Reference-Centric Models for Grounded Collaborative Dialogue
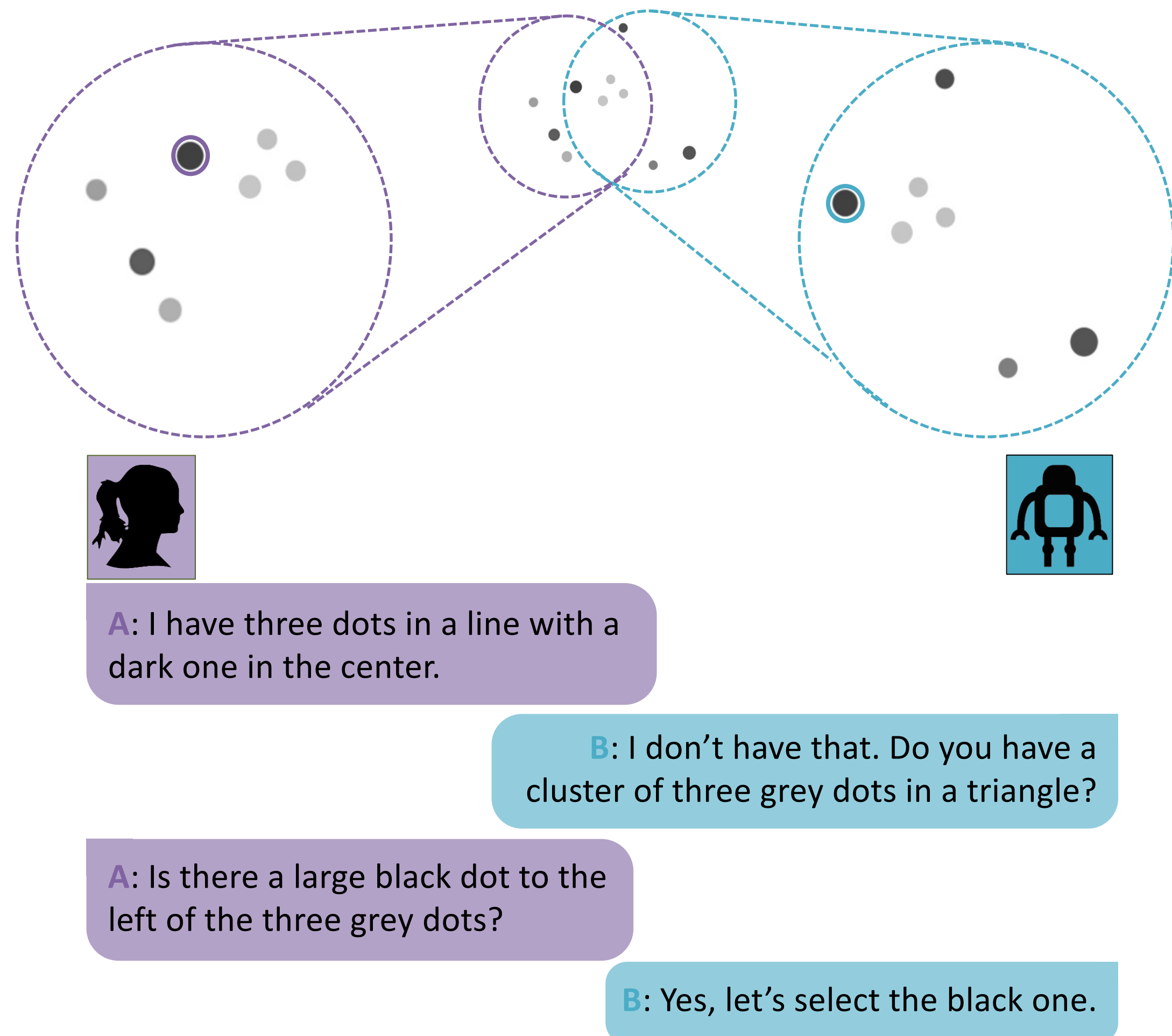
Daniel Fried[1], Justin Chiu[2], and Dan Klein[1]
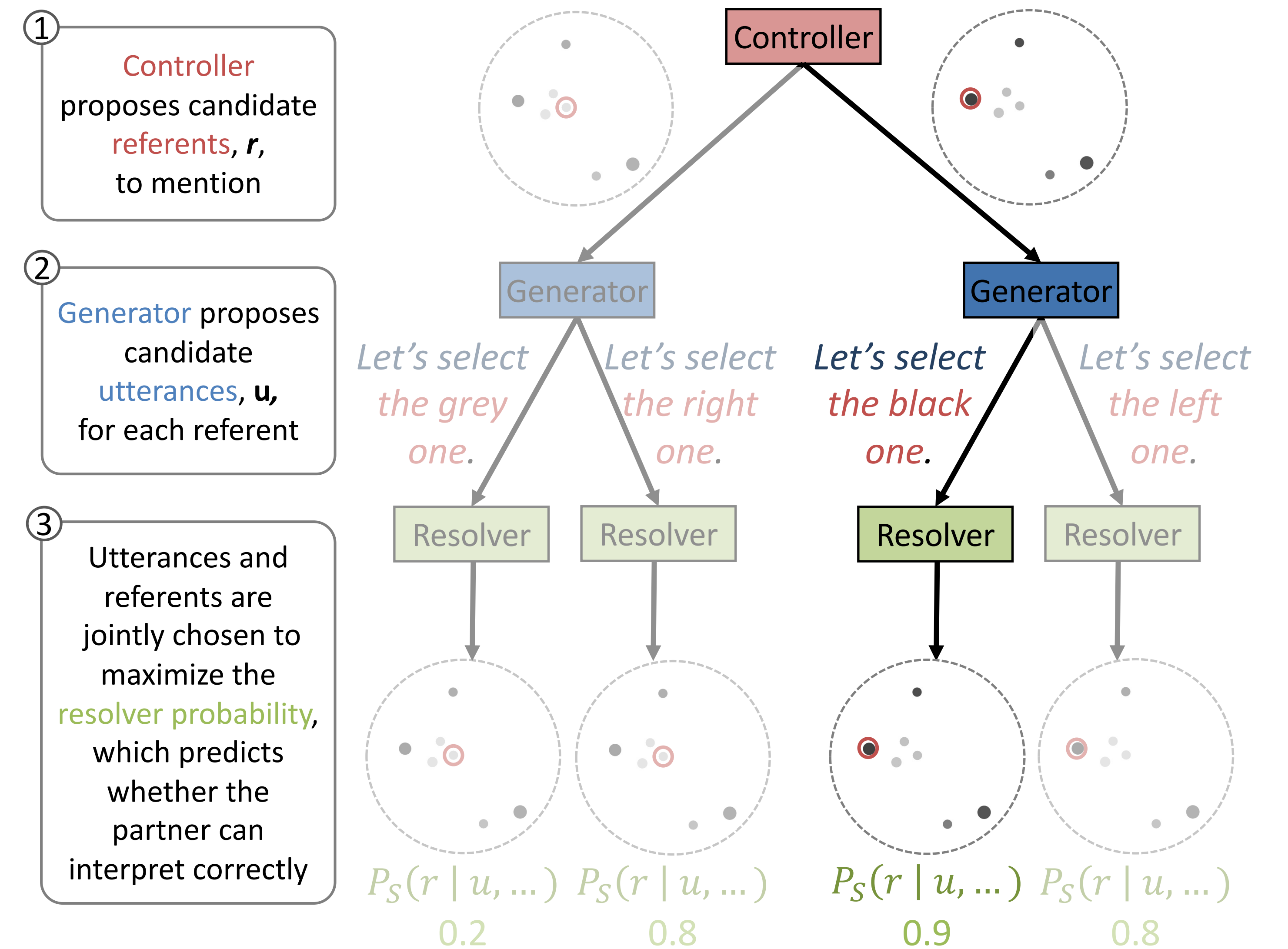
[1]UC Berkeley   [2]Cornell University

Grounded Collaborative Dialogue

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 cluster of three grey dots in a triangle?

A: Is there a large black dot to the left of the three grey dots?
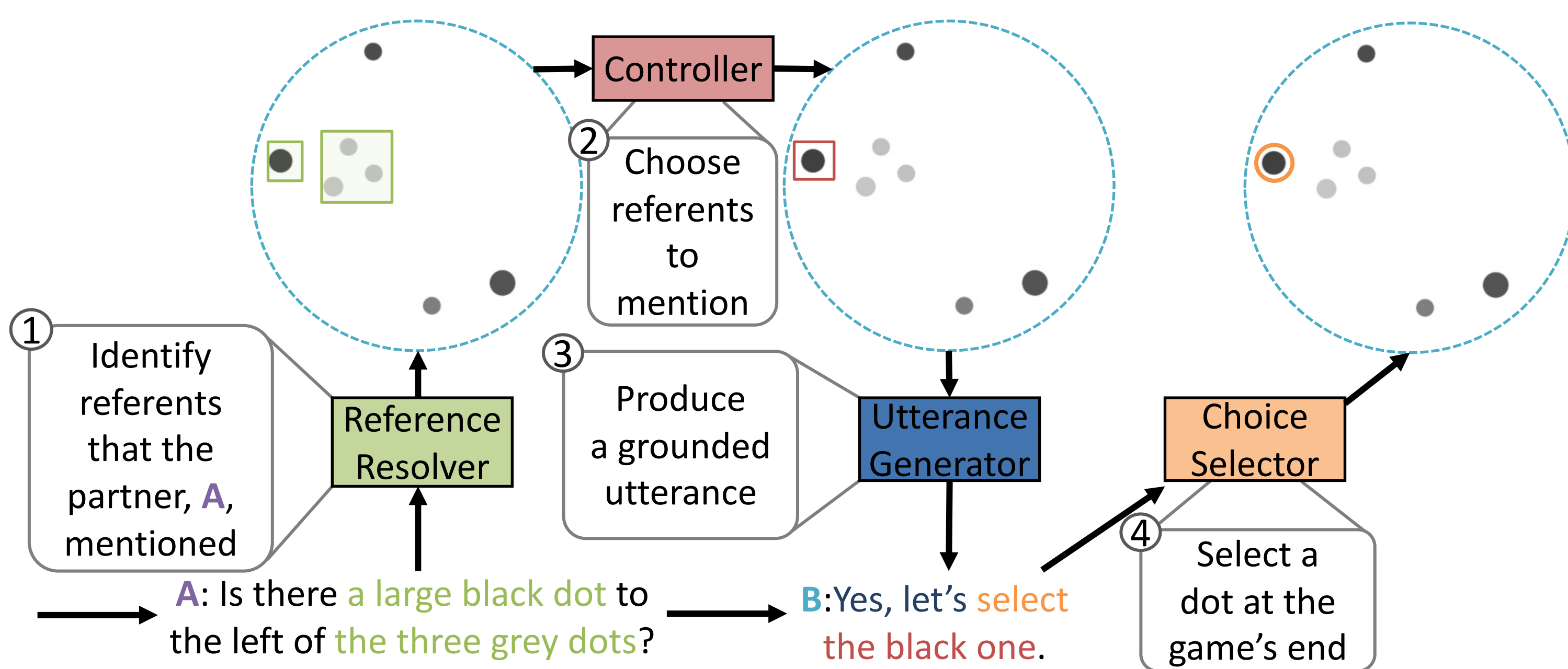
B: Yes, let's select the black one.

We focus on OneCommon [Udagawa & Aizawa 2019, 2020], a symmetric, **partially-observable collaborative reference** game. Represents a class of games [e.g. Haber et al. 2019] that require **strategic generation under uncertainty**.

Generation as a Pragmatic Game

① Controller proposes candidate referents, **r**, to mention

② Generator proposes candidate utterances, **u**, for each referent

③ Utterances and referents are jointly chosen to maximize the resolver probability, which predicts whether the partner can interpret correctly

Let's select the grey one.   Let's select the right one.   Let's select the black one.   Let's select the left one.

$P_S(r \mid u, \ldots)$ $P_S(r \mid u, \ldots)$ $P_S(r \mid u, \ldots)$ $P_S(r \mid u, \ldots)$
0.2   0.8   0.9   0.8

Modular Neural Architecture

① Identify referents that the partner, **A**, mentioned

② Choose referents to mention

③ Produce a grounded utterance

④ Select a dot at the game's end

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

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

We implement the referent prediction modules (Resolver, Controller, and Selector) using structured neural CRFs. The Generator is a grounded sequence-to-sequence module. All modules are trained on human—human dialogues (~5000 games) with annotated referents.

Evaluation
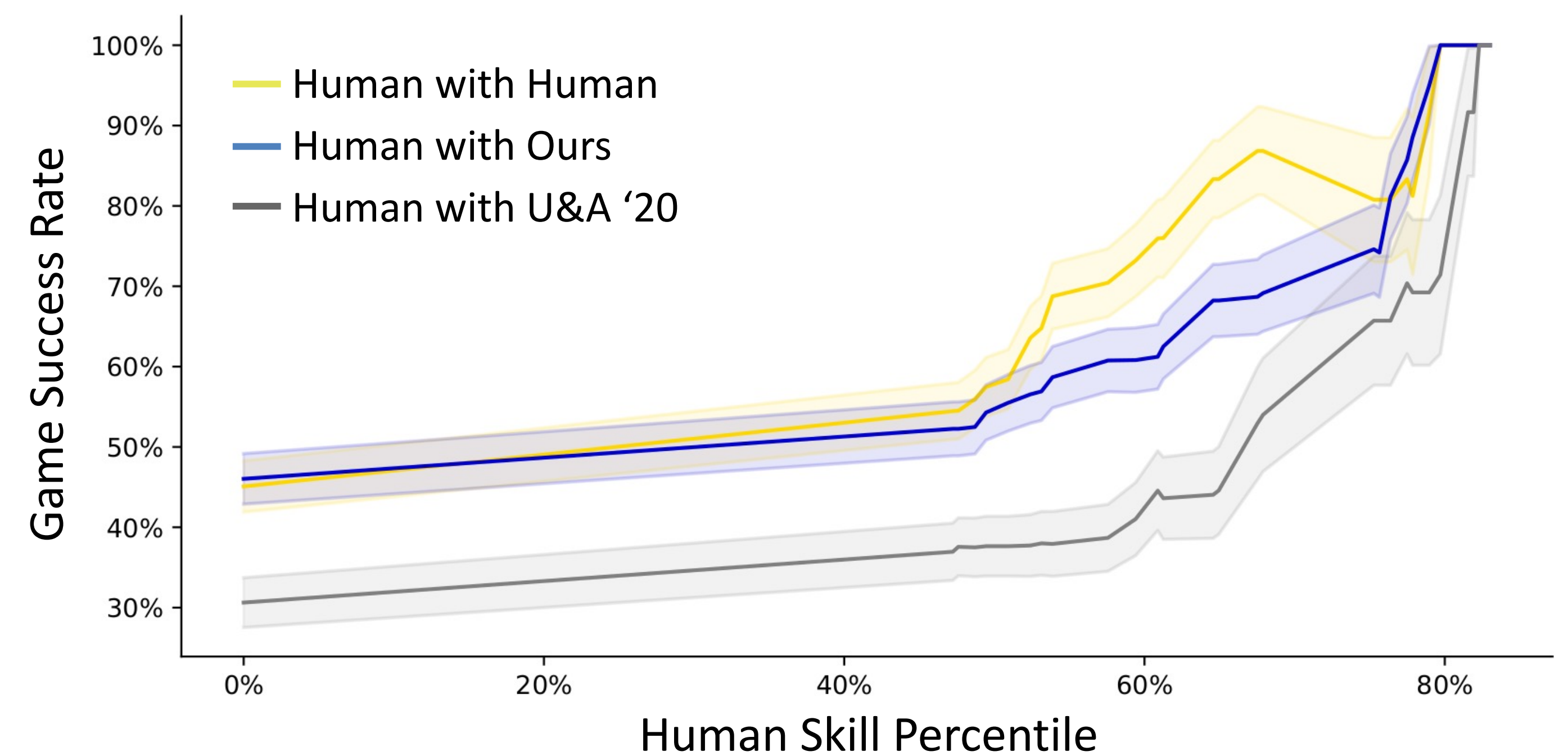
We compare our system to the past state of the art, an end-to-end neural dialogue model [U&A'20, Udagawa & Aizawa 2020] by pairing with human partners on Mechanical Turk. Our system obtains substantially higher success rates (y-axis) across humans of all skill levels (x-axis, partner success percentile).