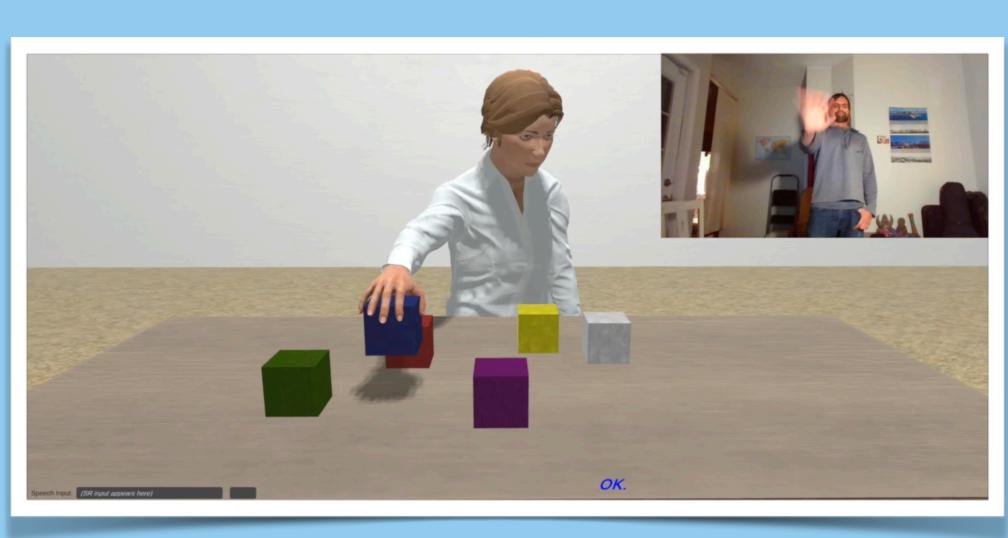
Diana's World: A Situated Multimodal Interactive Agent

Nikhil Krishnaswamy, Pradyumna Narayana, Rahul Bangar, Kyeongmin Rim, Dhruva Patil, David McNeely-White, Jaime Ruiz, Bruce Draper, Ross Beveridge, and James Pustejovsky



Real-time composition of language and gesture



Asynchronous and interruptible



Situationally aware



Object and event affordances



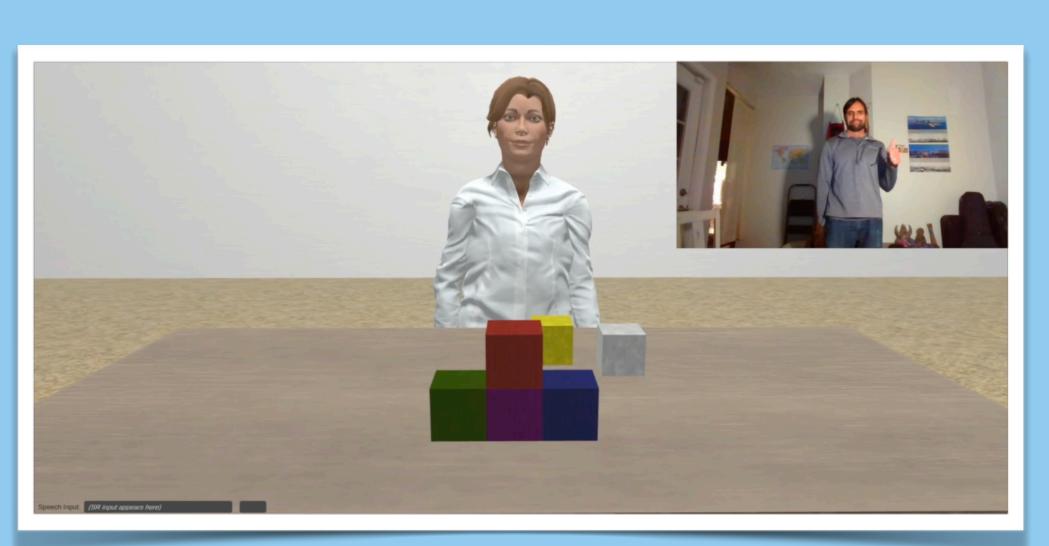


Iterative and continuous instructions

Dynamic contextual model

Diana is a multimodal interactive agent that is *co-situated* with a human user in a mixedreality environment. She can see both her virtual world and you in the real world. She interprets multi-channel inputs—including language, gesture, affect, and emotion—in real time, so she can play a proactive role in interactive collaborative tasks. Diana is situationally and environmentally aware, and participates in interactions based on a dynamically computed situational common ground created between her and her interlocutor.

<u>nkrishna@brandeis.edu</u> ross.beveridge@colostate.edu • jamesp@brandeis.edu



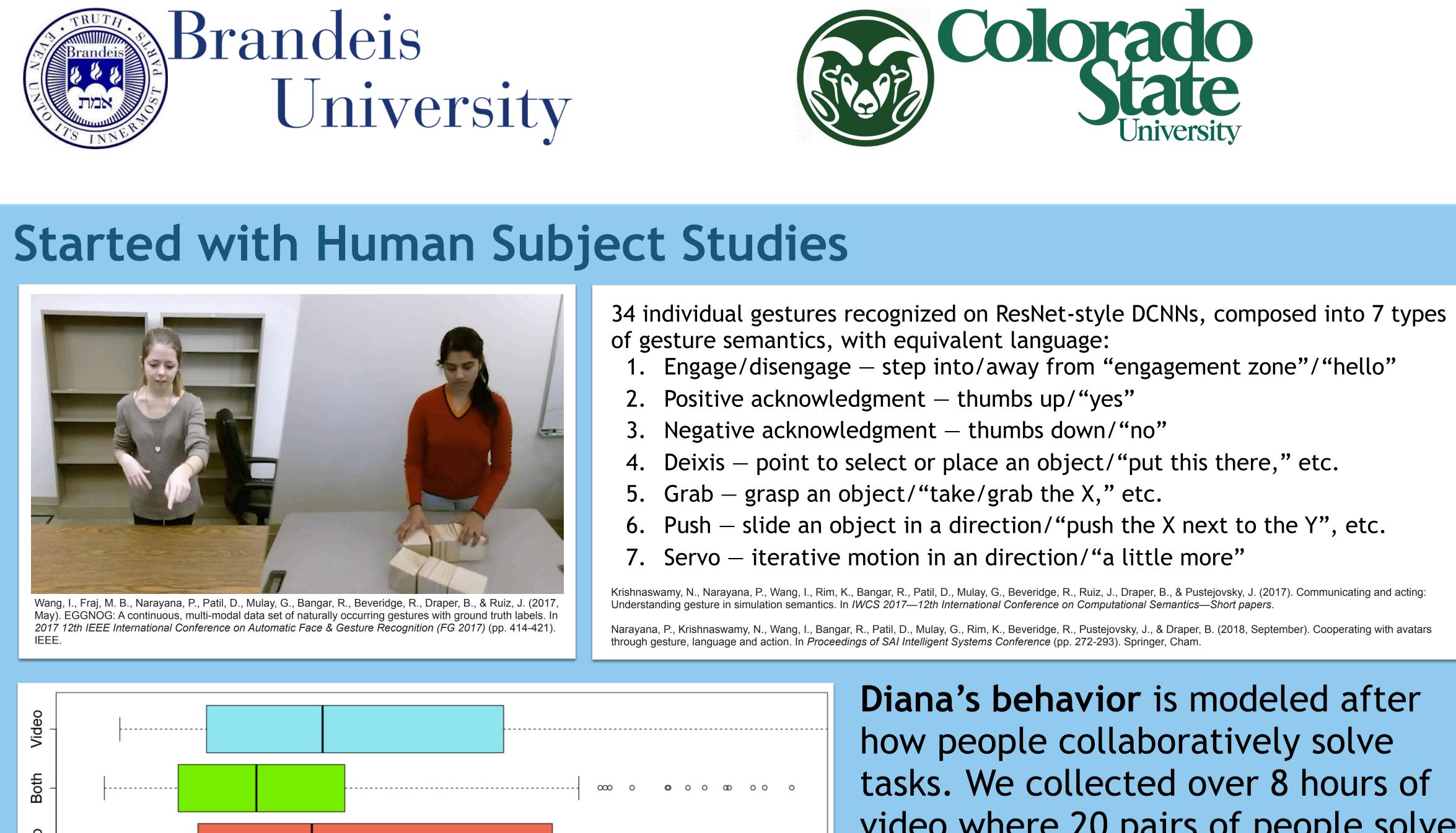
Affective and collaborative



One-shot learning



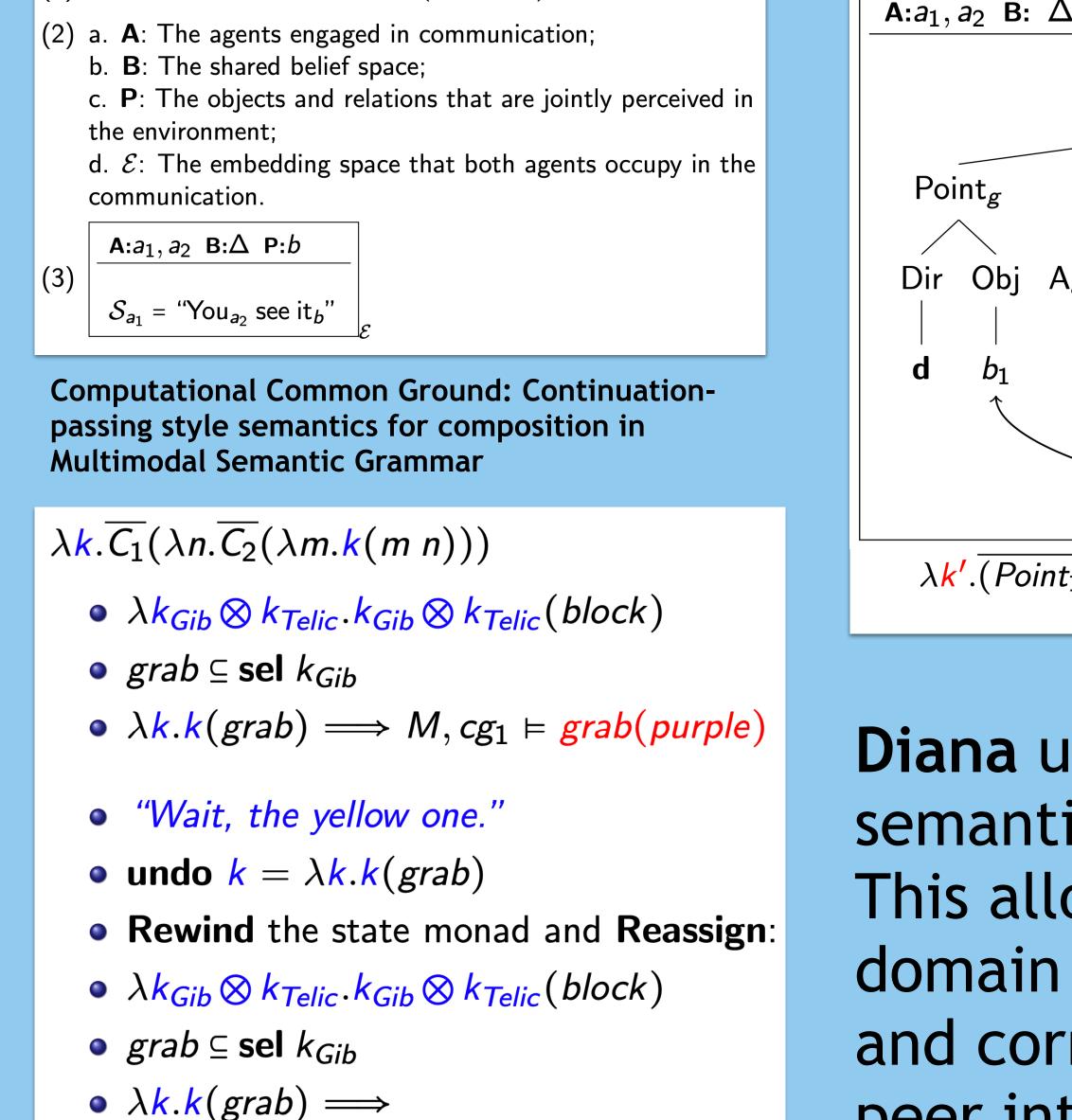
Multiple domains



Trial Time in Seconds Truncated at 200

Users complete tasks faster using both audio and visual channels

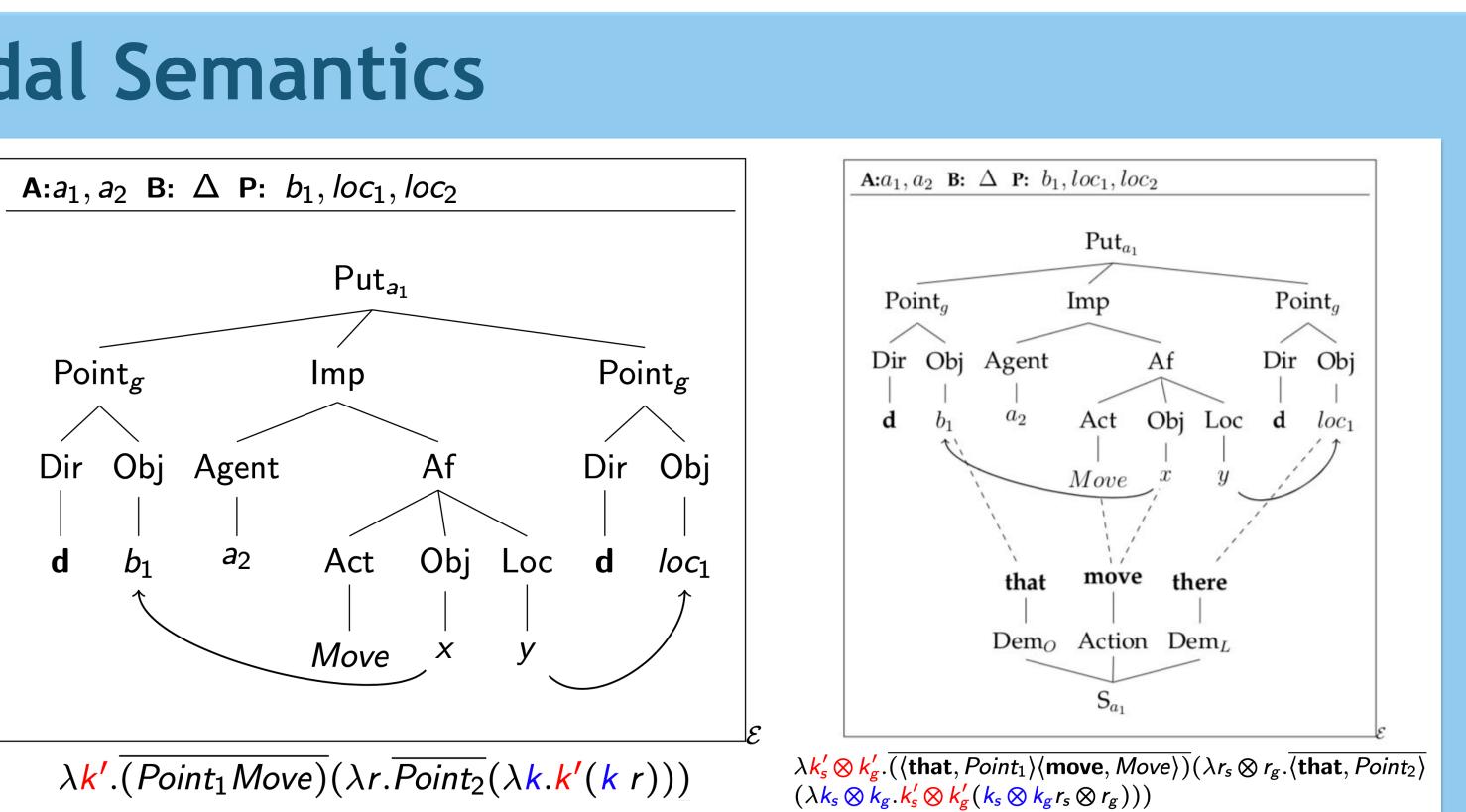
Depends Upon Multimodal Semantics



• $M, cg_1 \vDash grab(yellow)$

(1) State Monad: $M\alpha = State \rightarrow (\alpha \times State)$

Continuation semantics of correction



Unimodal and multimodal common ground structures

Diana uses a continuation-passing style semantics to aggregate and interpret inputs. This allows for robust interactions within the domain with the ability for the user to interrupt and correct actions, as they would in a peer-topeer interaction with another person.

UNIVERSITY of FLORIDA

video where 20 pairs of people solve 16 tasks each. This elicitation study deeply informs our design of Diana's use of gestures, language and vision.

