

Exploring Causal Overhypotheses in Active Learning

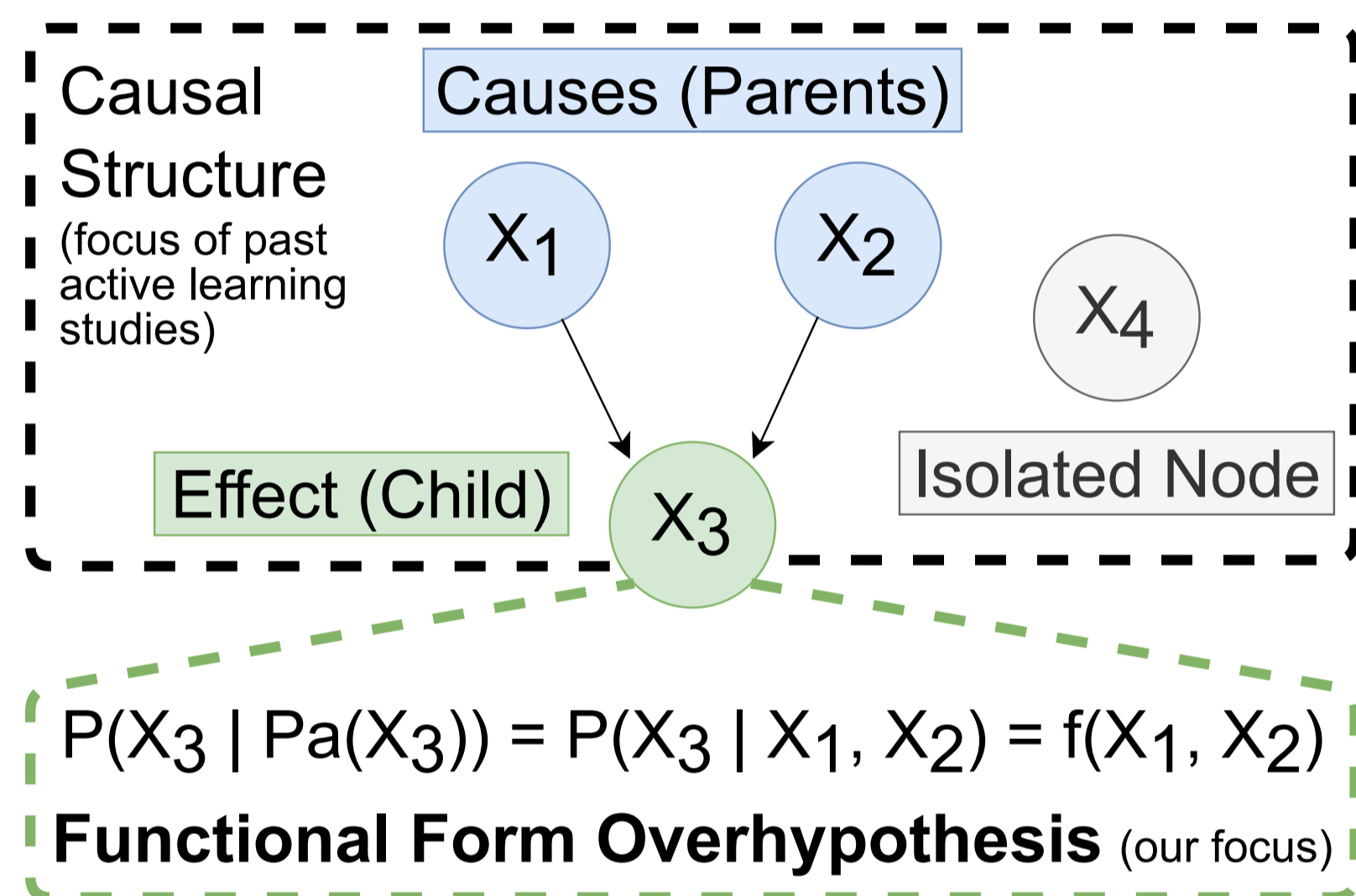
Chentian Jiang (chentian.jiang@ed.ac.uk) Christopher G. Lucas (clucas2@inf.ed.ac.uk)

Do people *actively* learn to learn?

That is: Do our actions seek out information that shapes our causal overhypotheses? Do these overhypotheses, in turn, influence our future actions?

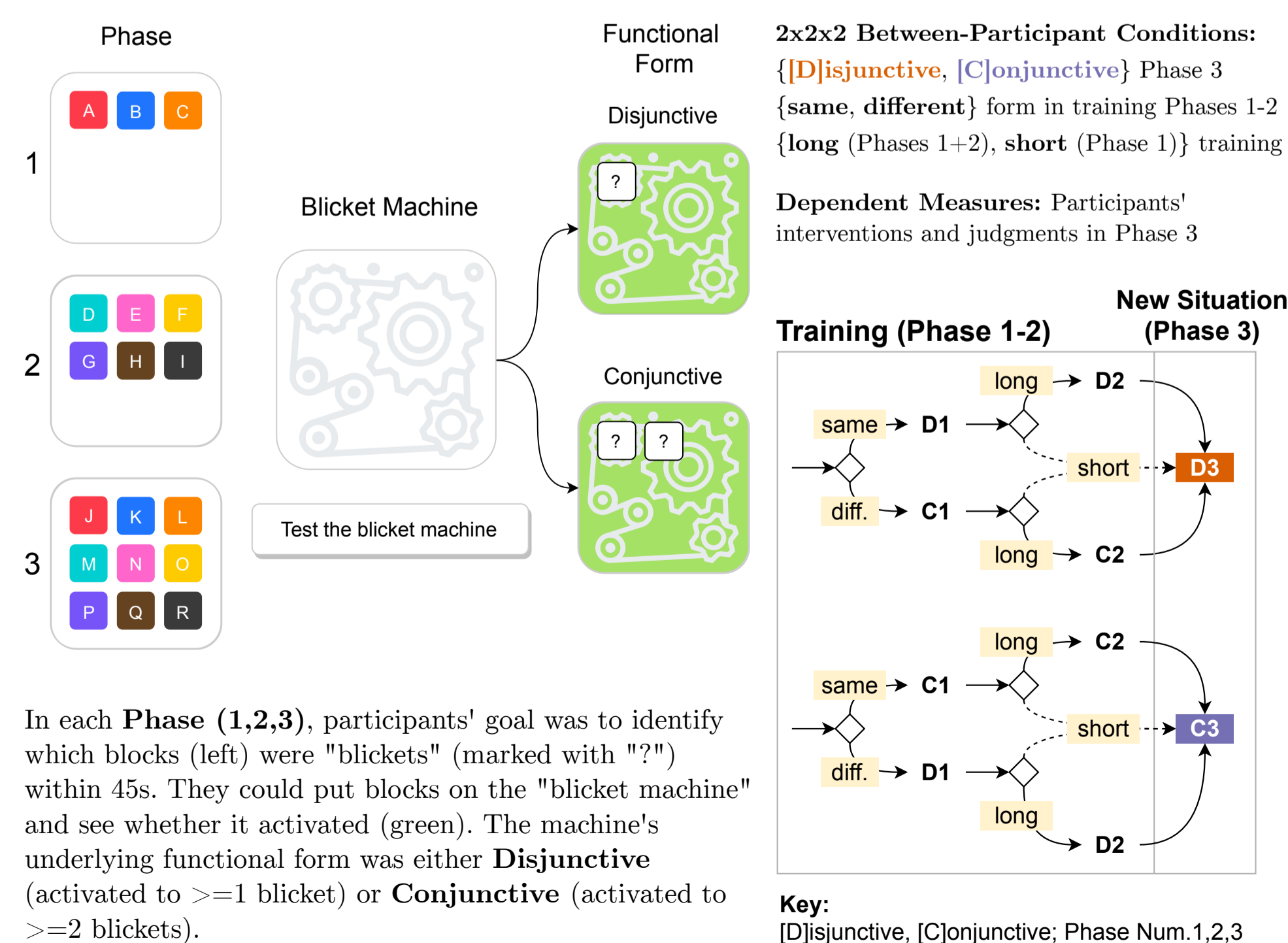
Causal overhypotheses are abstract beliefs about causal relationships that transfer across different contexts, accounting for the prior knowledge that people use to constrain their learning in future contexts¹⁻².

We focus on overhypotheses about the **functional form** of causal relationships, examining people's ability to learn functional forms through active interventions, and how learning about functional forms can influence subsequent active learning.



Active Blicket Experiment (212 MTurk Participants)

Active extension to Lucas and Griffiths's (2010) blicket experiments³.



In each **Phase (1,2,3)**, participants' goal was to identify which blocks (left) were "blickets" (marked with "?") within 45s. They could put blocks on the "blicket machine" and see whether it activated (green). The machine's underlying functional form was either **Disjunctive** (activated to ≥ 1 blicket) or **Conjunctive** (activated to ≥ 2 blickets).

Q1: Can people learn overhypotheses about the functional form in an *active* setting?

A1: Mostly yes, as measured in a new disjunctive situation.

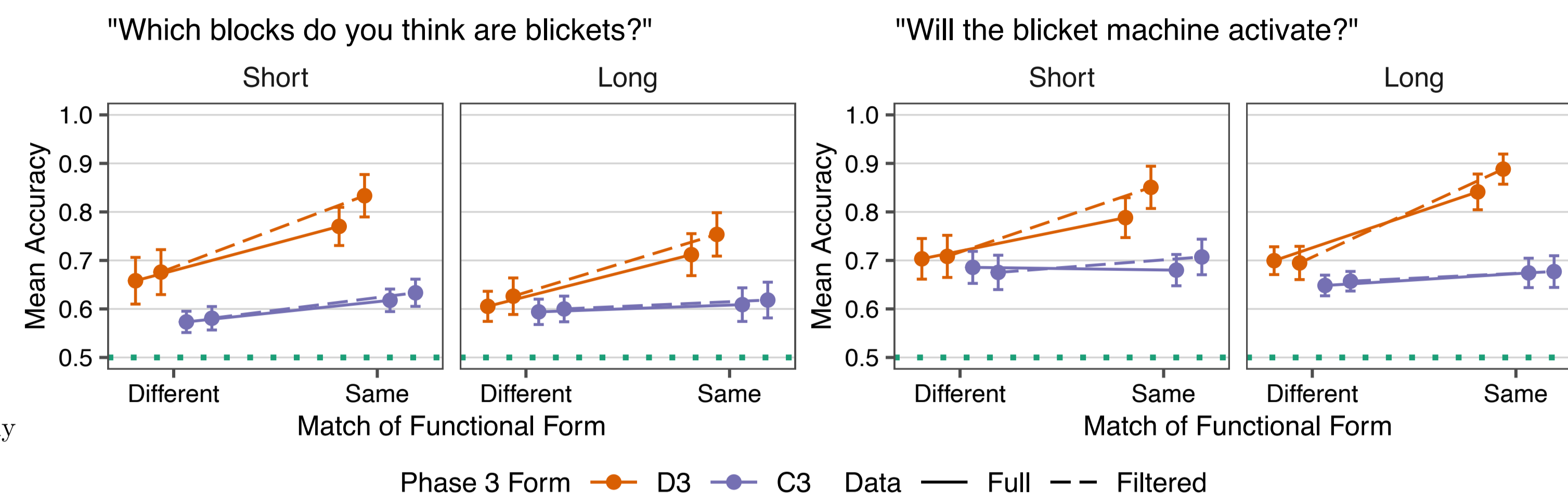
If people use a positive testing strategy⁴ that presupposes a disjunctive functional form³, their active learning may not reveal conjunctive forms. But we found this was not the case in a new disjunctive situation (D3):

Yes:

In a new disjunctive situation (D3), people's judgments were more accurate following training with the same form (disjunctive) than with a different form (conjunctive).

Unclear results due to task difficulty:

In a new conjunctive situation (C3), the effect of training with the same (conjunctive) vs. different (disjunctive) form was not significant. This was likely because C3's task was too difficult to learn in 45s, regardless of whether people learned overhypotheses.



Given a generally positive A1, we pursued a more detailed question about *how exactly* people learn overhypotheses actively:

Q2: How do people choose interventions that help them learn overhypotheses?

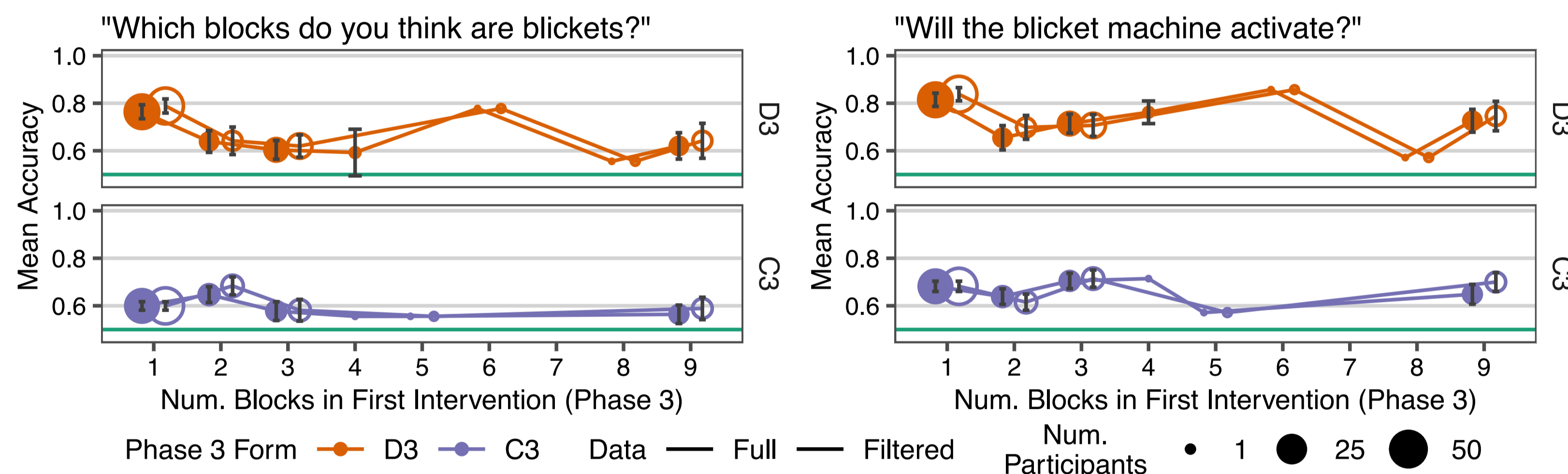
A2 (left plot): Choosing 1 object helps disj. learning; 2 helps conj. learning.

Left plot (significant):

Num. blocks in the first intervention of Phase 3 predicted (logistic model) subsequent blicket judgments. Specifically, the plot shows a singleton block indicated better **disjunctive** blicket judgments, while a pair of blocks indicated better **conjunctive** blicket judgments.

Right plot (not significant):

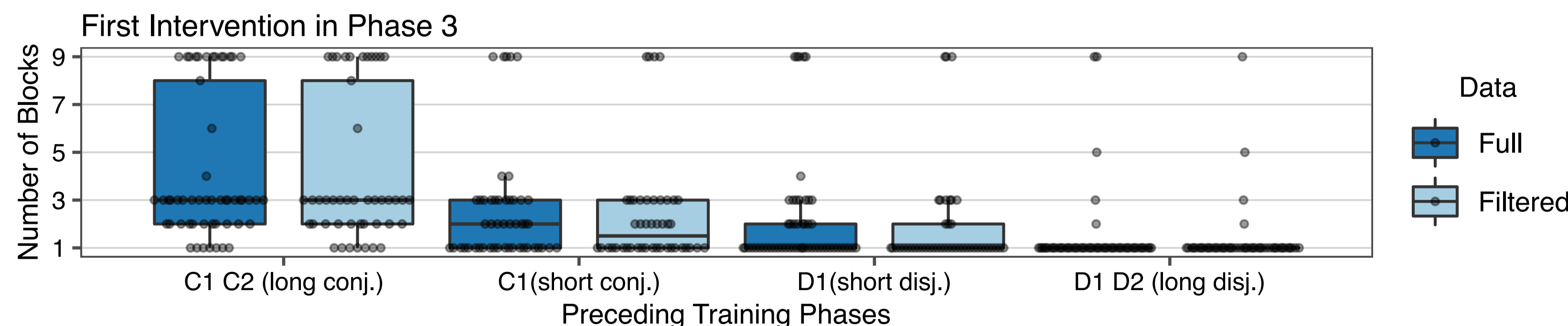
Num. blocks in the first intervention of Phase 3 was not a significant predictor (logistic model) for activation judgments.



Q3: Complementary to Q2, how do learned overhypotheses affect interventions?

A3: Conjunctive training predicts choosing more objects than disjunctive training.

The training form (disjunctive or conjunctive) and training length (long or short) predicted (linear model) the num. blocks in the first intervention of Phase 3 (new situation). The plot shows that num. blocks decreases from long conjunctive to long disjunctive training.



Ongoing work with Bayesian modeling: Do people use previously learned overhypotheses to choose the most informative (as measured by information gain) interventions?

References

- Lucas, C. G., Gopnik, A. & Griffiths, T. L. Developmental differences in learning the forms of causal relationships. in Proceedings of the Annual Meeting of the Cognitive Science Society vol. 32 (2010).
- Kemp, C., Perfors, A. & Tenenbaum, J. B. Learning overhypotheses with hierarchical Bayesian models. Developmental Science 10, 307-321 (2007).
- Lucas, C. G. & Griffiths, T. L. Learning the Form of Causal Relationships Using Hierarchical Bayesian Models. Cognitive Science 34, 113-147 (2010).
- Coenen, A., Rehder, B. & Gureckis, T. M. Strategies to intervene on causal systems are adaptively selected. Cognitive Psychology 79, 102-133 (2015).