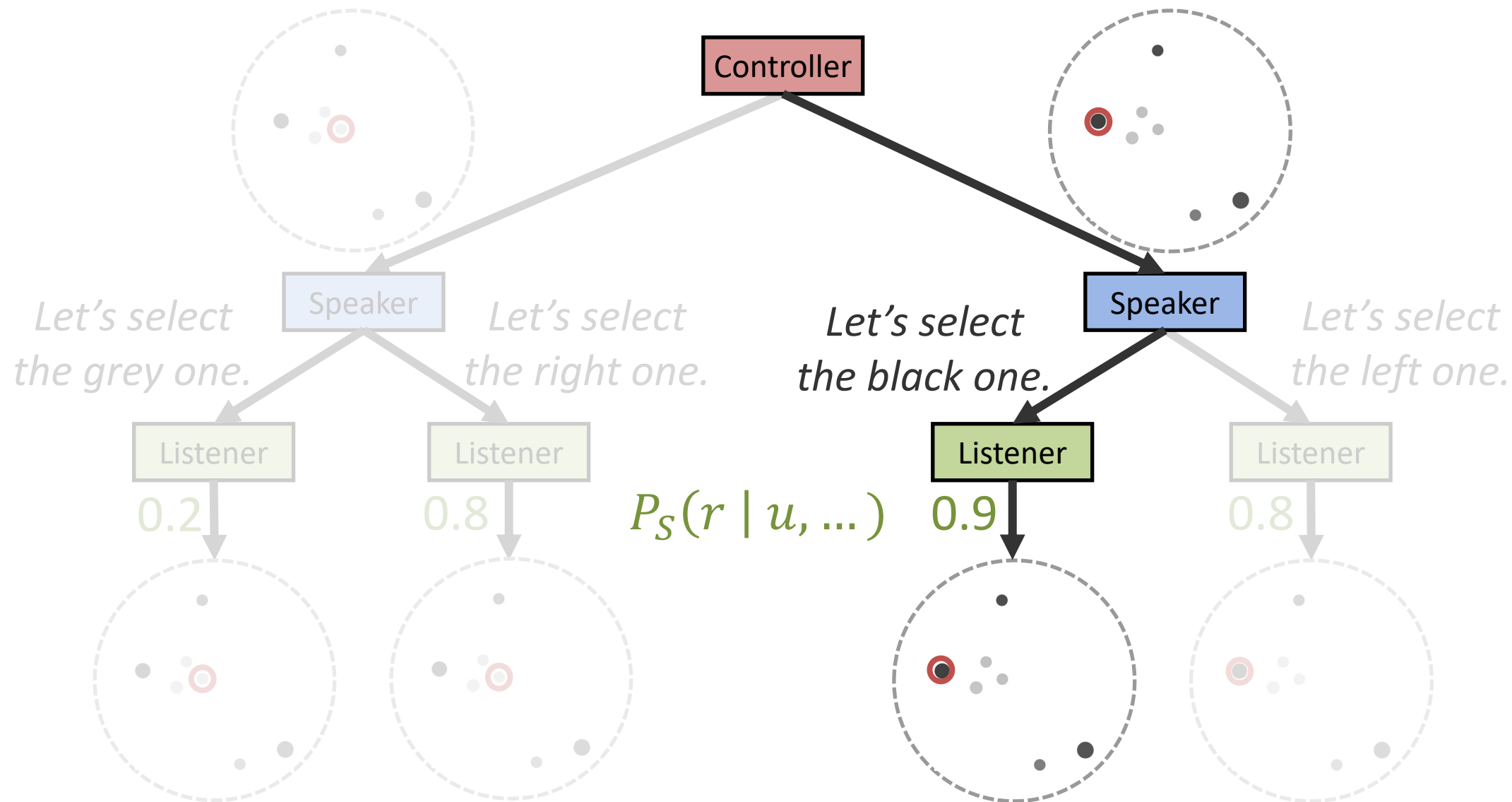
+ Pragmatic informativity

Listener





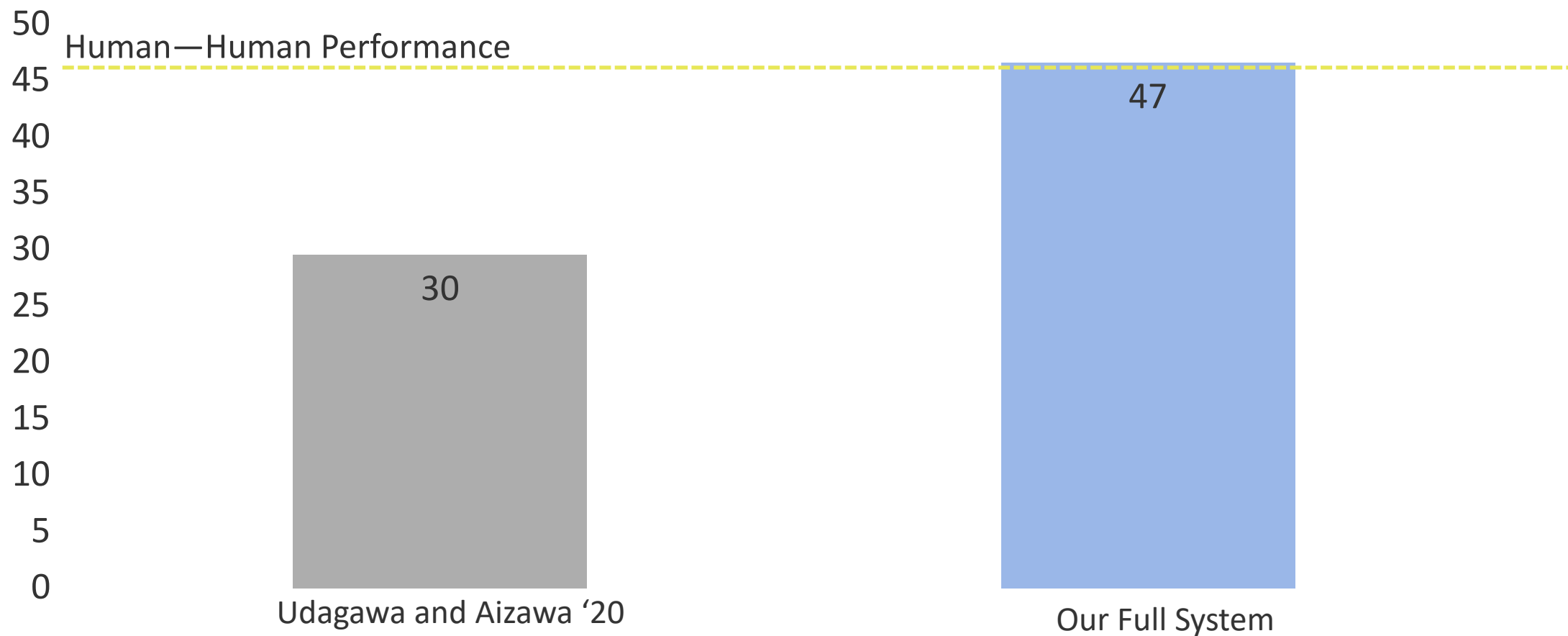
Pragmatic Generation





Full System Evaluation

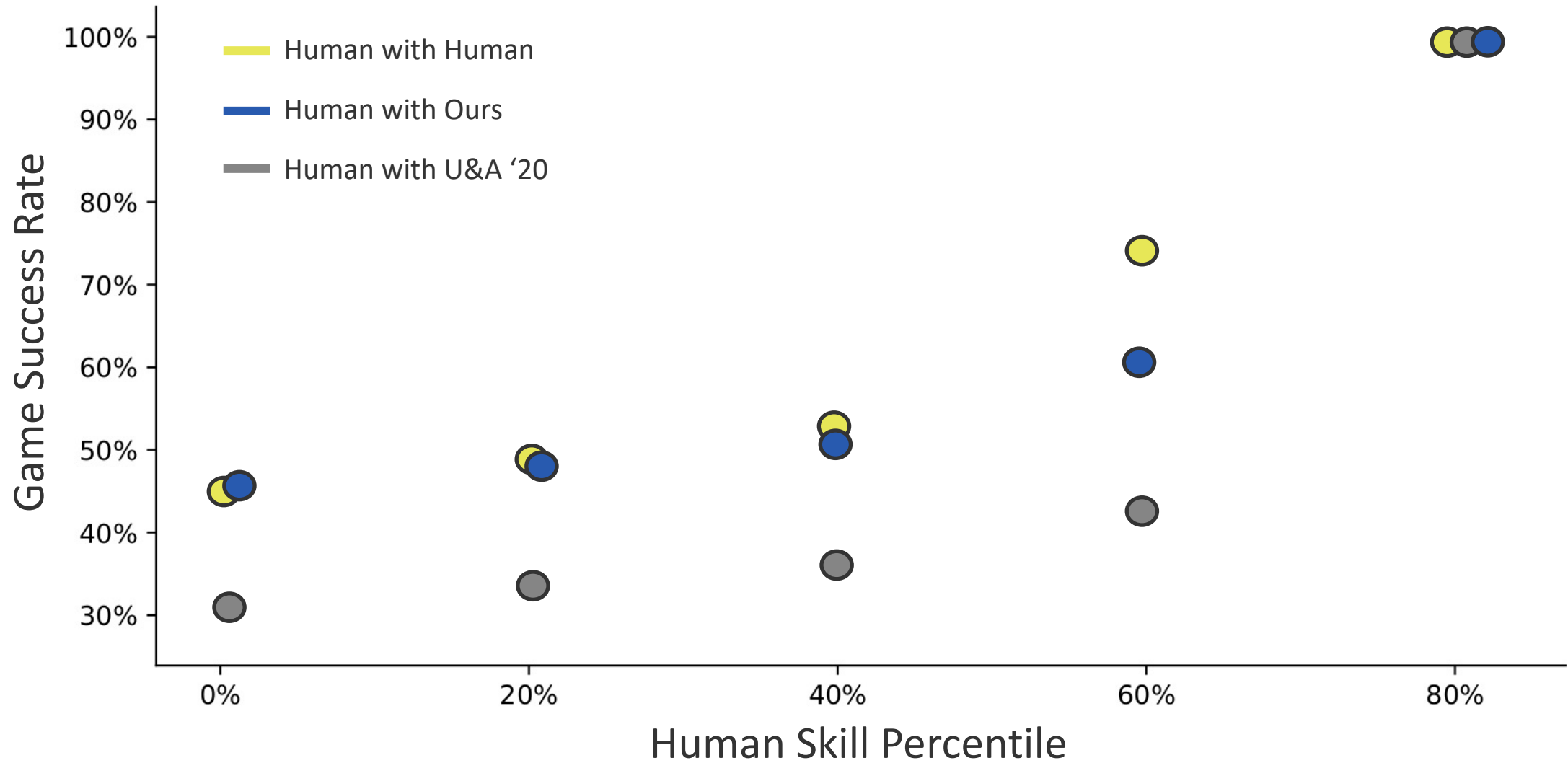
Game Success in Pairings with Humans



See paper for ablations showing that structure and pragmatics help substantially.

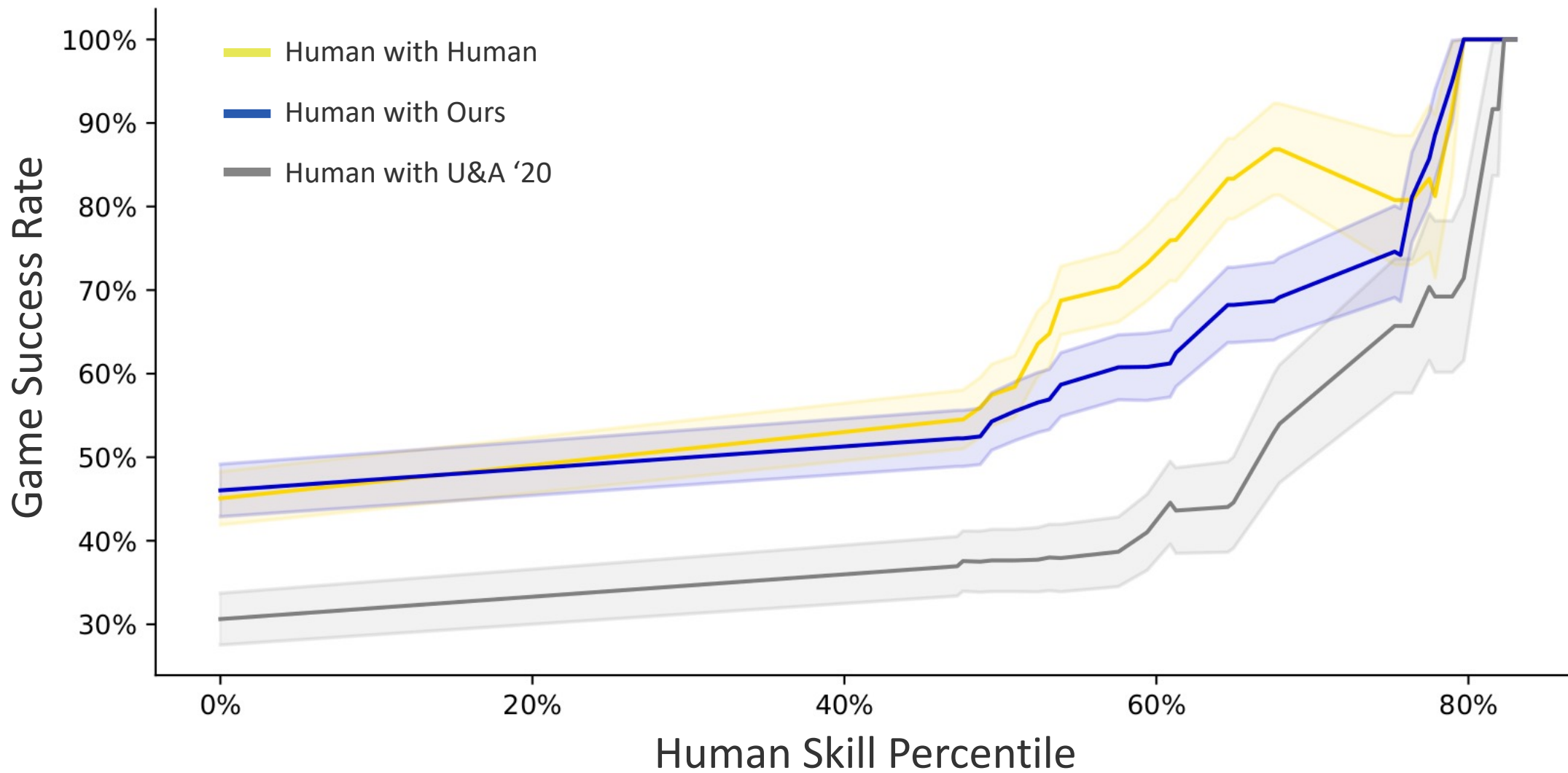


Success by Human Skill Level



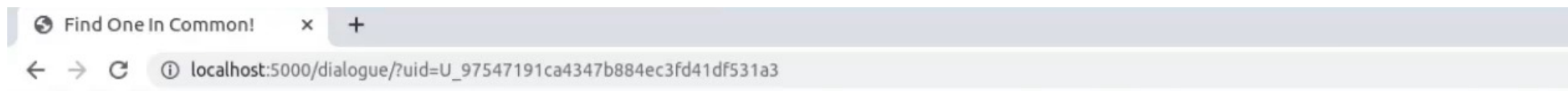


Success by Human Skill Level





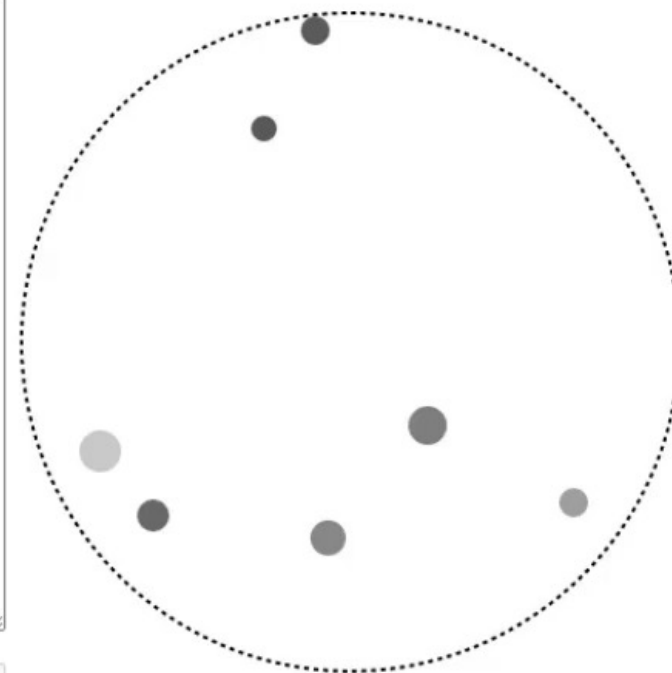
Demo



Time Remaining: 6:00

[02/12/21 08:57:44] <You entered the room.>
[02/12/21 08:57:46] <Your partner has joined the room.>

Your view

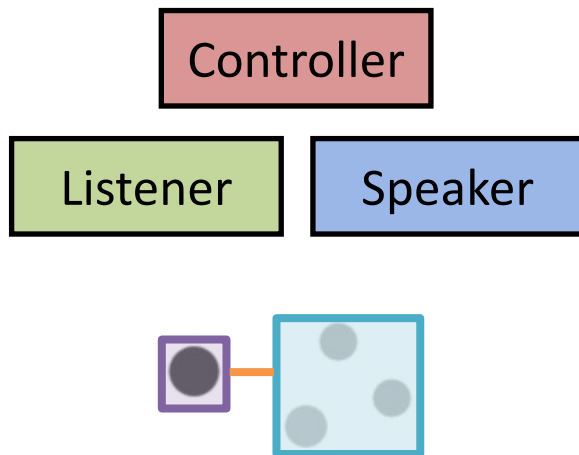


Waiting on your partner to take a turn...

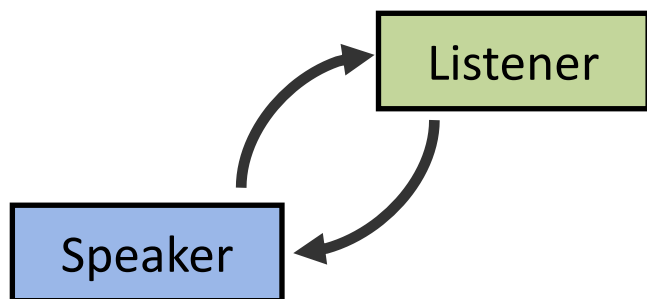




Takeaways



Structure still benefits grounded neural dialogue models.



Pragmatic modeling makes grounded dialogue more effective.

Thank you!



Poster: Sunday Nov 7th,
7-9pm AST / 4-6pm Pacific

github.com/dpfried/onecommon