Understanding "why"

Causation, Counterfactuals, and Imagination

@tobigerstenberg

http://cicl.stanford.edu

Why I'm here today

Tobias Gerstenberg @tobigerstenberg · Mar 6

Looking forward to it!

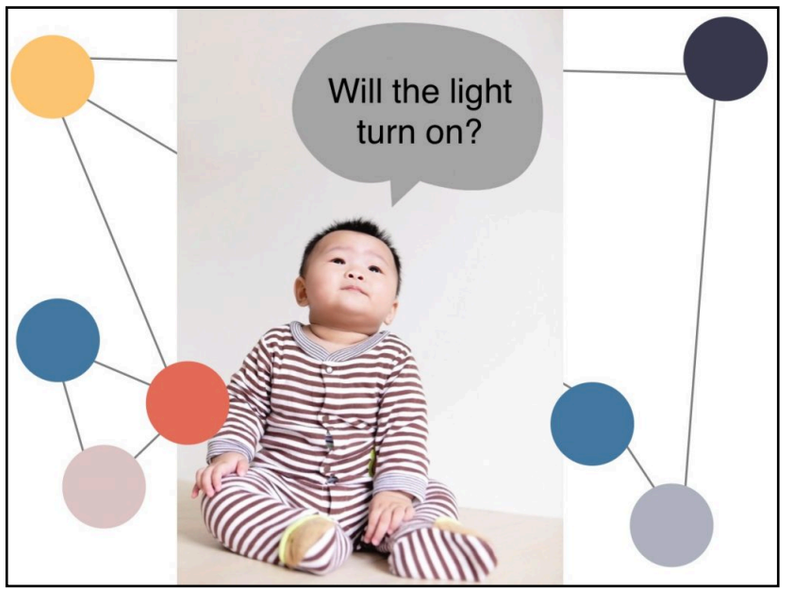
> **Elias Bareinboim** @eliasbareinboim
> The WHY-19 will be happening from Mar/25-27 @ Stanford. The theme this year is "Beyond Curve Fitting: Causation, Counterfactuals, Imagination-based AI". We have great speakers, including J. Pearl, Y. Bengio, K. Imai, J. Ioannidis. Don't miss!! why19.causalai.net #bookofwhy

Hi Tobias, how are you doing? I saw your tweet and realized that you could be a great speaker and/or panelist to our symposium. Would you be interested in joining? If so, I'll send a formal invite.

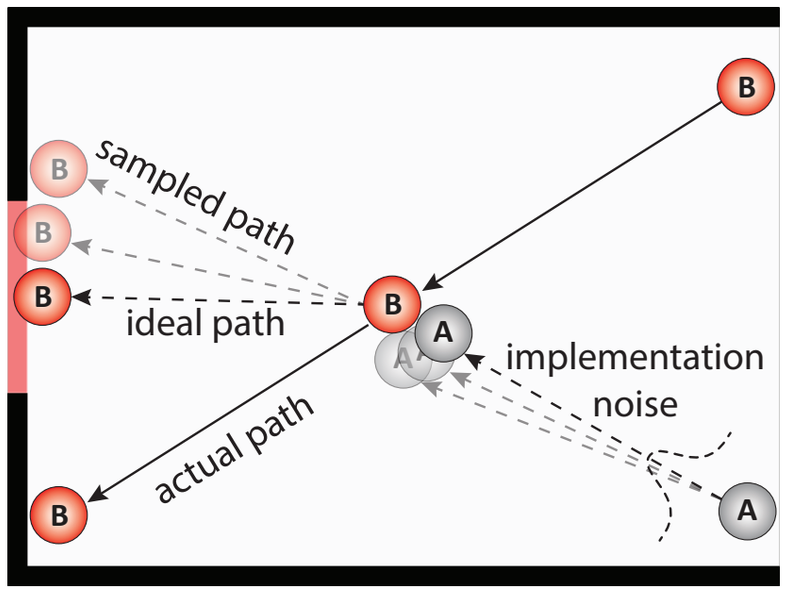
Mar 7

C i C Causality in Cognition

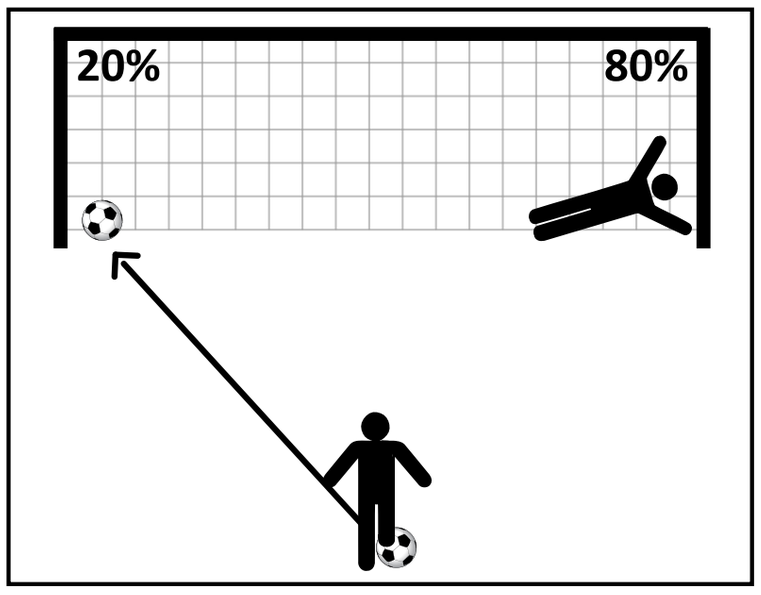
Our lab studies the role of causality in our understanding of the world, and of each other.



learning



reasoning



judgment

@tobigerstenberg

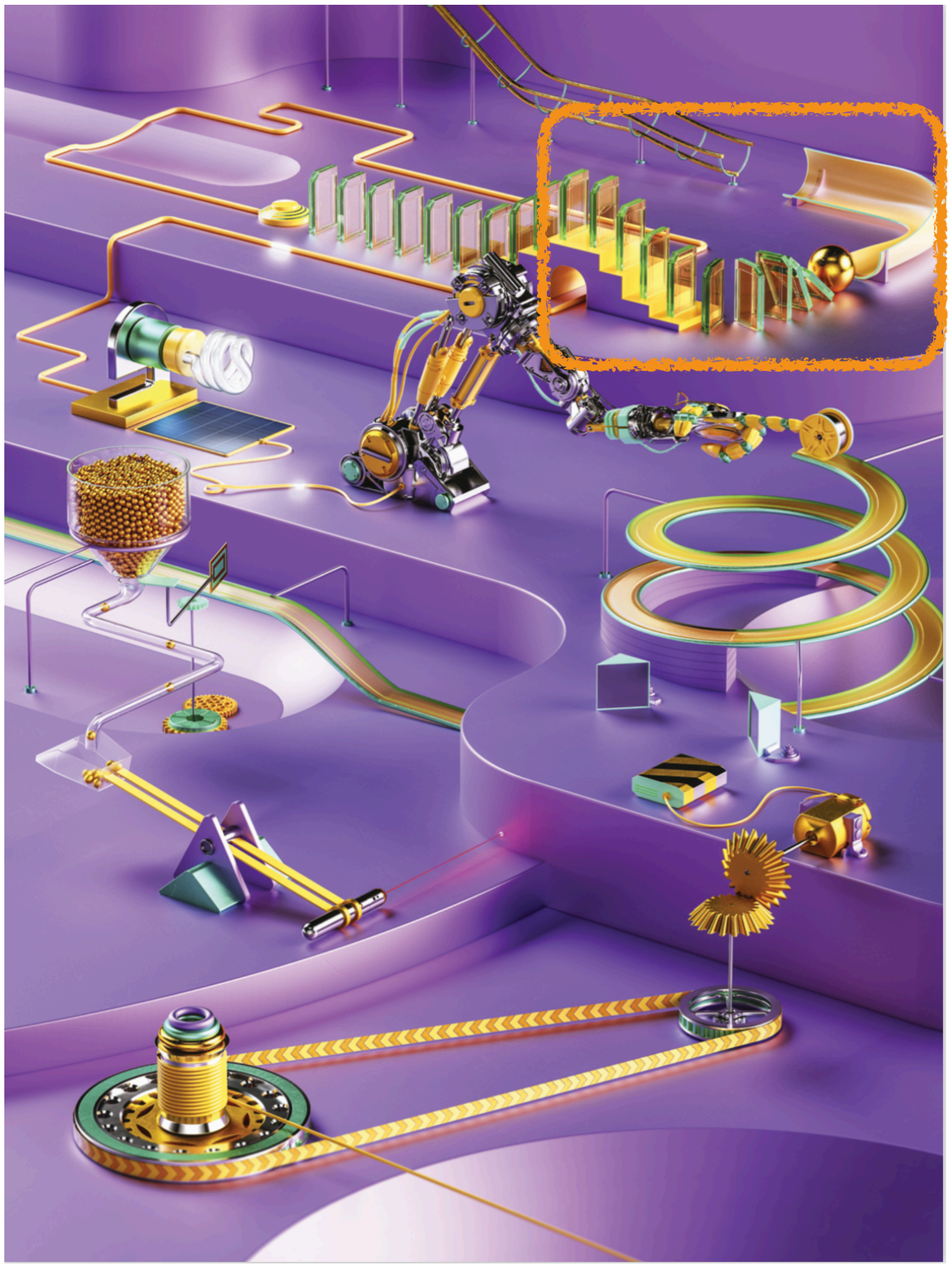
http://cicl.stanford.edu

DOI:10.1145/3241036

The kind of causal inference seen in natural human thought can be "algorithmitized" to help produce human-level machine intelligence.

BY JUDEA PEARL

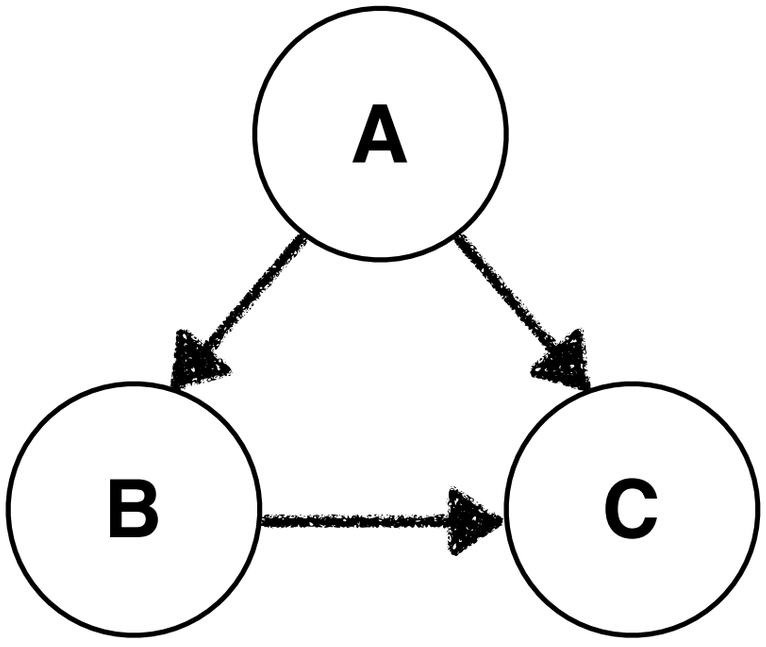
The Seven Tools of Causal Inference, with Reflections on Machine Learning



Pearl, J. (2019). The seven tools of causal inference, with reflections on machine learning. *Communications of the ACM*, 62(3), 54-60.

?

!



causal networks

objects space

forces time

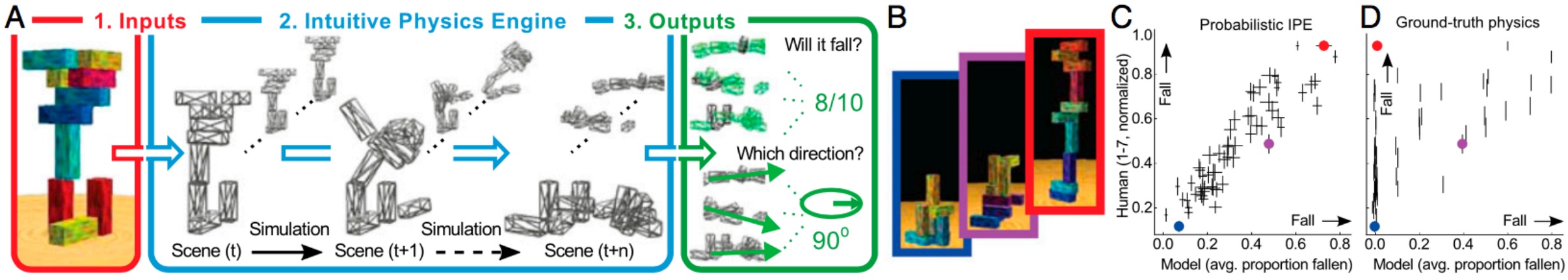
mechanisms

probabilistic programs

Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought. In *The Conceptual Mind: New Directions in the Study of Concepts*. MIT Press.

Gerstenberg, T. & Tenenbaum, J. B. (2017). Intuitive Theories. In *Oxford Handbook of Causal Reasoning*. Oxford University Press.

Intuitive theories as probabilistic programs



Battaglia, Hamrick & Tenenbaum (2013) Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences*

Beyond structural equations

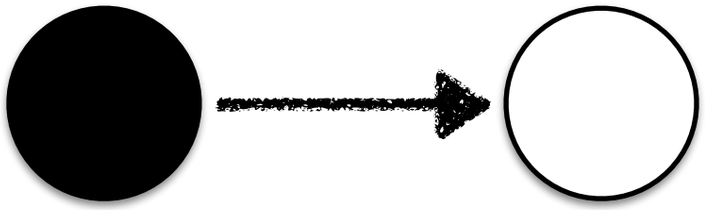
How do we **do** ( ) in a probabilistic program?

How do we simulate counterfactuals?

Outline

Causal judgments

1.



colliding balls

Multi-modal inference

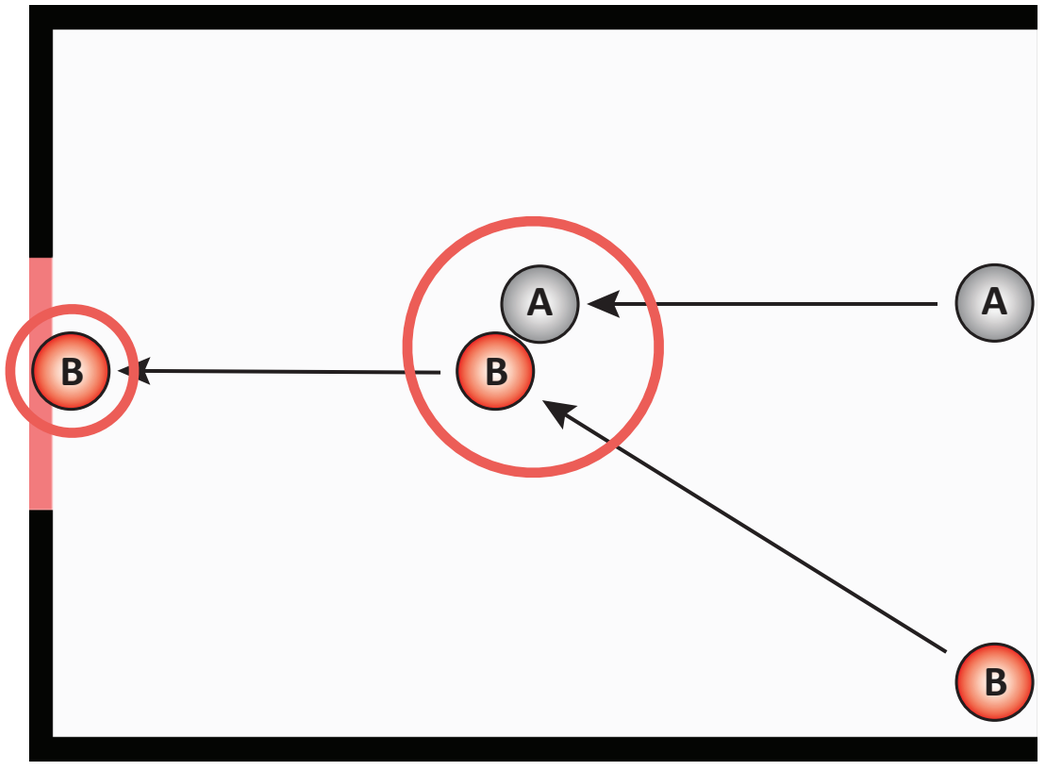
2.

dropping ball

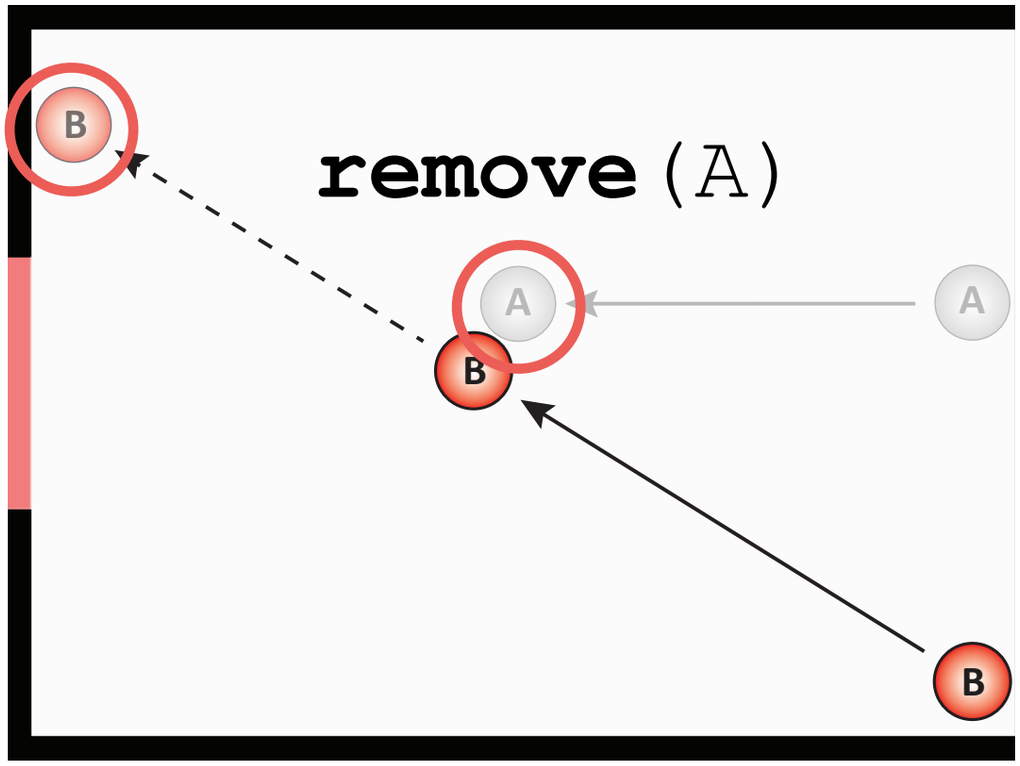
Did A cause B to go through the gate?

Counterfactual Simulation Model

What happened?



What would have happened?



Actual situation

B went through the gate

≠

Counterfactual situation

B would have missed the gate

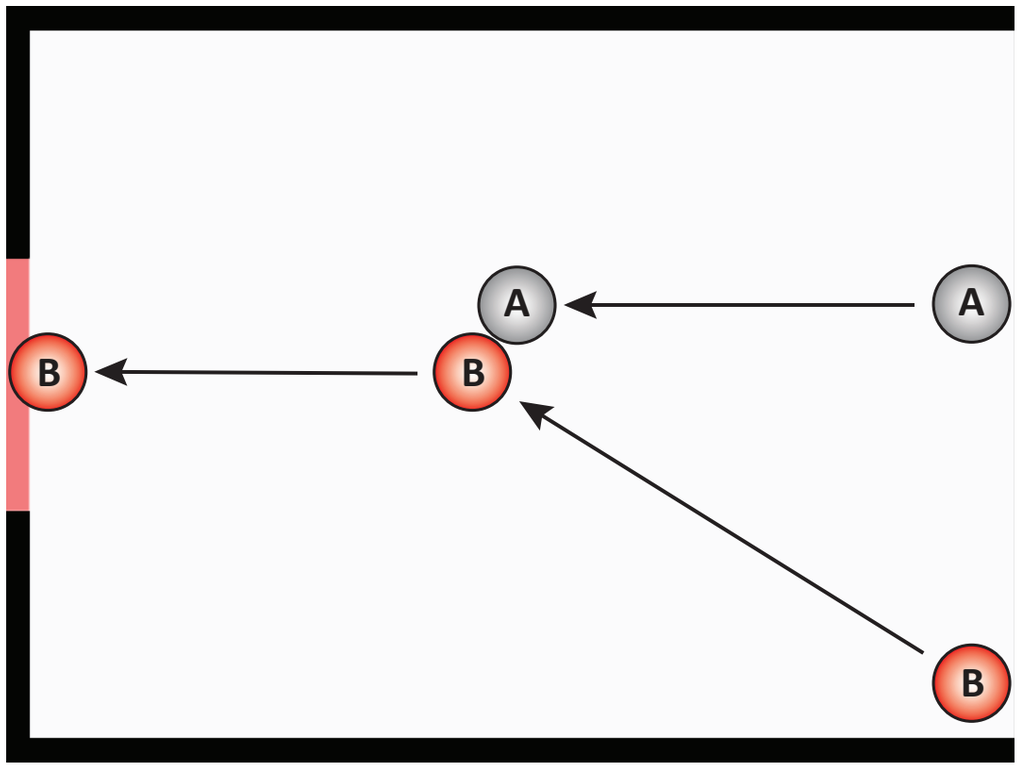
Gerstenberg, Goodman, Lagnado, & Tenenbaum (2012) Noisy Newtons: Unifying process and dependency accounts of causal attribution. *Cognitive Science Proceedings*

Gerstenberg, Goodman, Lagnado, & Tenenbaum (2014) From counterfactual simulation to causal judgment. *Cognitive Science Proceedings*

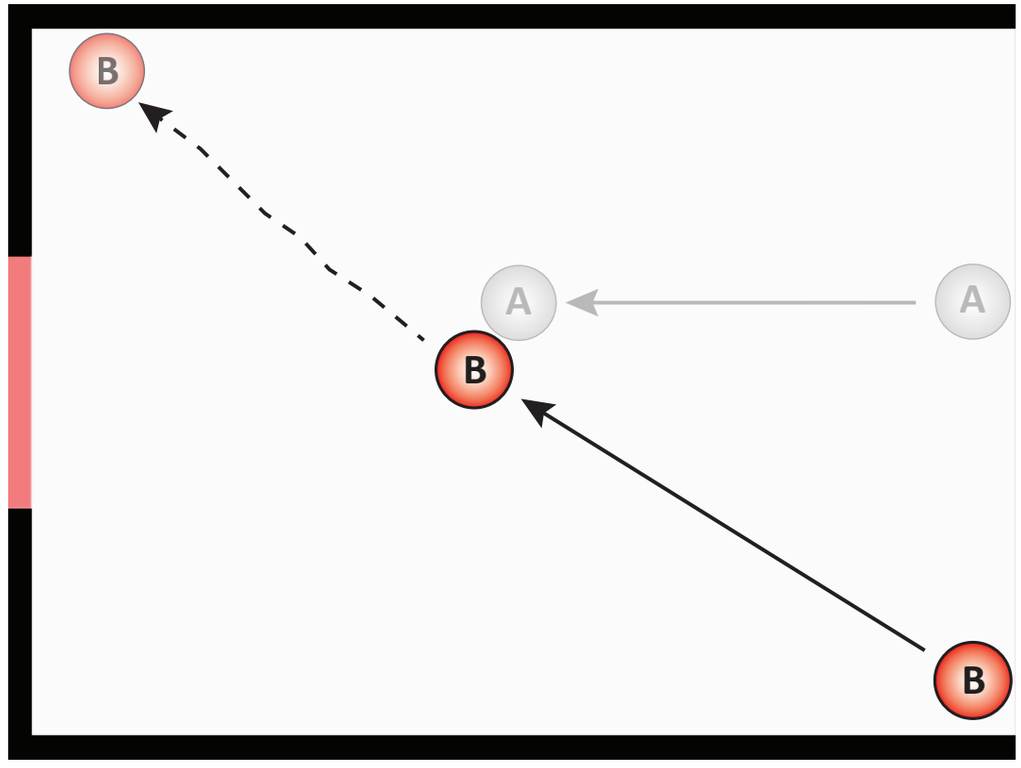
Gerstenberg, Goodman, Lagnado, & Tenenbaum (2015) How, whether, why: Causal judgments as counterfactual contrasts. *Cognitive Science Proceedings*

Quantitative predictions

What happened?



What would have happened?



Actual situation

B went through the gate

?

≠

Counterfactual situation

B would have missed the gate ✓

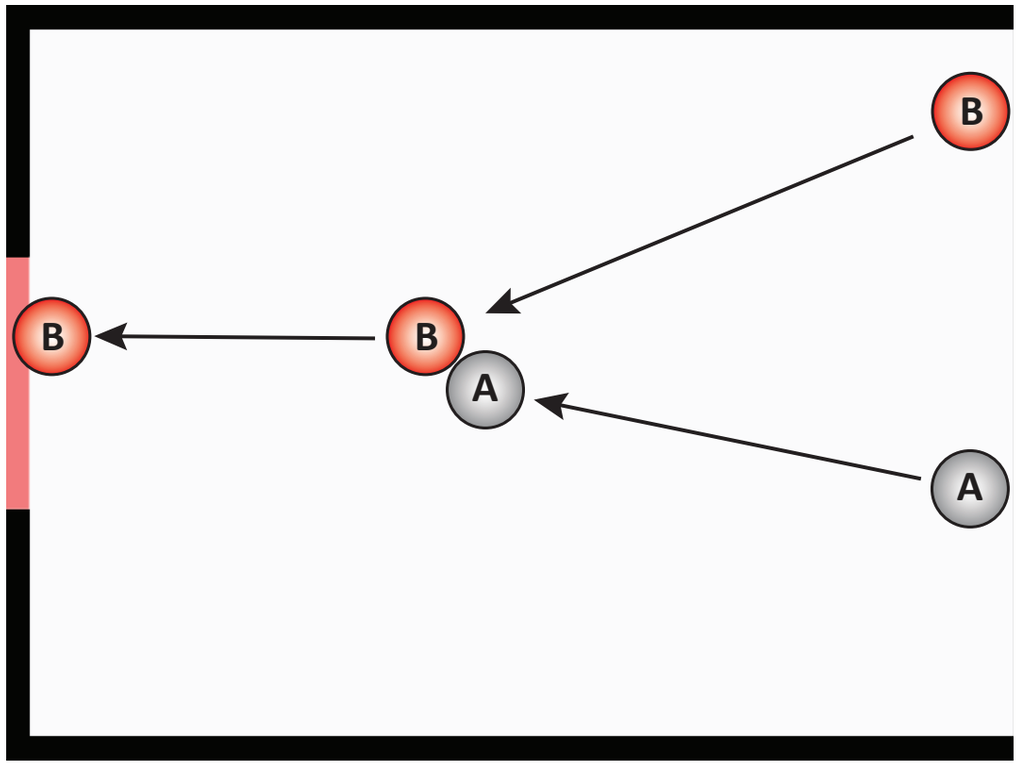
B would have missed the gate ✓

B would have missed the gate ✓

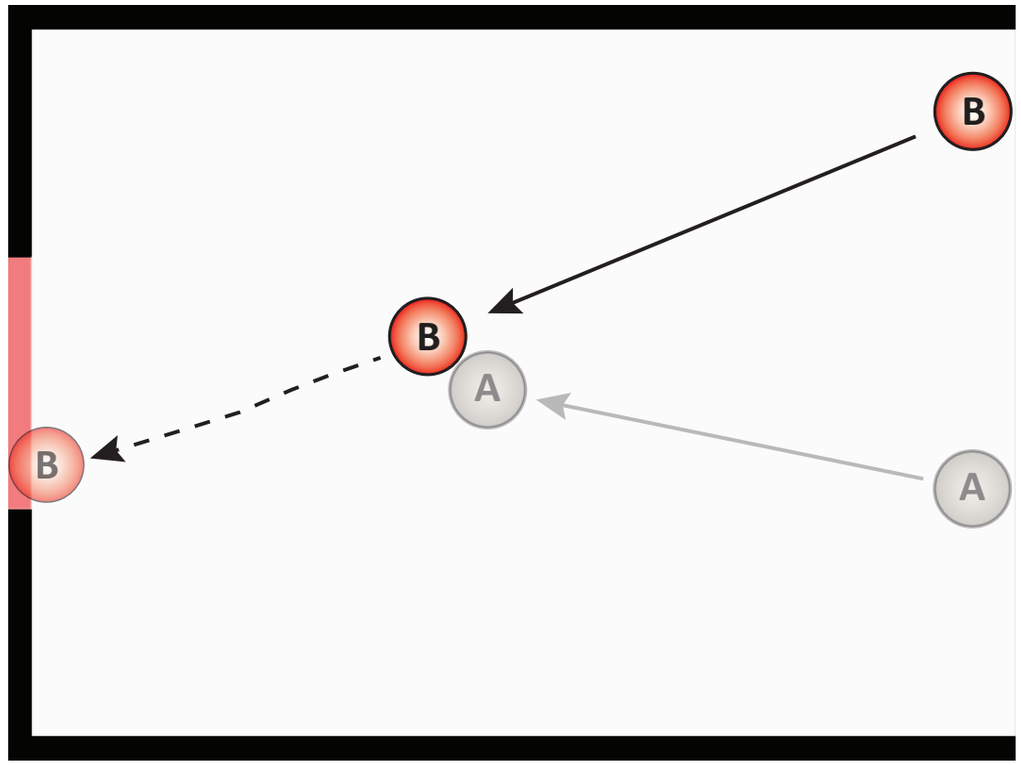
Gerstenberg, Goodman, Lagnado, & Tenenbaum (2012) Noisy Newtons: Unifying process and dependency accounts of causal attribution. Cognitive Science Proceedings

Quantitative predictions

What happened?



What would have happened?



Actual situation

B went through the gate

?

≠

Counterfactual situation

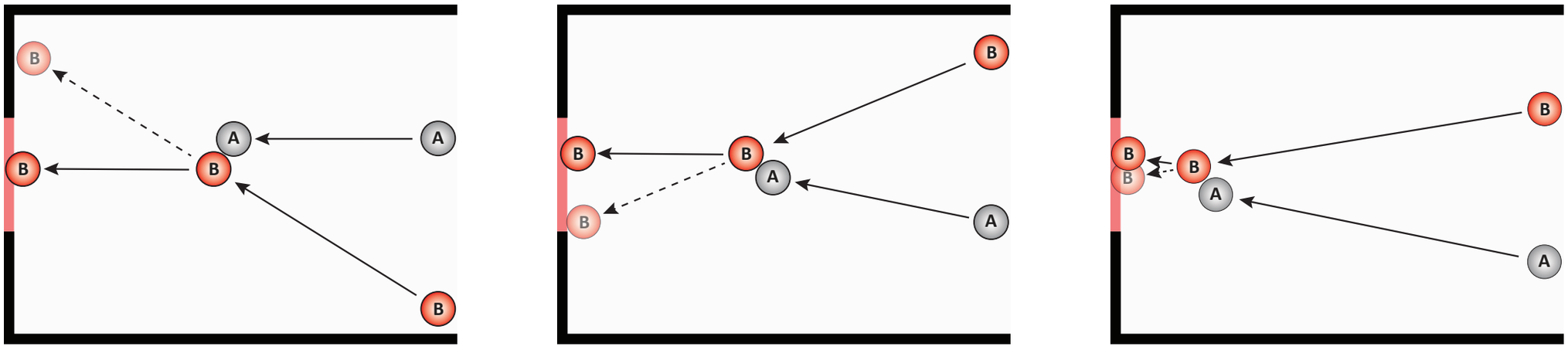
B would have **missed** the gate ✓

B would have **gone through** gate ✗

B would have **gone through** gate ✗

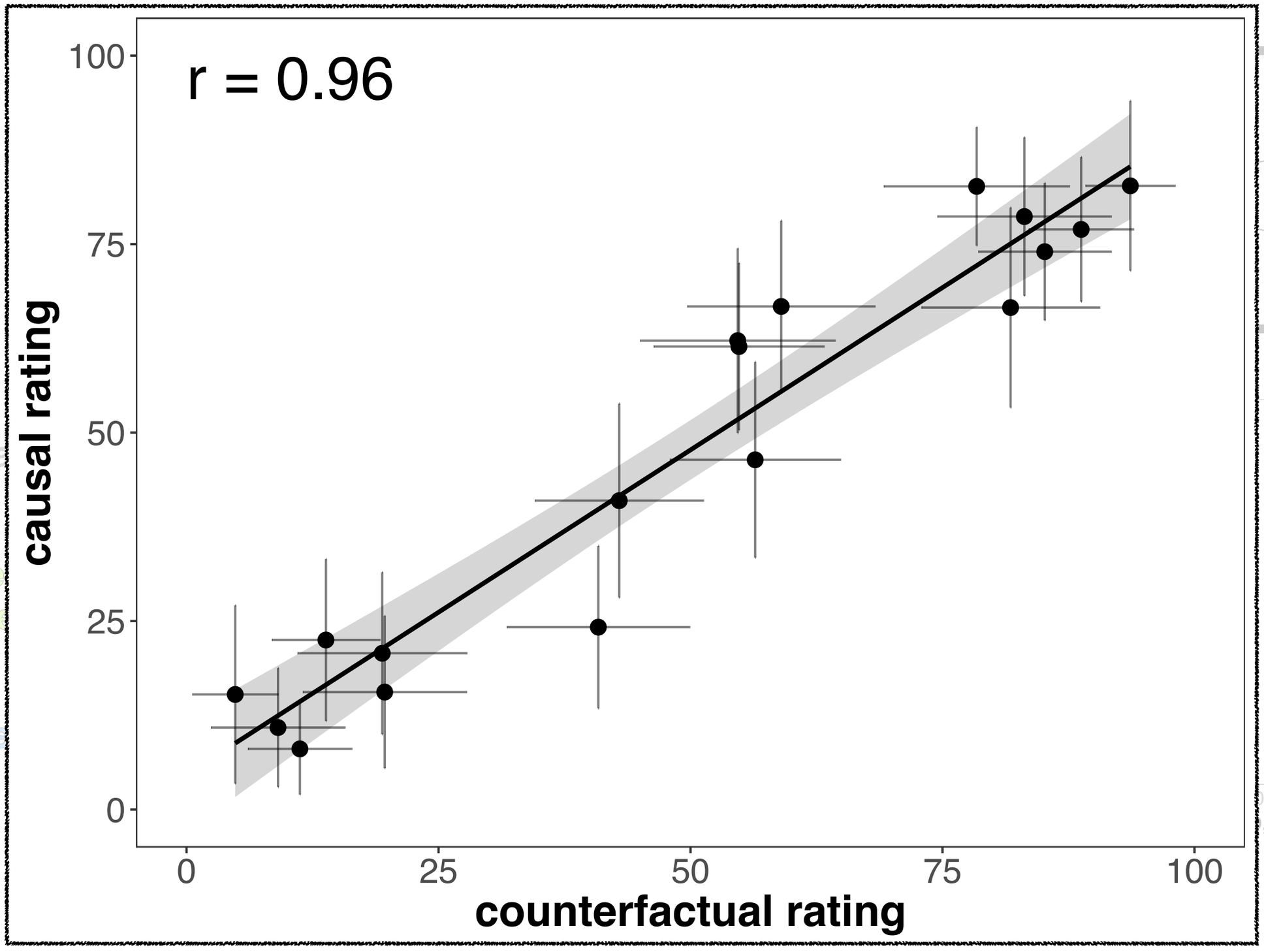
Gerstenberg, Goodman, Lagnado, & Tenenbaum (2012) Noisy Newtons: Unifying process and dependency accounts of causal attribution. Cognitive Science Proceedings

Quantitative predictions



Gerstenberg, Goodman, Lagnado, & Tenenbaum (2012) Noisy Newtons: Unifying process and dependency accounts of causal attribution. Cognitive Science Proceedings

Quantitative predictions

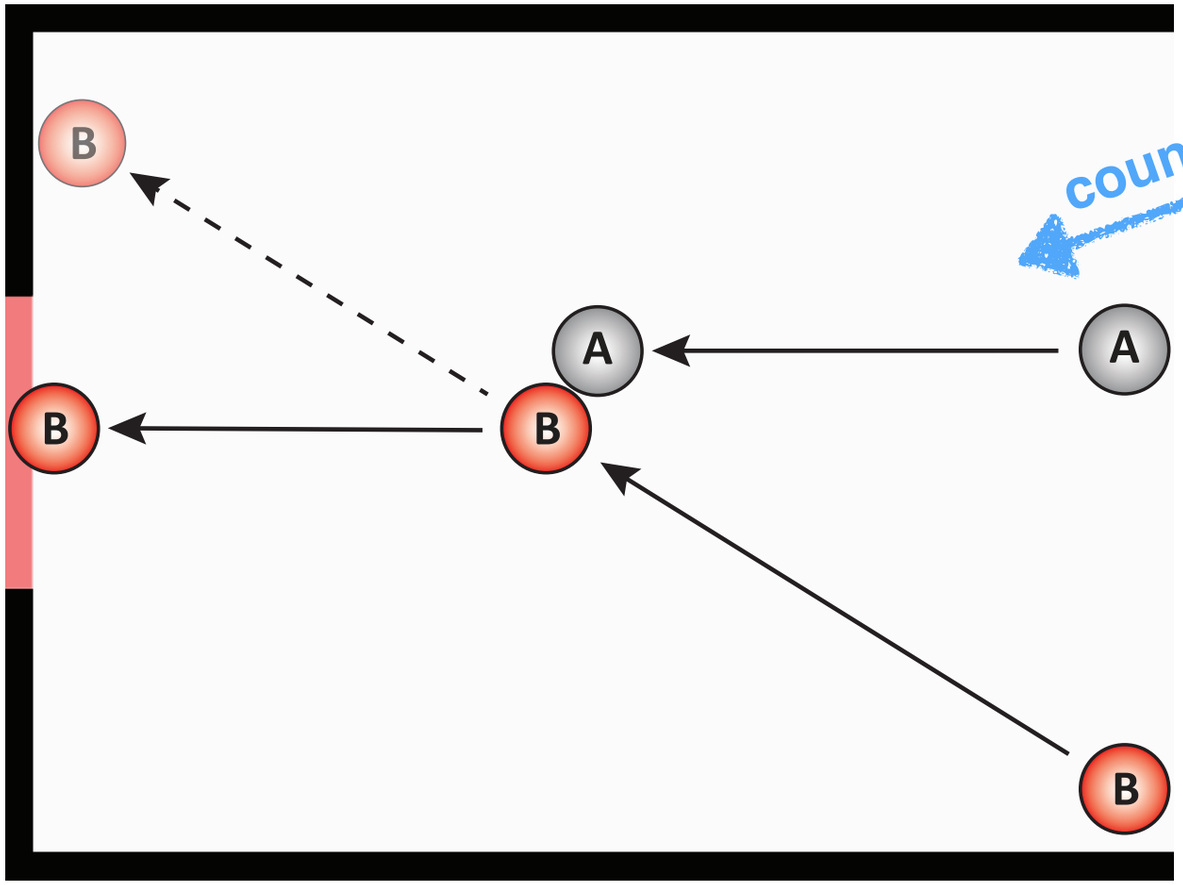


Gerstenberg, Goodman, Lagnado, & Tenenbaum (2012) Noisy Newtons: Unifying process and dependency accounts of causal attribution. Cognitive Science Proceedings

Causal system

represent

Causal judgment



Probabilistic program

```
//Define table with walls
function createTable(wall.x,wall.y,wall.length,wall.width){...}
//Define balls
function createBalls(x.position,y.position,x.velocity,y.velocity){...}

//Define world
function createWorld(table, ball1, ball2){
    createTable(...);
    createBalls(...);
    return(world)
}

//Define actual world
function simulateWorld(world){
    while (time < endOfClip){
        // When did the balls collide?
        if (collision){
            collisionTime = time;
        }
        time++;
    }
    if (ball2.x < 0){outcome = 1} //Did ball go through the gate?
    return (outcome)
}

//Define counterfactual world
function simulateCounterfactual(world){
    while (time < endOfClip){
        if (time == collisionTime-1){
            remove(ball1);
        }
        if (time >= collisionTime){
            addNoise(ball2); //add noise to ball trajectory
        }
        time++;
    }
    if (ball2.x < 0){outcome = 1} //Did ball go through the gate?
    return (outcome)
}

//Run actual world
actualWorld = createWorld(table, ball1, ball2);
outcomeActual = simulateWorld(actualWorld);

//Run counterfactual world
counterfactualWorld = createWorld(table, ball1, ball2);
outcomeCounterfactual = simulateWorld(counterfactualWorld);

//Test for causation
if(outcomeActual != outcomeCounterfactual){
    cause = 1;
}
```

objects

processes

uncertainty

comparison

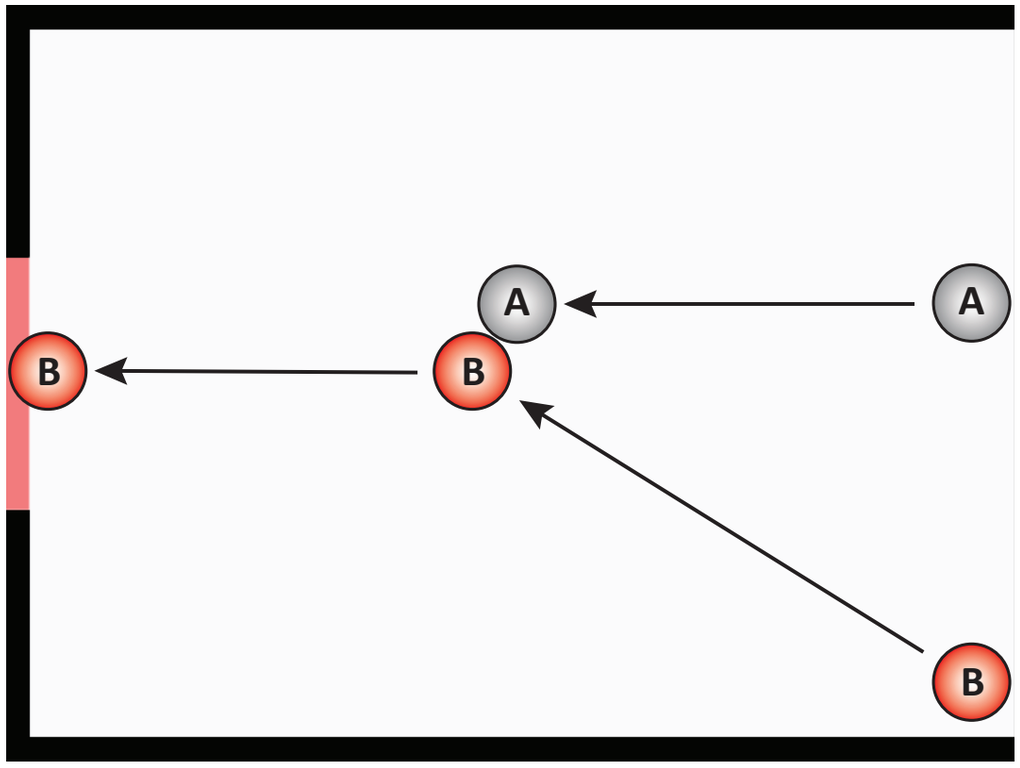
Chater & Oaksford (2013) Programs as causal models: Speculations on mental programs and mental representation. Cognitive Science

Goodman, Tenenbaum, & Gerstenberg (2015) Concepts in a probabilistic language of thought. The Conceptual Mind: New Directions in the Study of Concepts

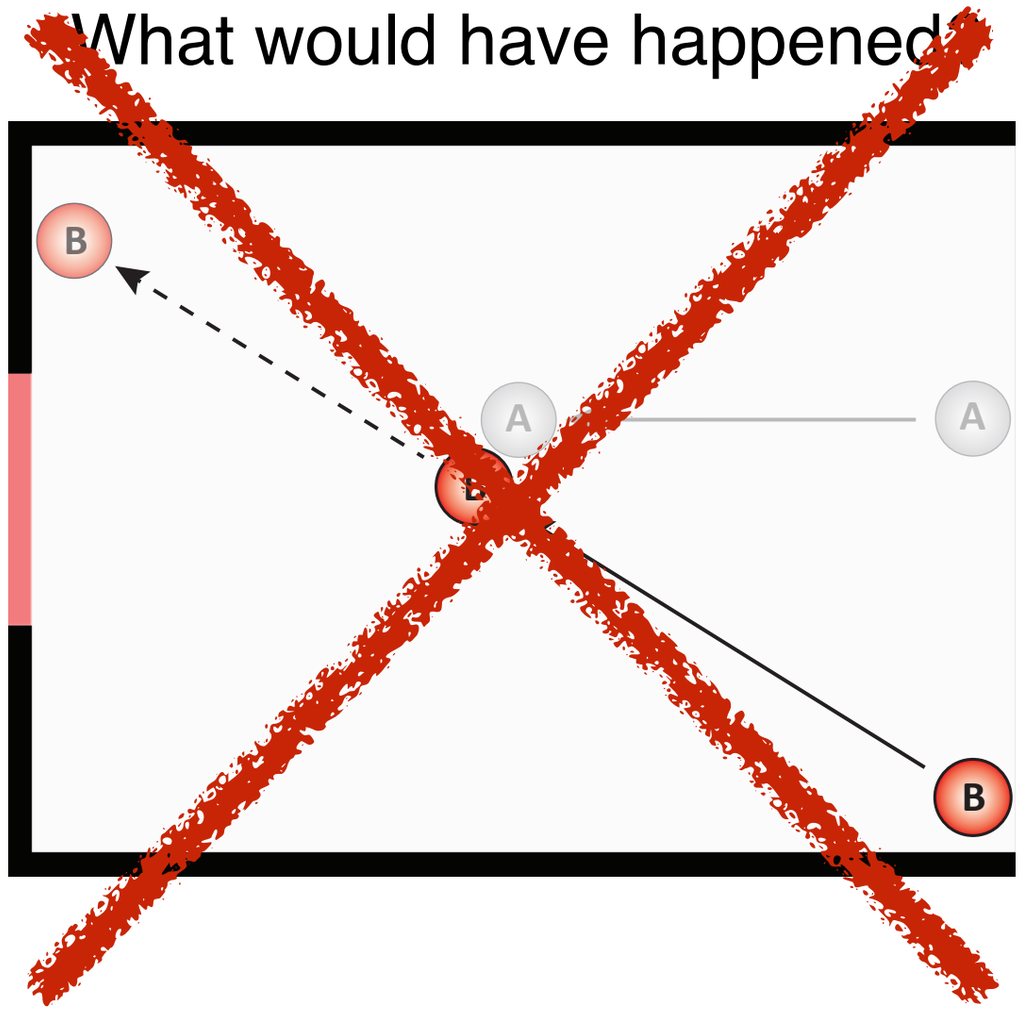
What else?

Actualist theories of causation

What happened?



What would have happened?



Hall (2004) Two concepts of causation. Causation and Counterfactuals

Paul & Hall (2013). Causation: A User's Guide

Salmon (1994) Causality without counterfactuals. Philosophy of Science

Talmy (1988) Force dynamics in language and cognition. Cognitive Science

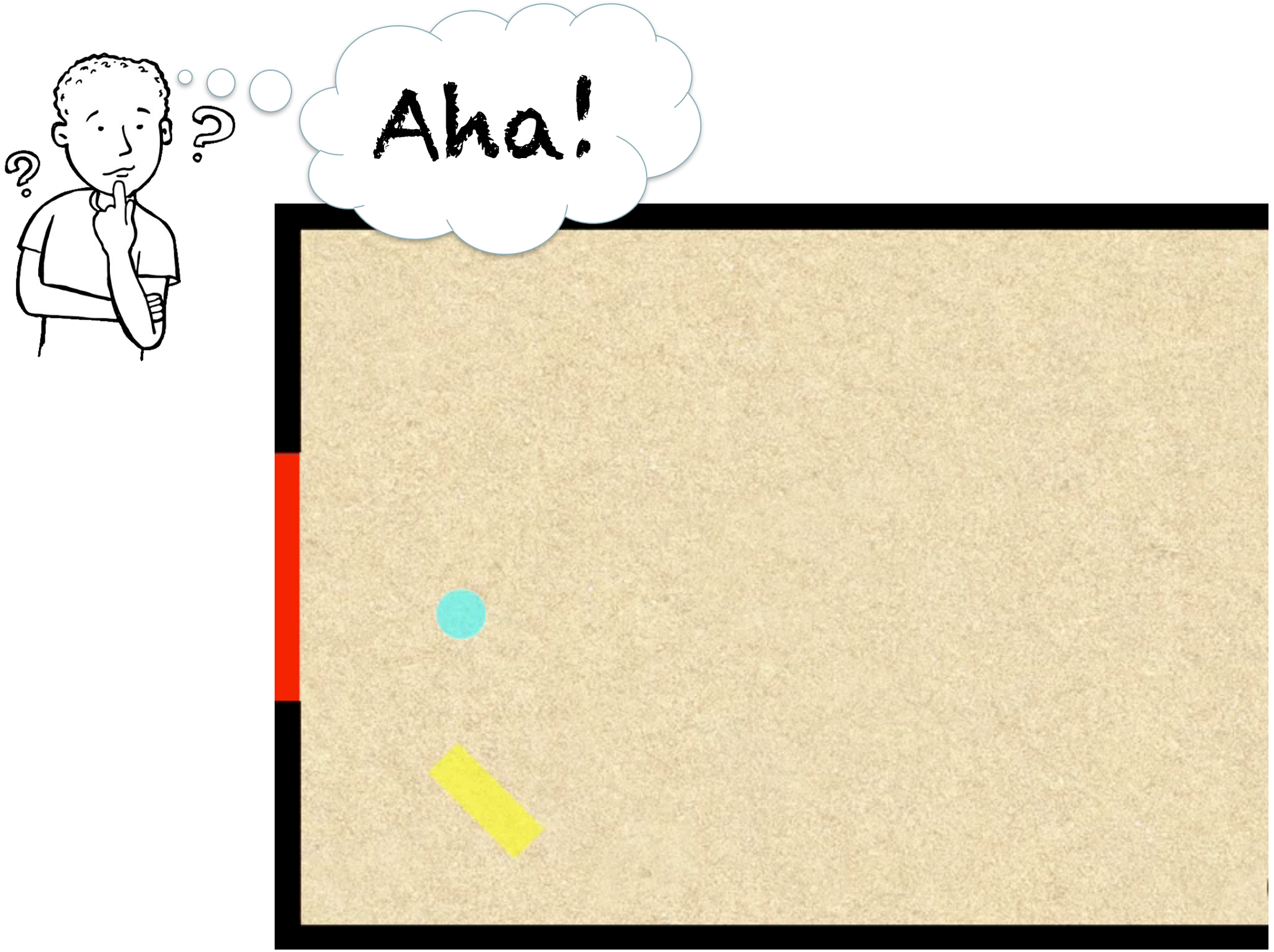
Walsh & Sloman (2011) The meaning of cause and prevent: The role of causal mechanism. Mind & Language

Wolff (2007) Representing causation. Journal of Experimental Psychology: General

Counterfactuals are necessary

Gerstenberg, Goodman, Lagnado, & Tenenbaum (2014) From counterfactual simulation to causal judgment. Cognitive Science Proceedings

Counterfactuals are necessary



Gerstenberg, Goodman, Lagnado, & Tenenbaum (2014) From counterfactual simulation to causal judgment. Cognitive Science Proceedings

Counterfactuals are necessary

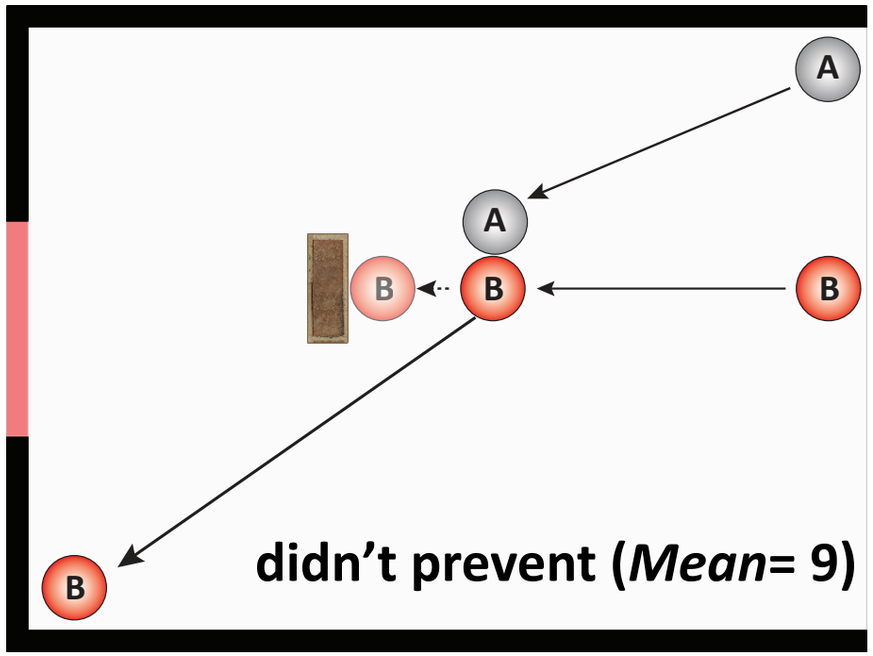
Did A prevent B from going through the gate?

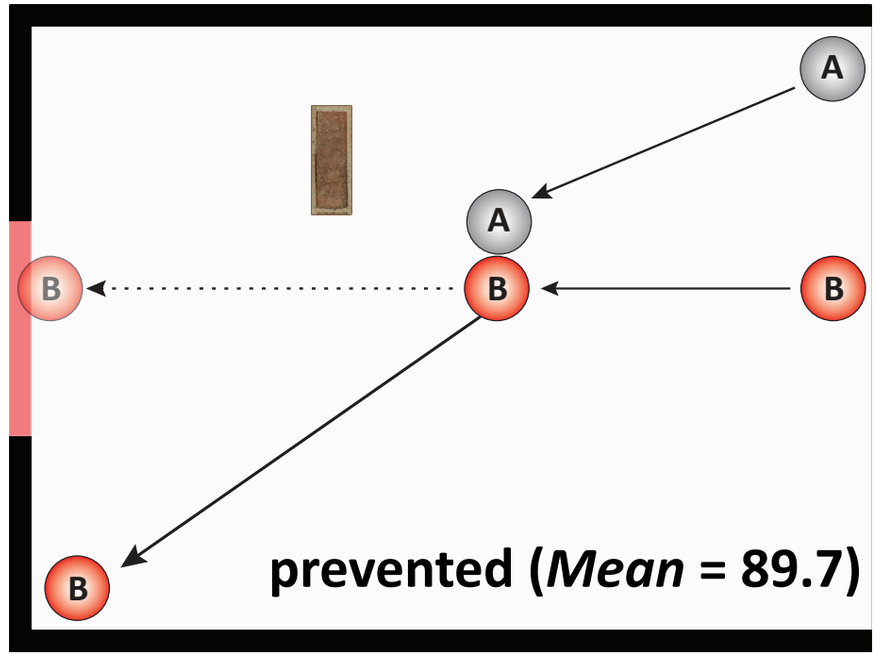
Actual

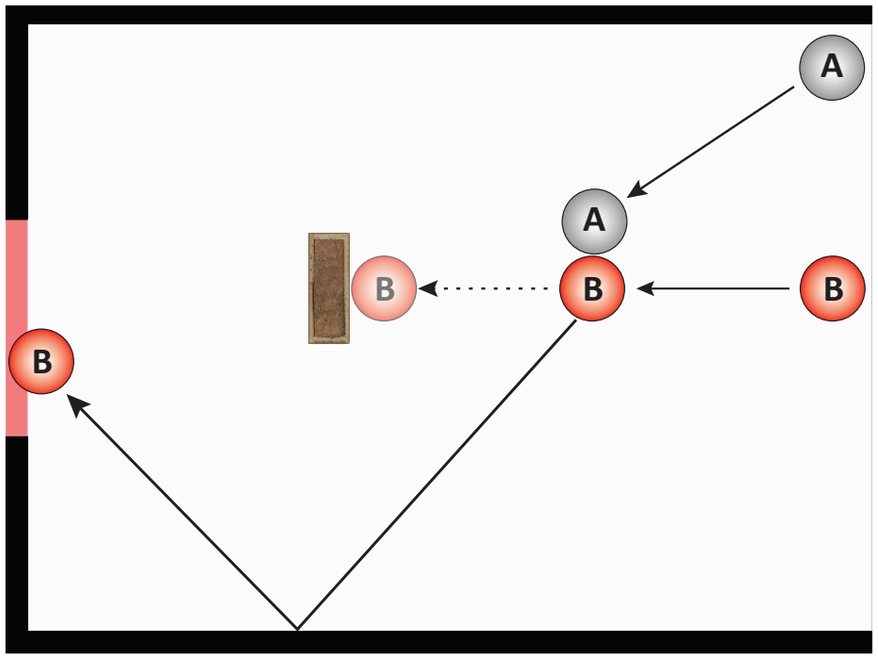
Counterfactual

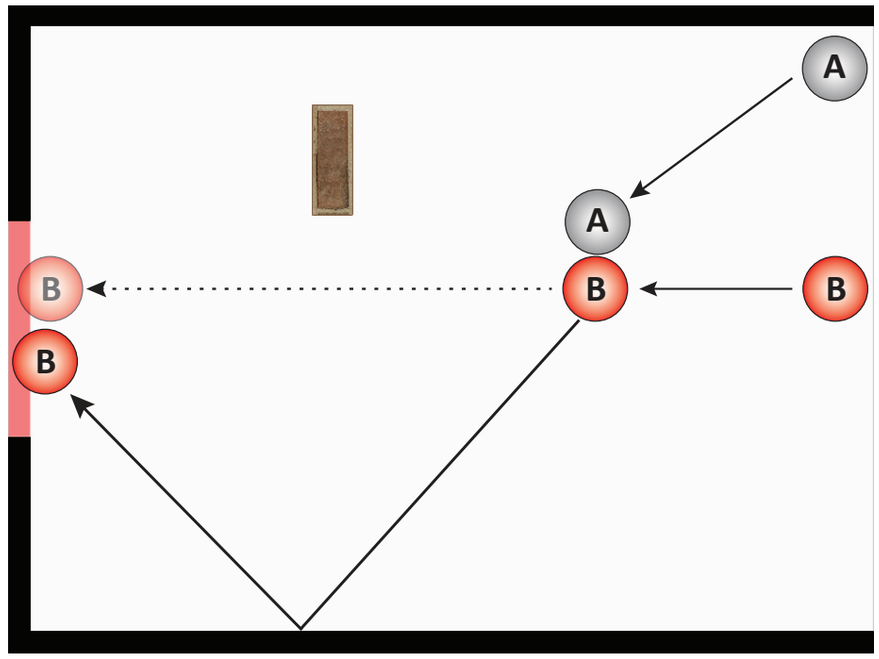
Gerstenberg, Goodman, Lagnado, & Tenenbaum (2014) From counterfactual simulation to causal judgment. Cognitive Science Proceedings

Counterfactuals are necessary







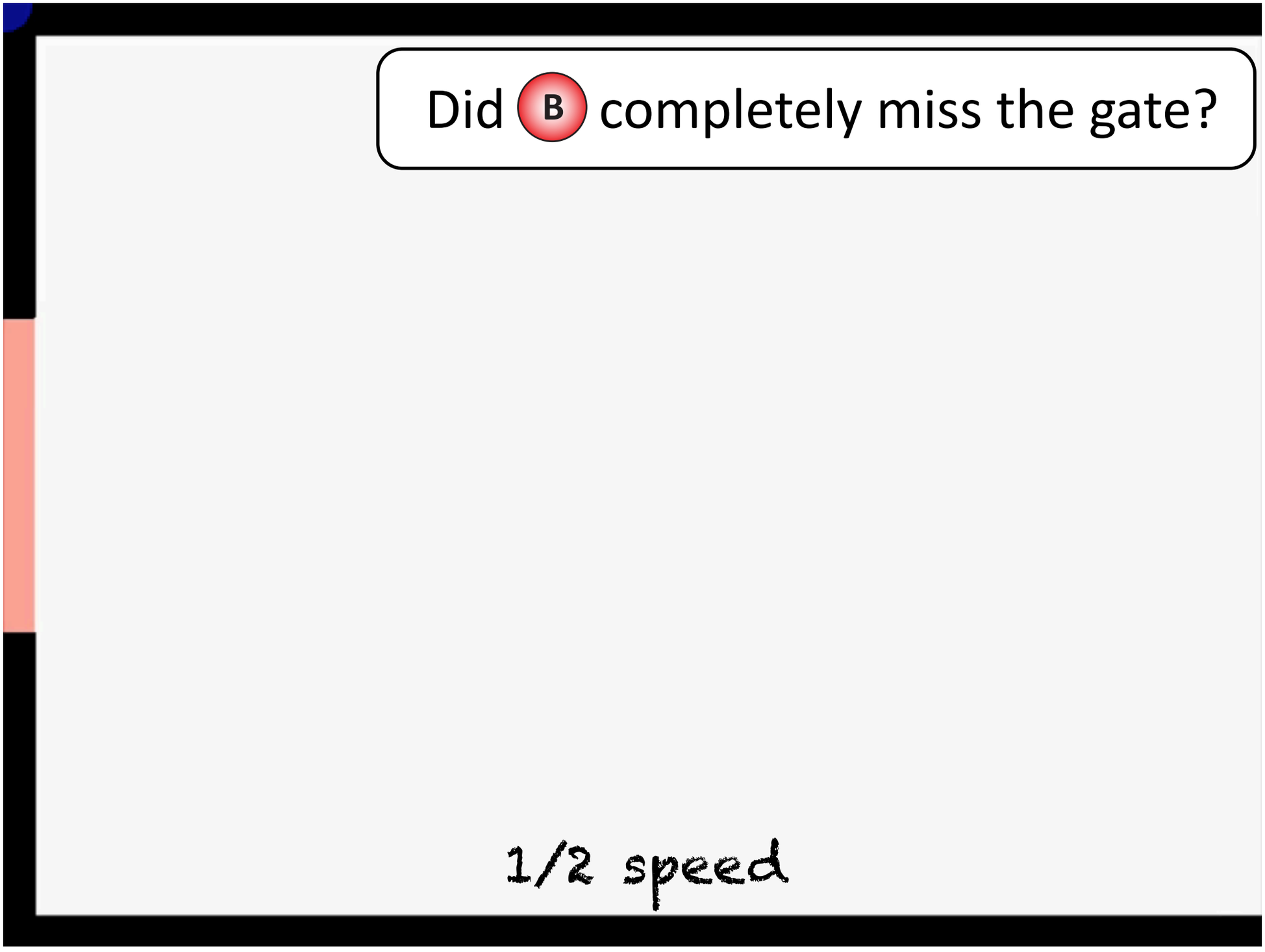


Gerstenberg, Goodman, Lagnado, & Tenenbaum (2014) From counterfactual simulation to causal judgment. Cognitive Science Proceedings

Spontaneous counterfactual simulation

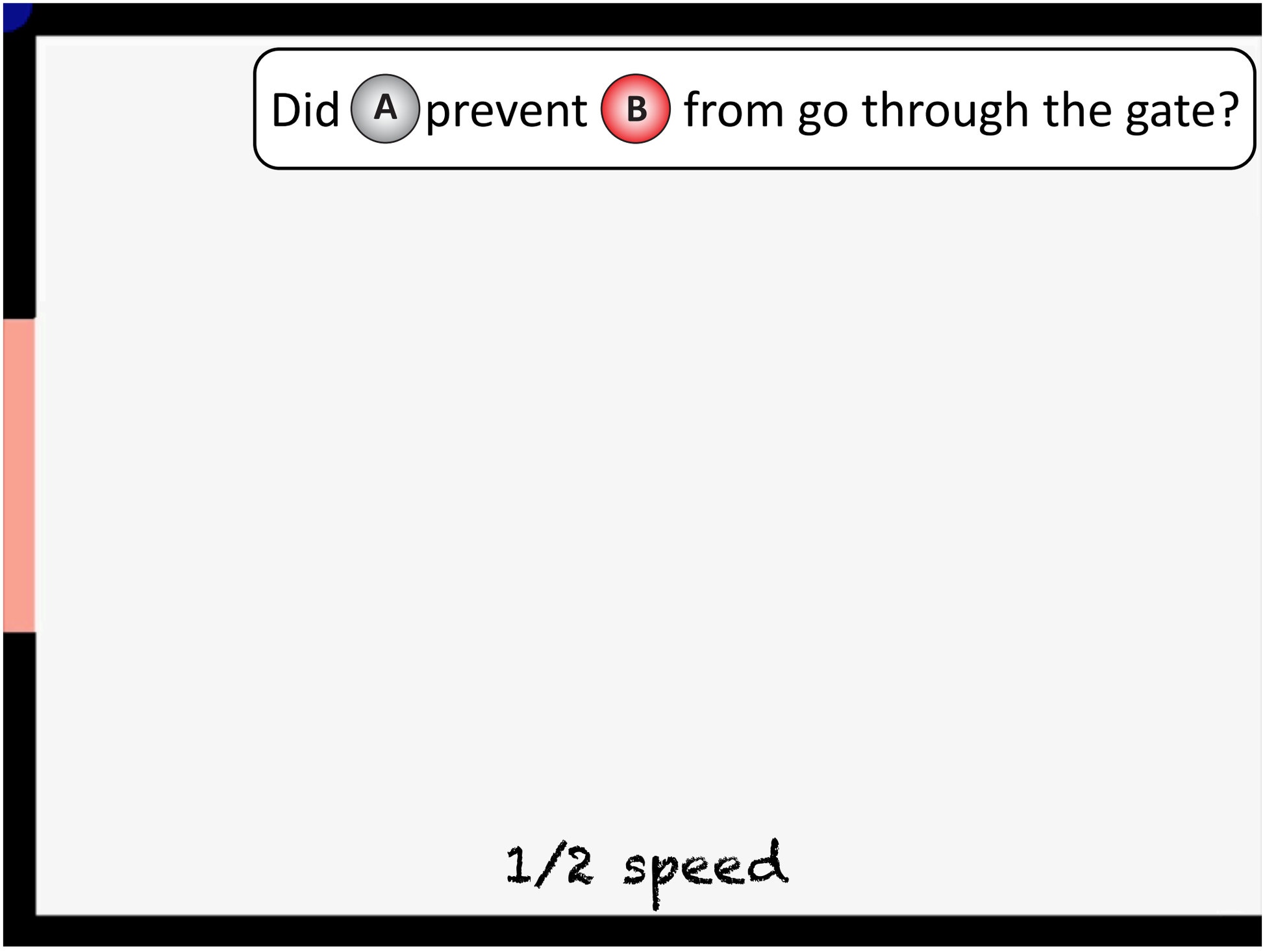
Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. *Psychological Science*

Spontaneous counterfactual simulation



Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. *Psychological Science*

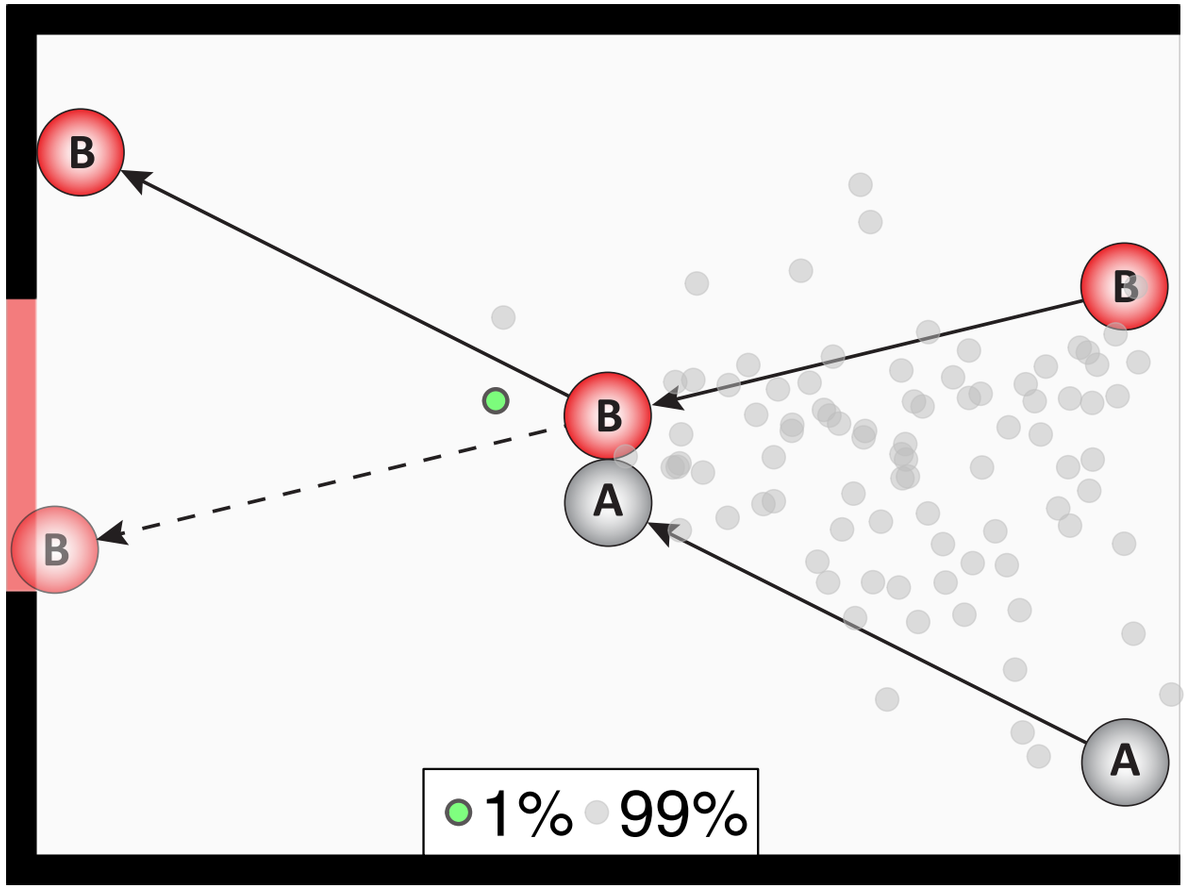
Spontaneous counterfactual simulation



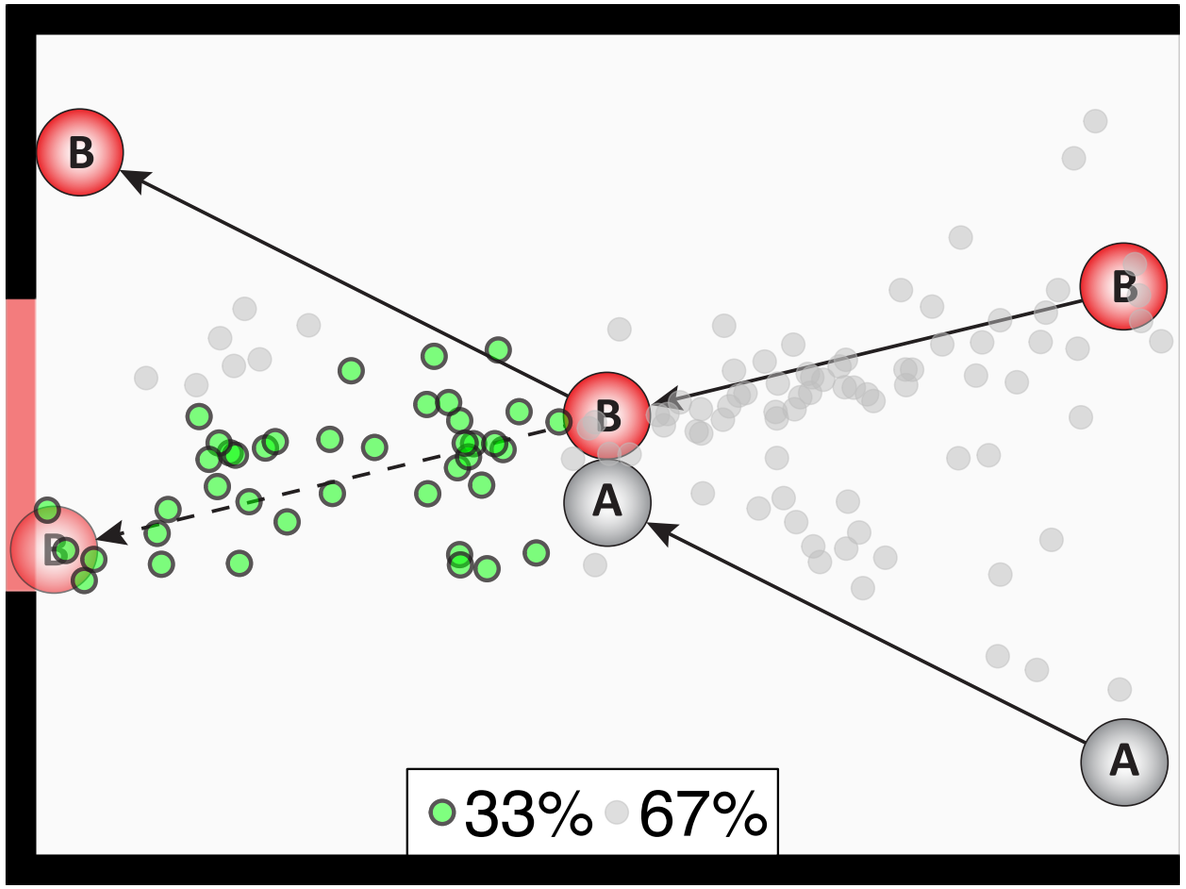
Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. *Psychological Science*

Spontaneous counterfactual simulation

Did B completely miss the gate?

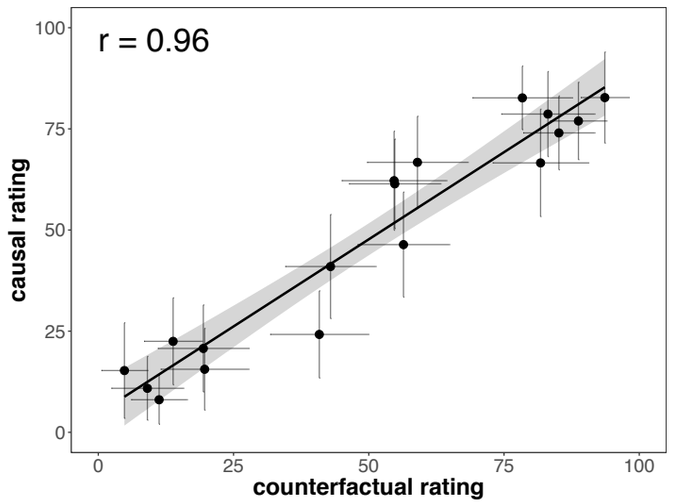


Did A prevent B from go through the gate?

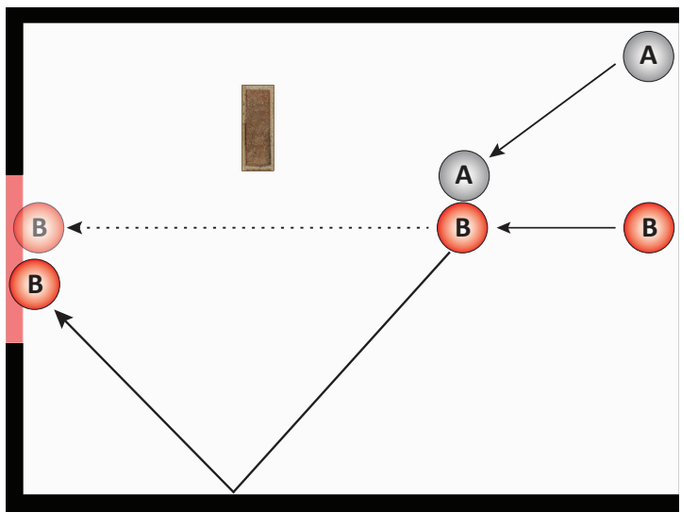


Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. *Psychological Science*

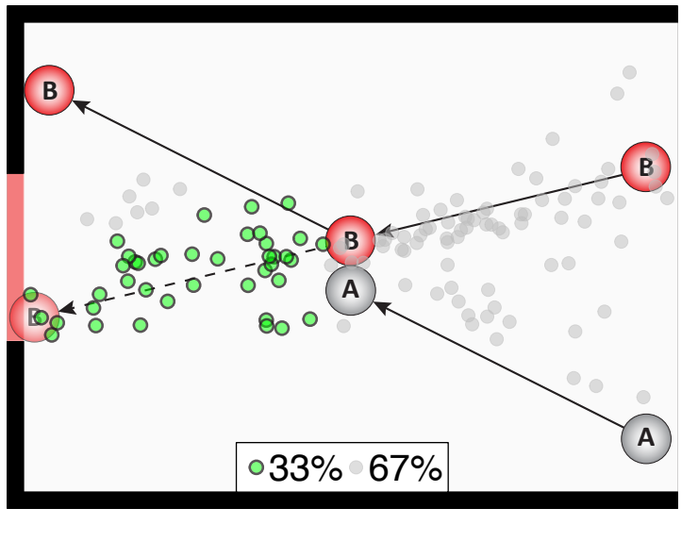
Counterfactual simulation model of causal judgment



- causal judgments are well-explained by the observer's beliefs about whether the candidate cause made a difference to the outcome



- counterfactual contrasts are necessary for explaining people's causal judgments



- people spontaneously engage in counterfactual simulation when making causal judgments

Outline

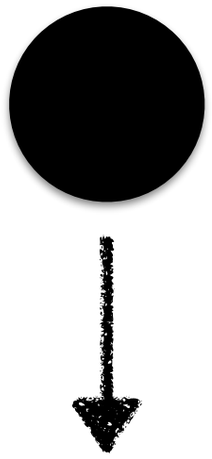
Causal judgments

1.

colliding balls

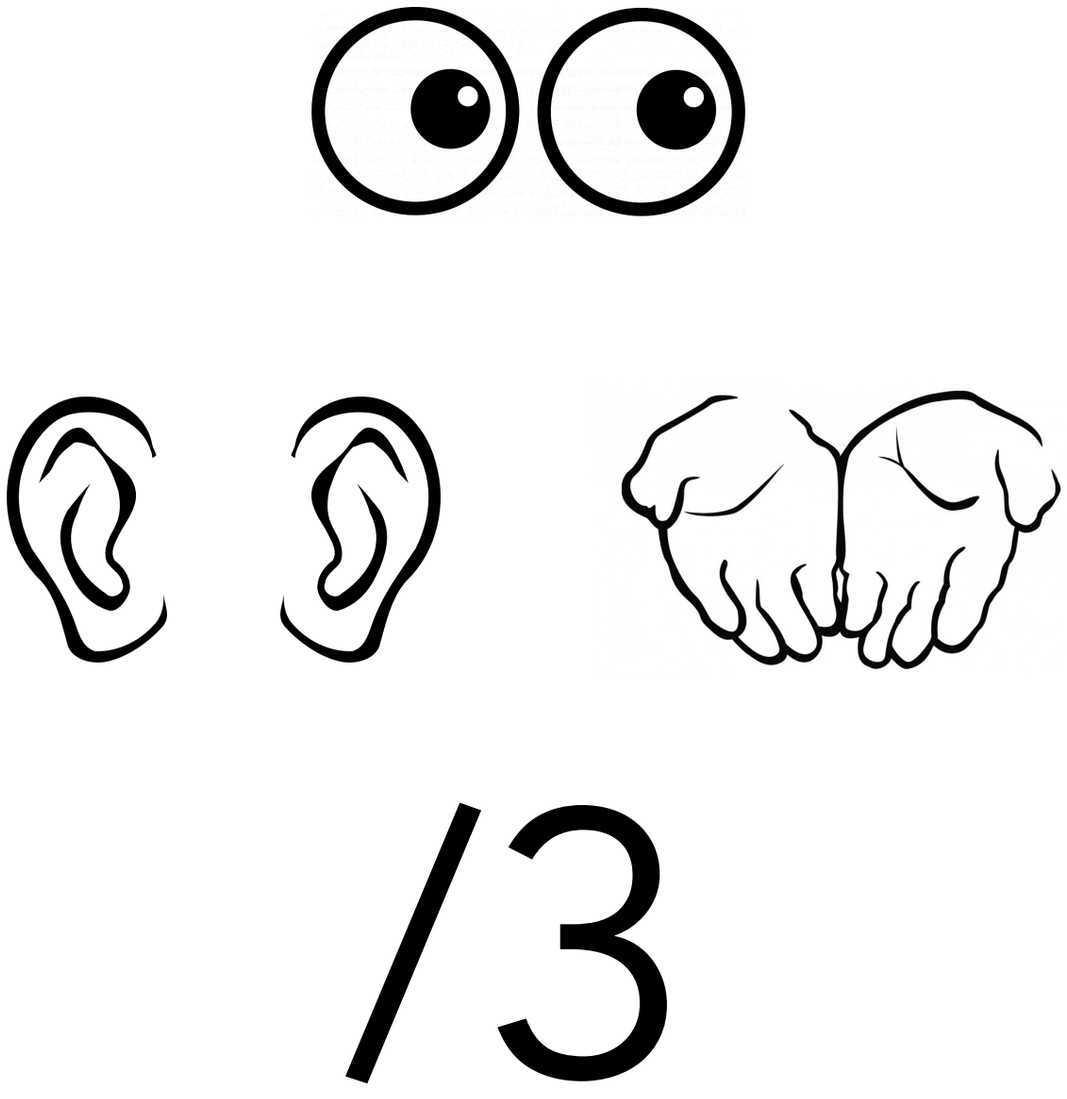
Multi-modal inference

2.



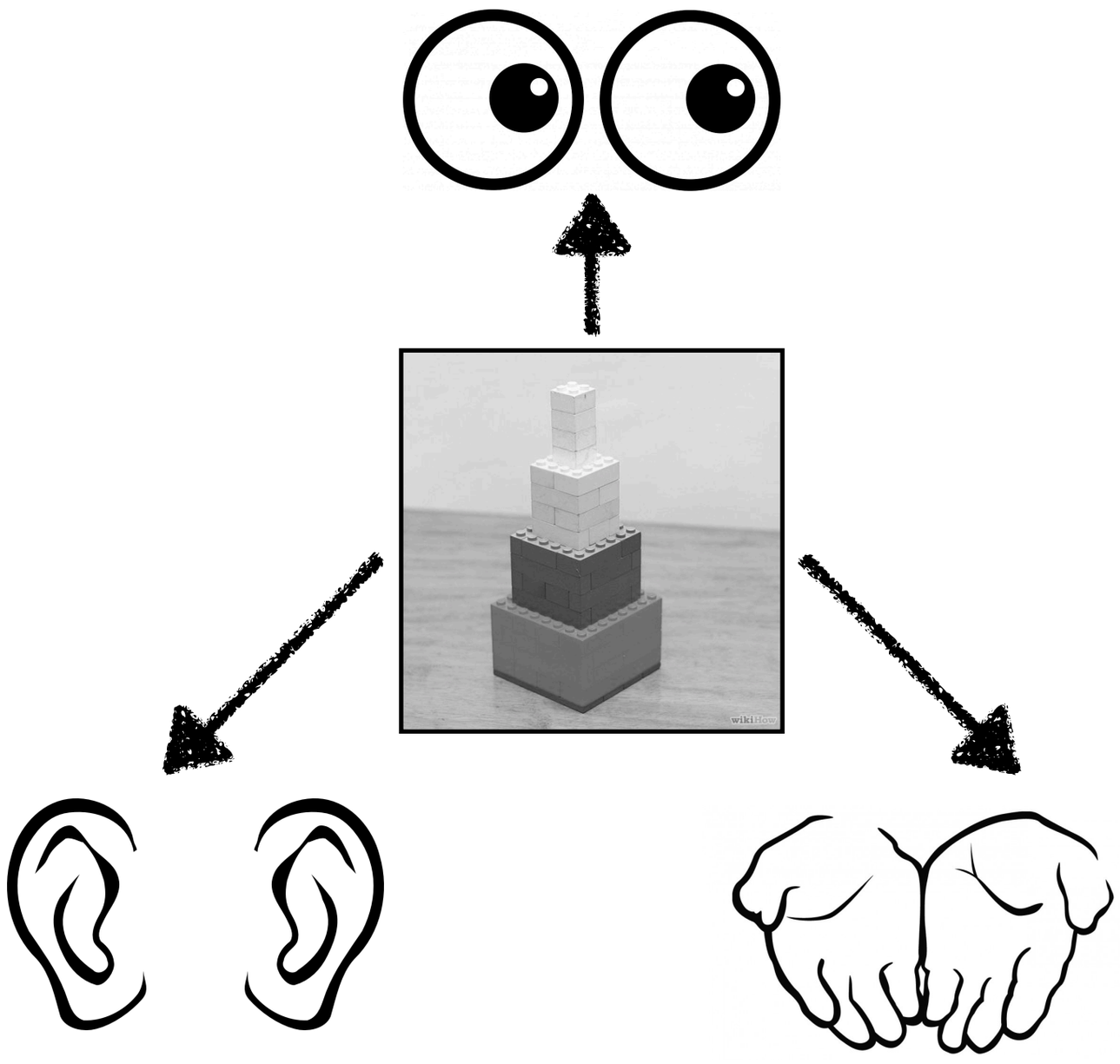
dropping ball

Causal inference: Multi-modal integration through mental simulation



/3

statistical integration



integration via mental simulation

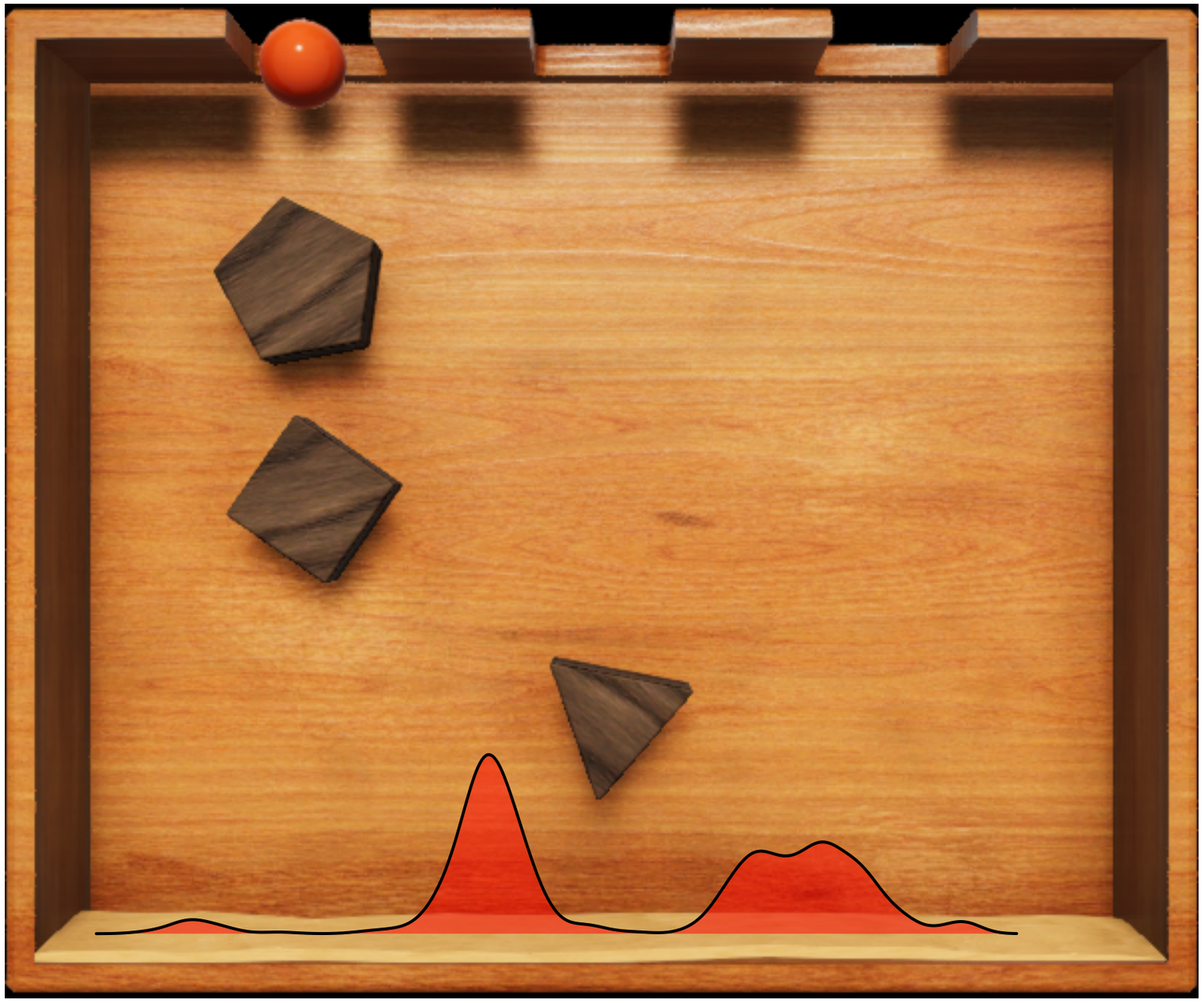
Siegel, Magid, Tenenbaum, & Schulz (2014) Black boxes: Hypothesis testing via indirect perceptual evidence. *CogSci Proceedings*

Yildirim (2014) From perception to conception: learning multisensory representations. *PhD thesis*

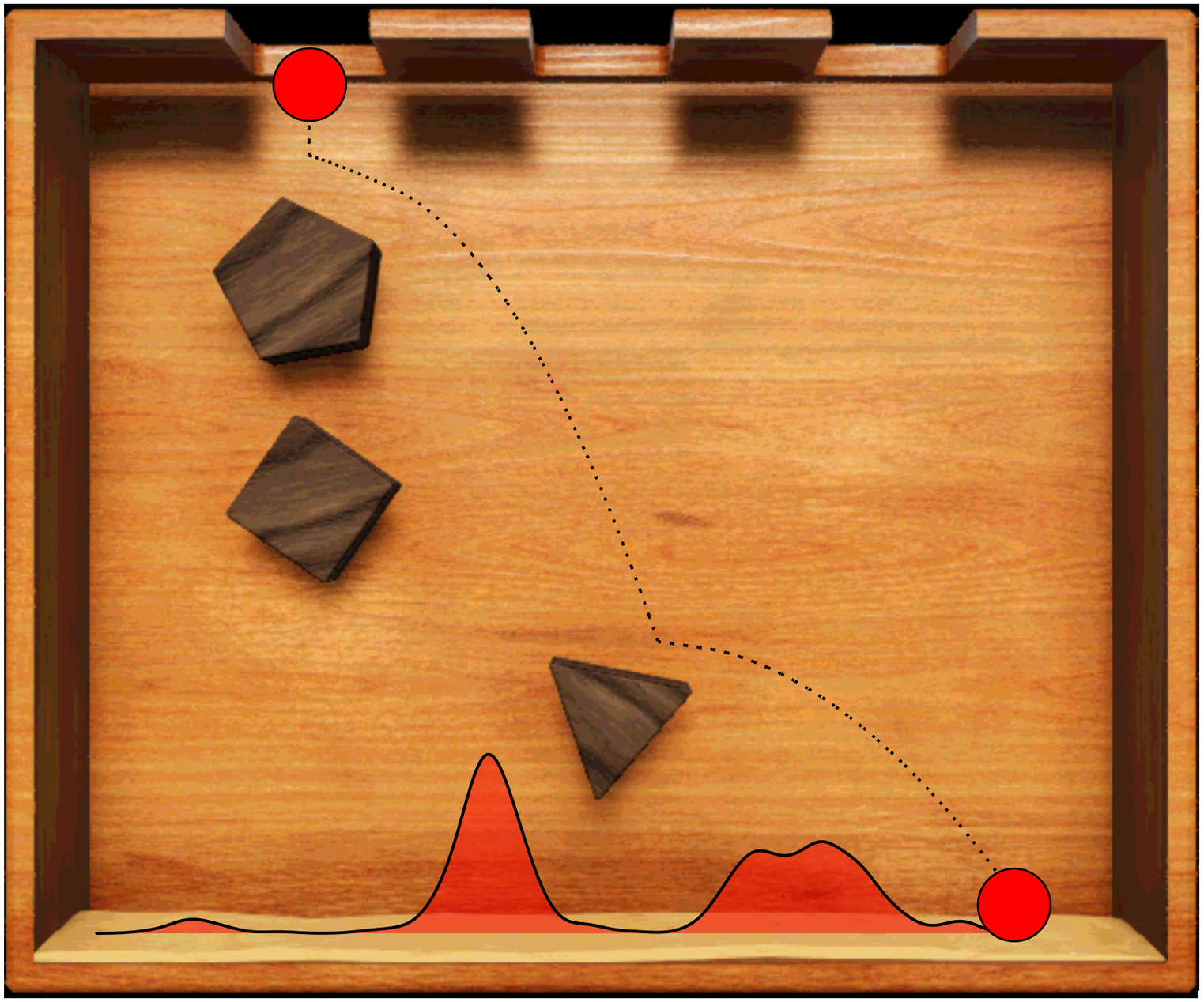
Prediction: Where will the ball land?

CREEPY HAND

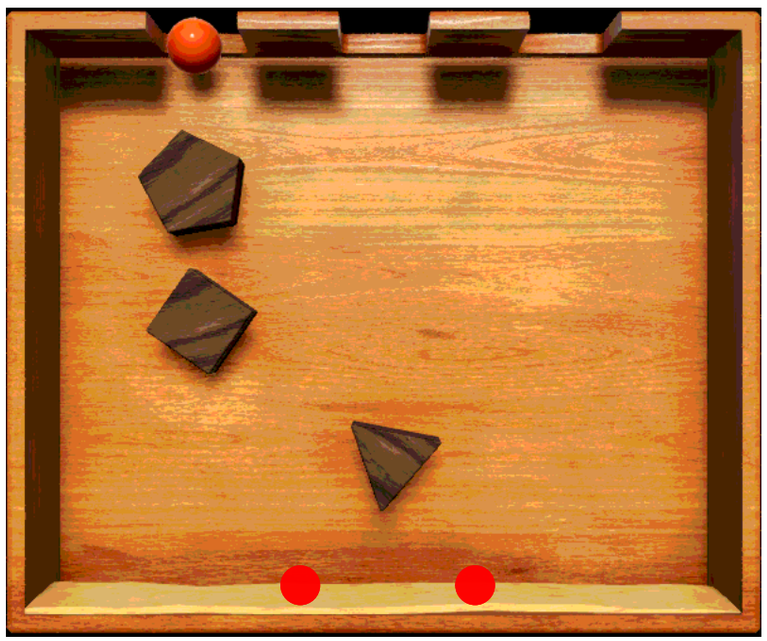
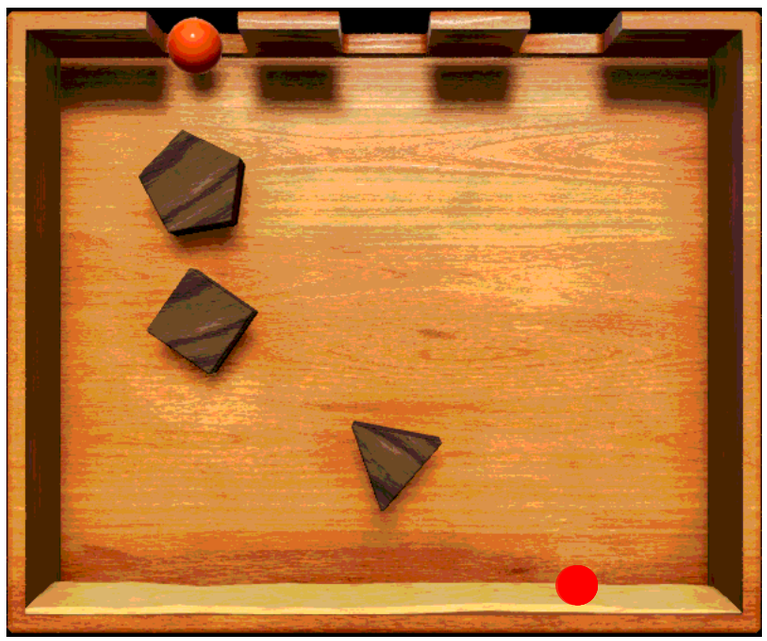
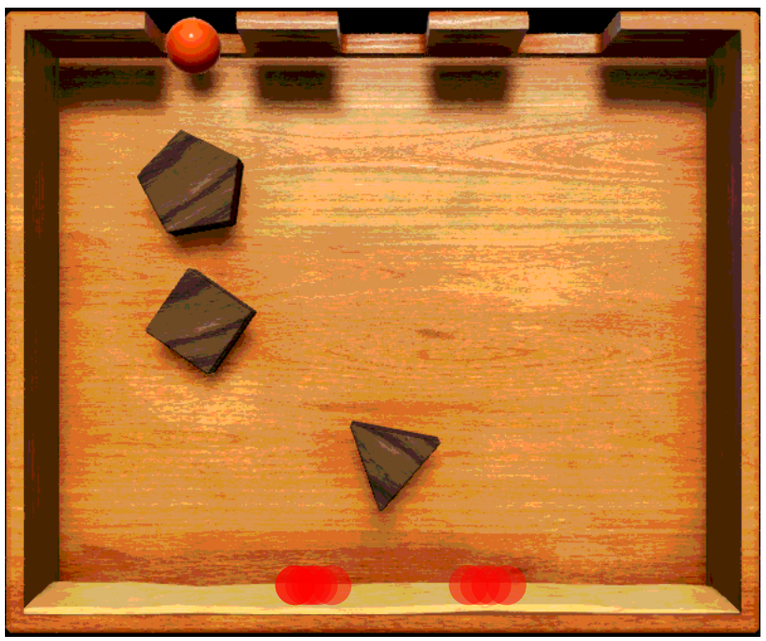
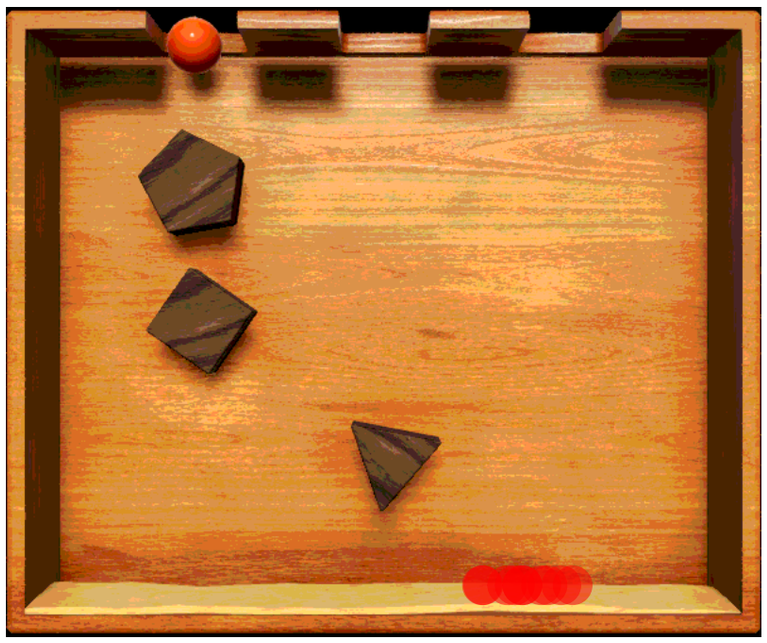
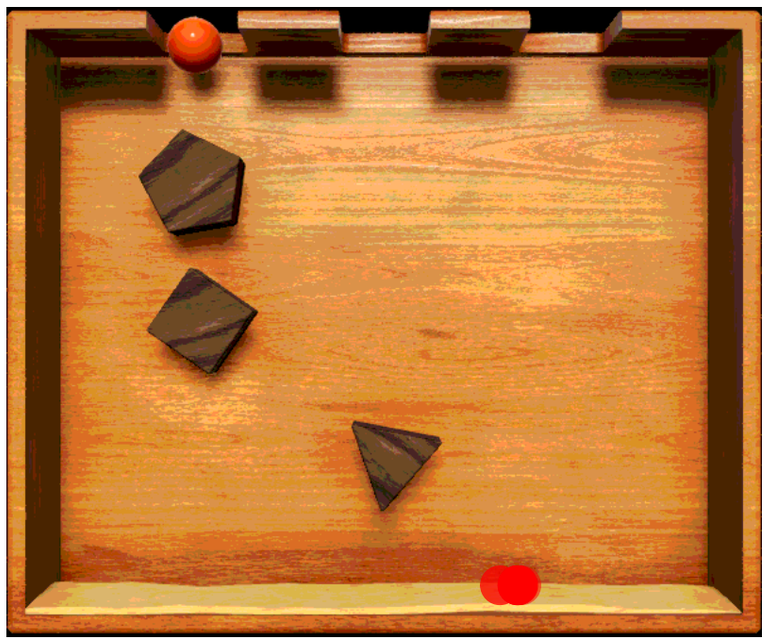
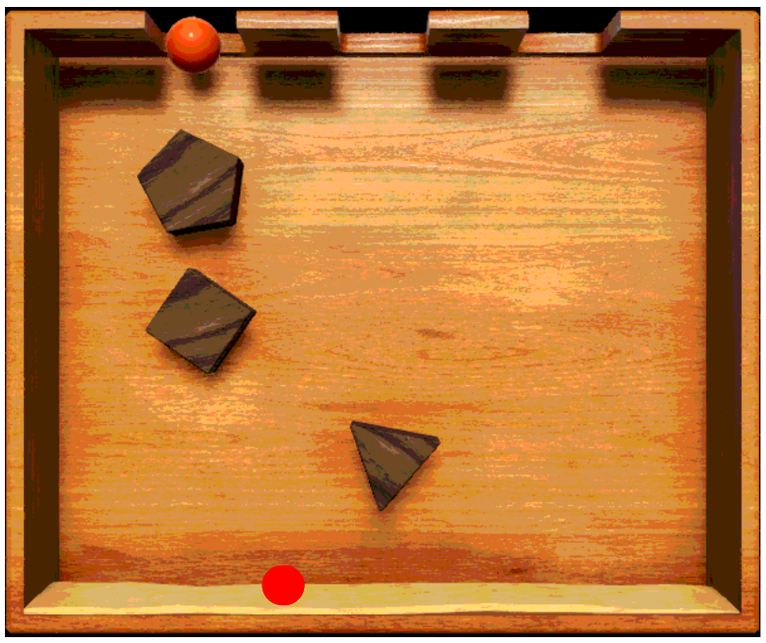
Prediction: Where will the ball land?



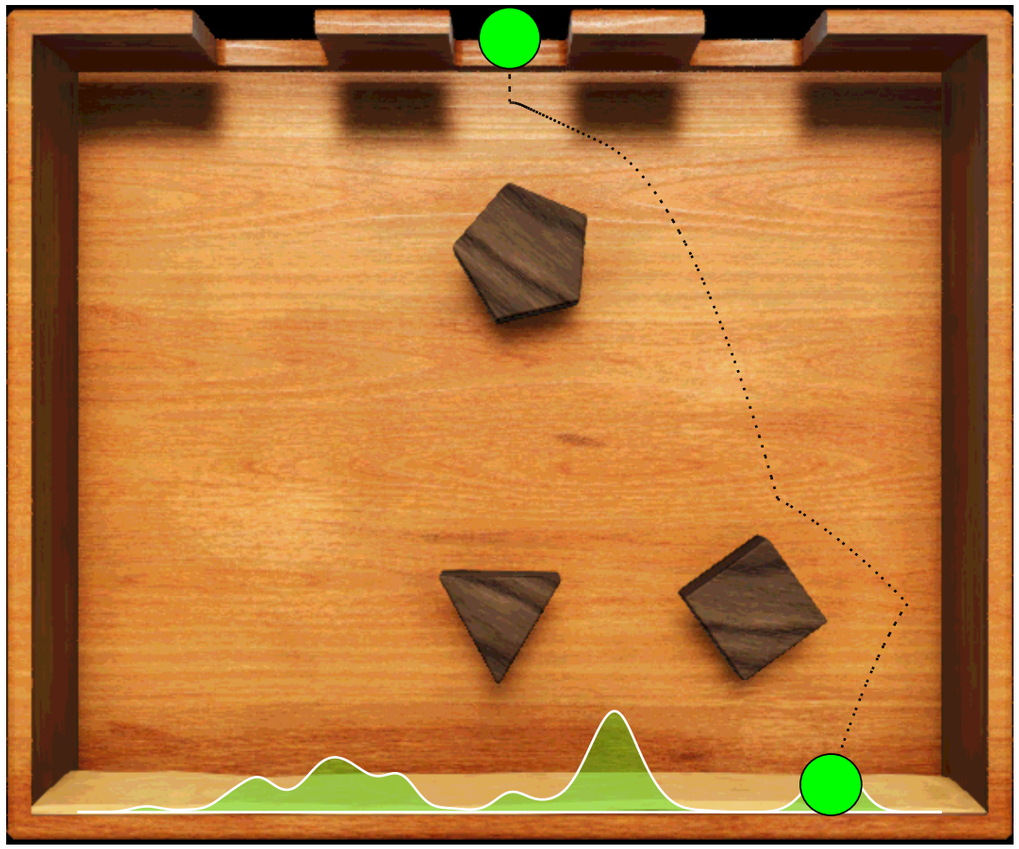
Prediction: Where will the ball land?



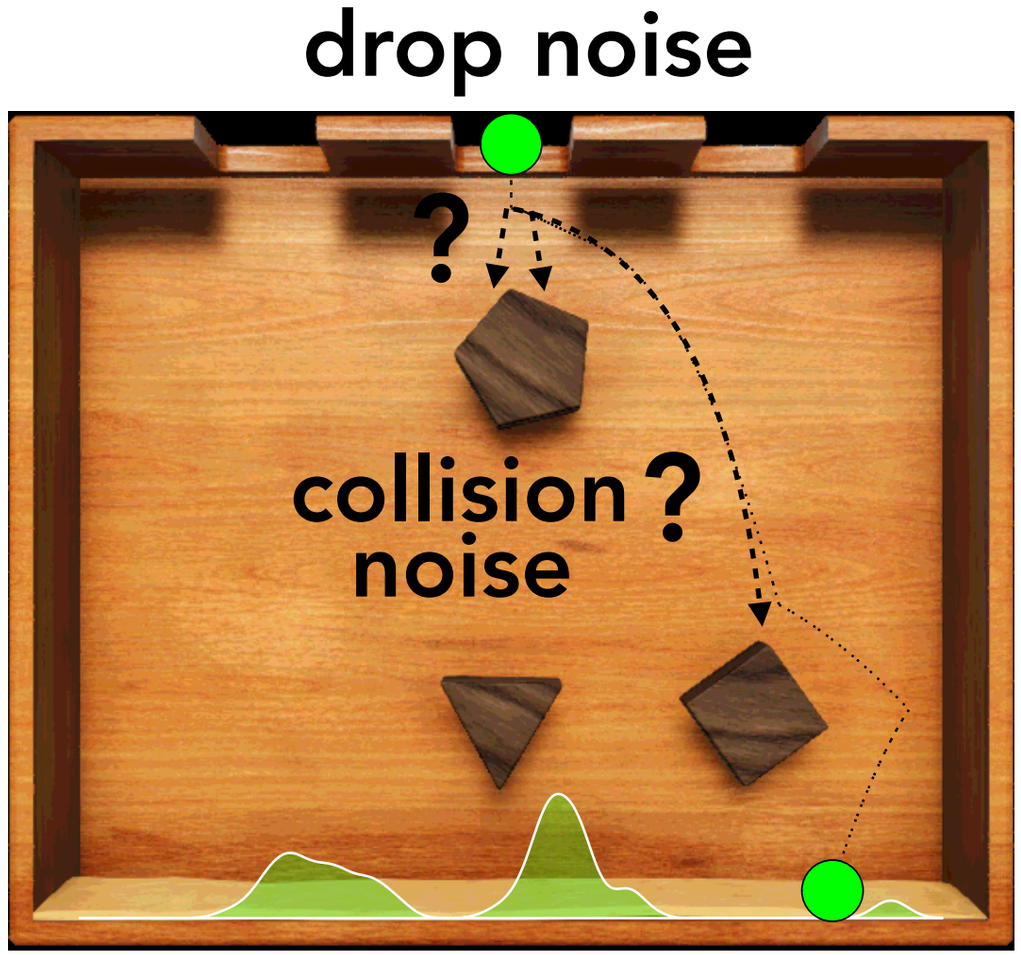
Prediction: Where will the ball land?

Prediction: Where will the ball land?



people

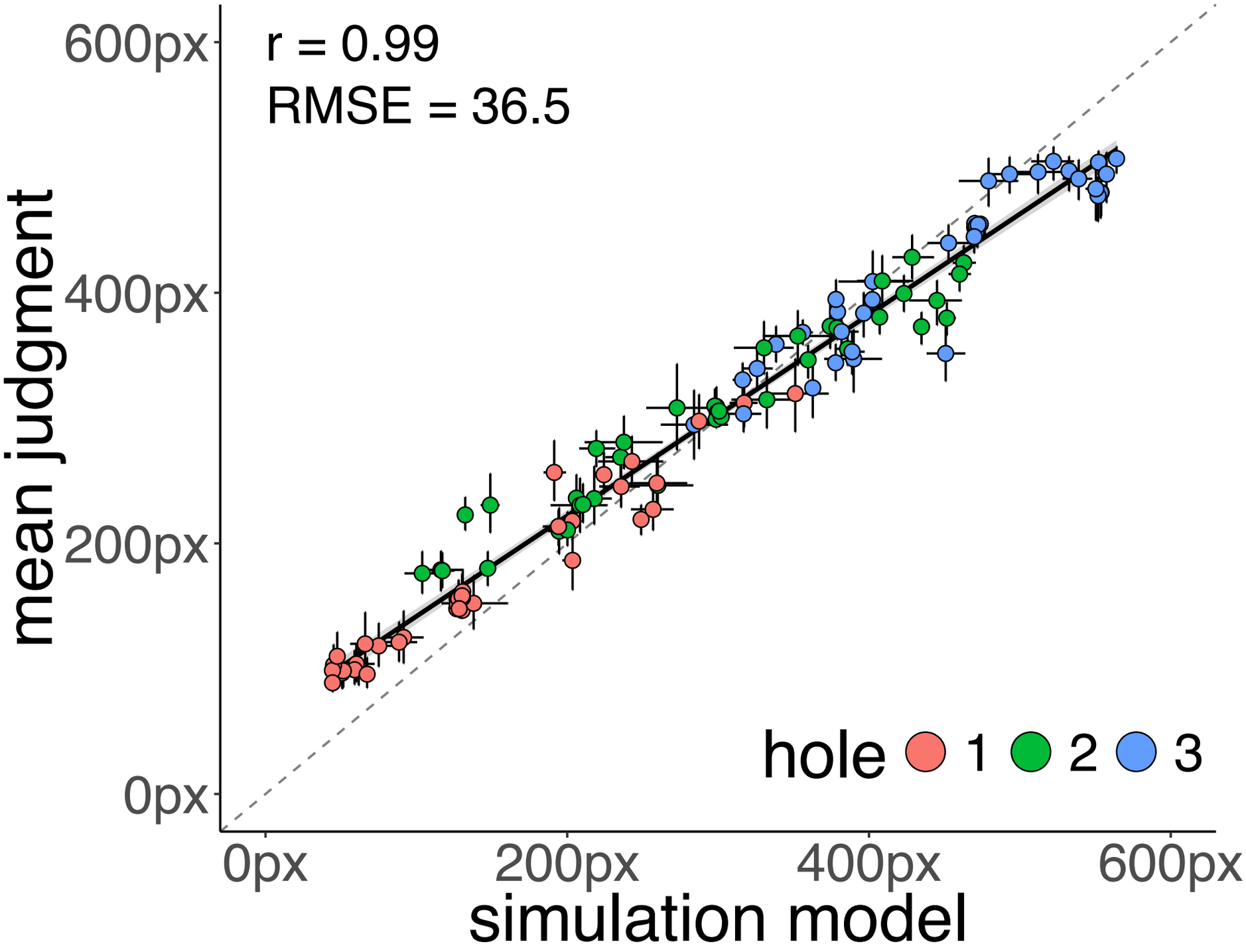


model

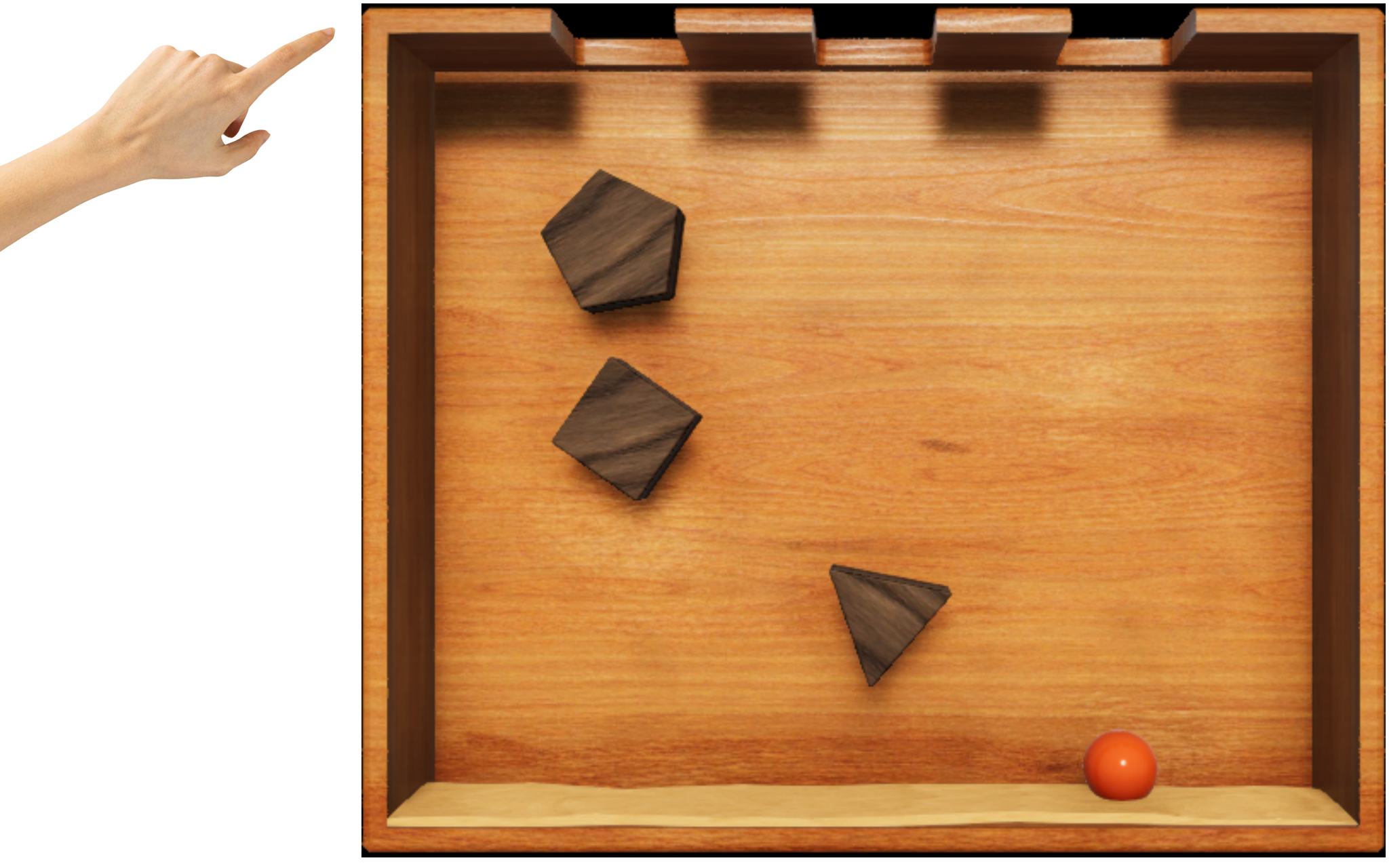
Ullman, Spelke, Battaglia, & Tenenbaum (2017) Mind Games: Game Engines as an Architecture for Intuitive Physics. *Trends in Cognitive Sciences*

Smith & Vul (2013) Sources of uncertainty in intuitive physics. *Topics in Cognitive Science*

Prediction: Where will the ball land?



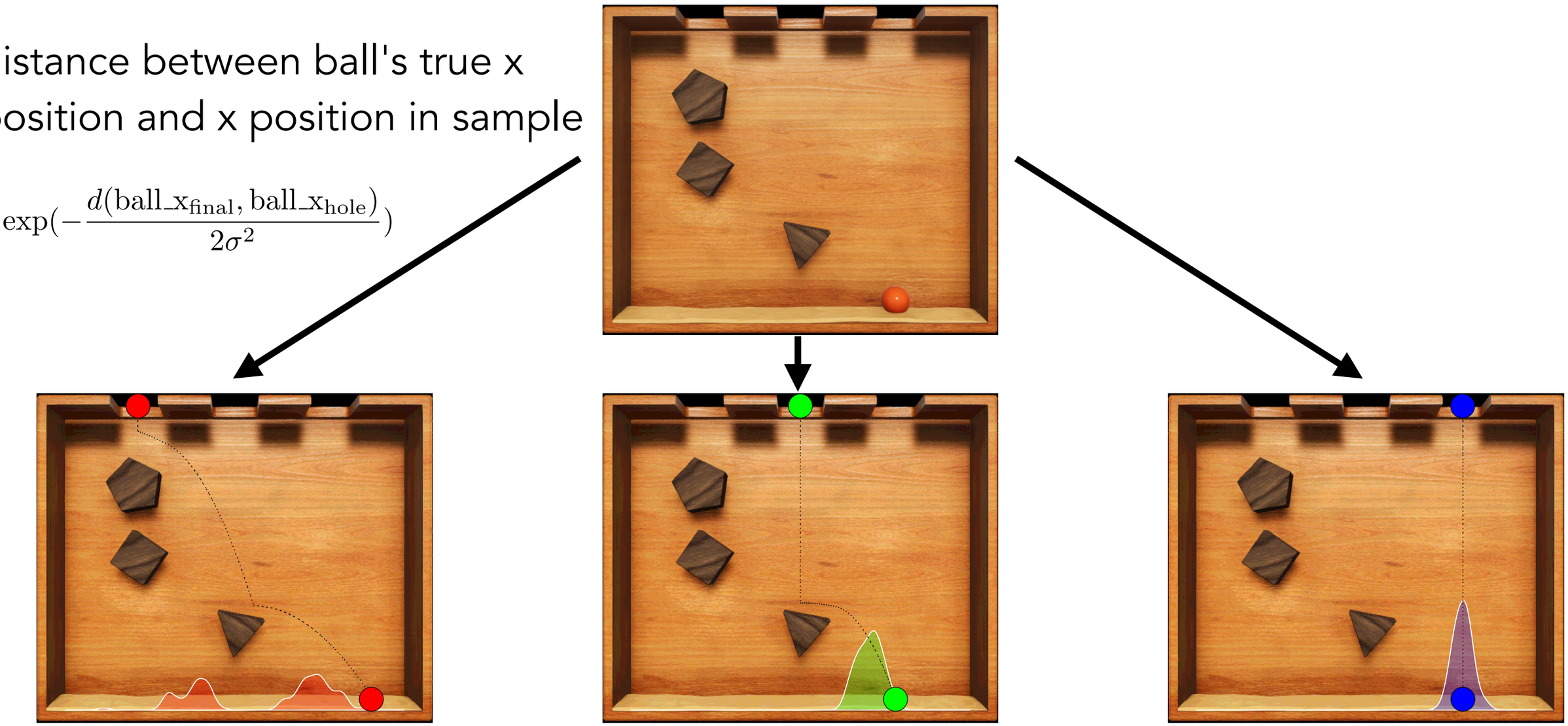
Inference: In which hole was the ball dropped?

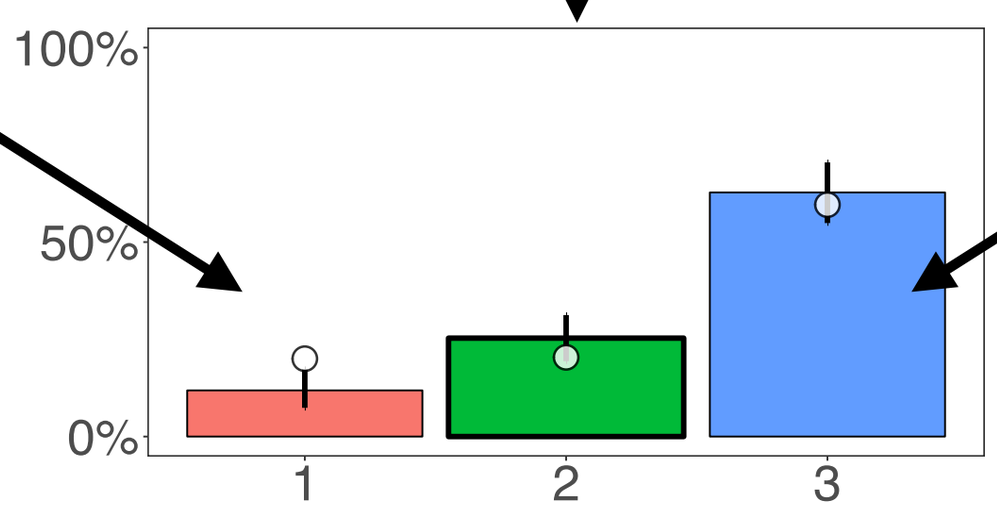


Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

$$\exp(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2})$$



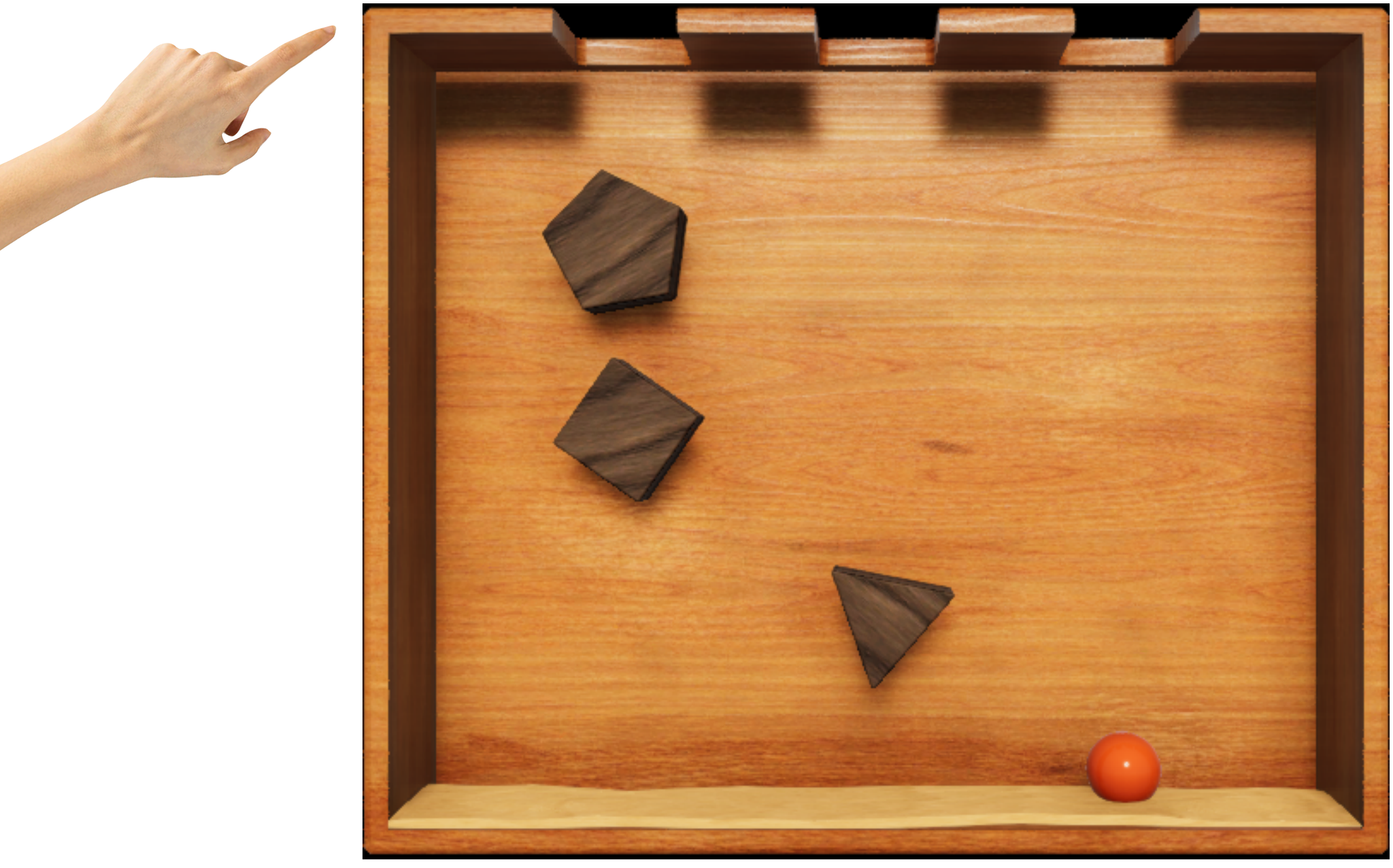


data

model prediction

Inference: In which hole was the ball dropped?

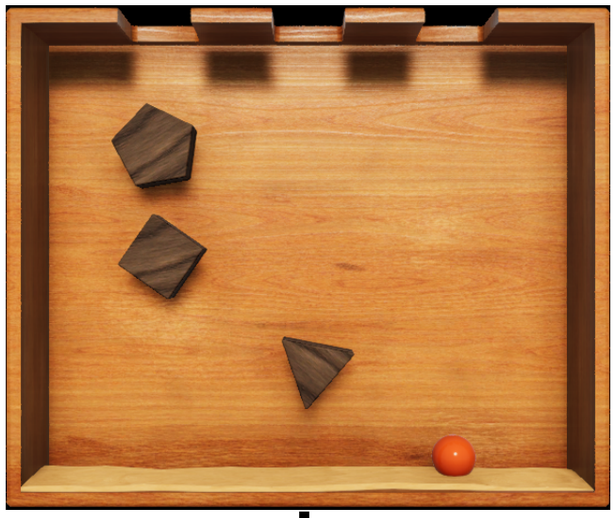
Inference: In which hole was the ball dropped?

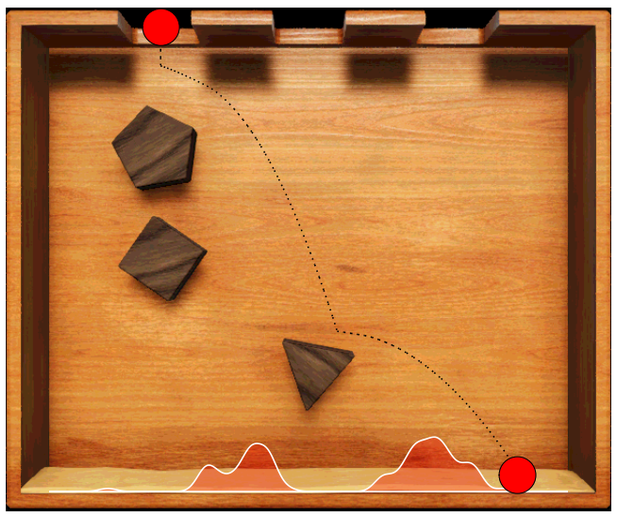


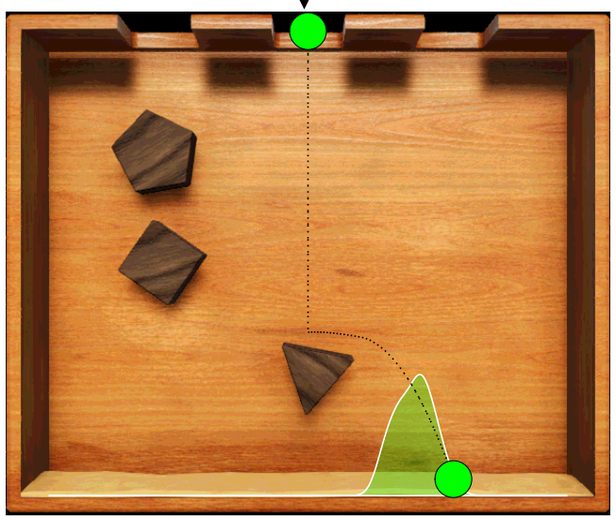
Inference: In which hole was the ball dropped?

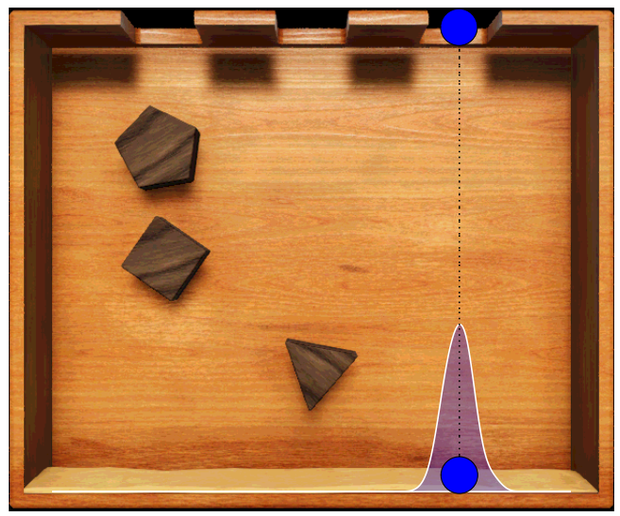
distance between ball's true x position and x position in sample

$$\exp(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2})$$

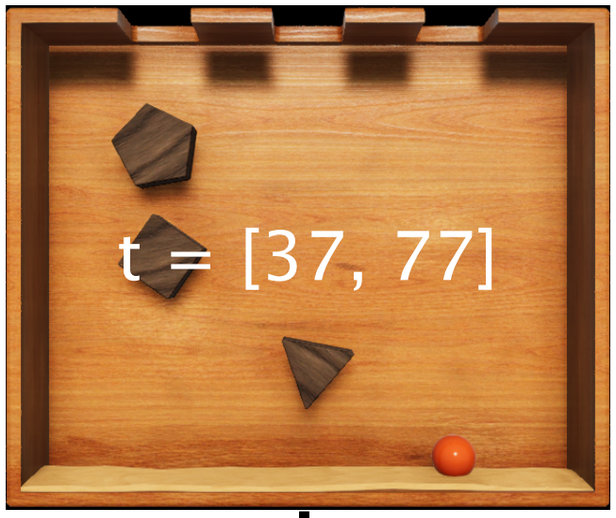






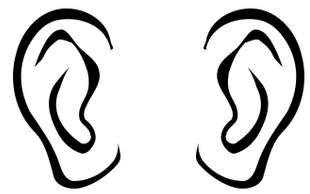


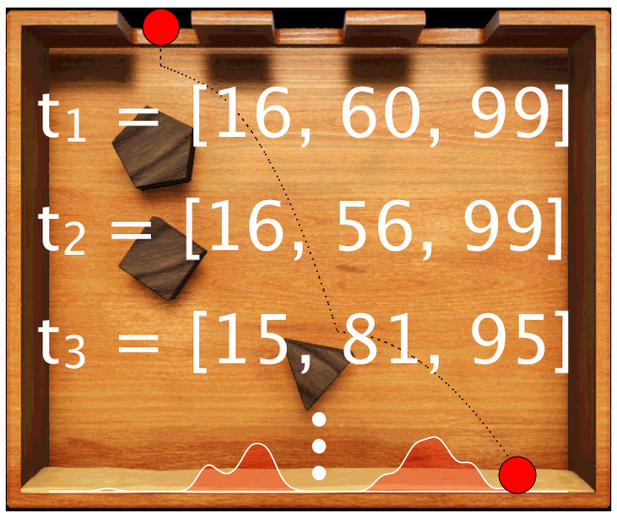
Inference: In which hole was the ball dropped?

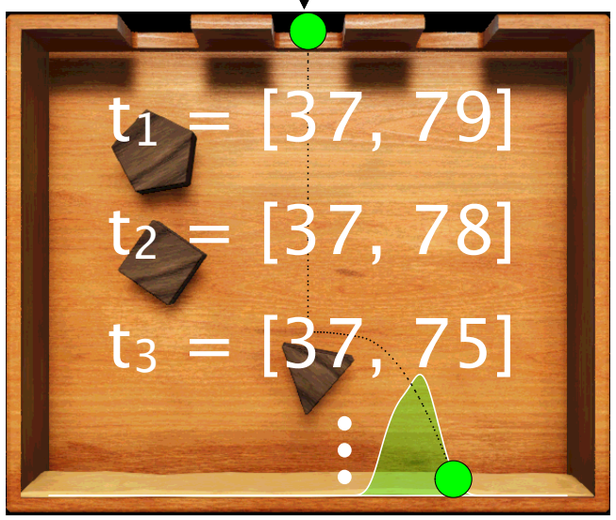


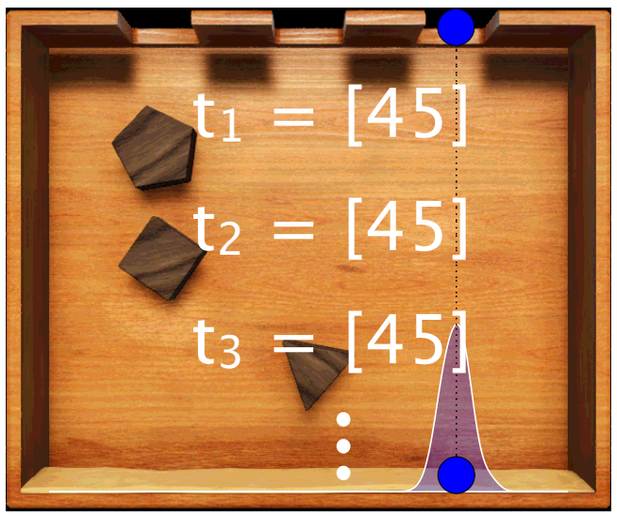
average temporal distance between time points

$$\frac{\sum_i^N \exp(-\frac{d(\text{sound\_true}_i, \text{sound\_simulation}_i)}{2\sigma^2})}{N}$$









t = [37, 77]

+ penalty

t_1 = [16, 60, 99]

t = [37, 77]

t_1 = [37, 79]

t = [37, 77]

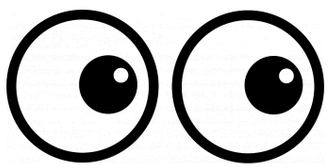
+ penalty

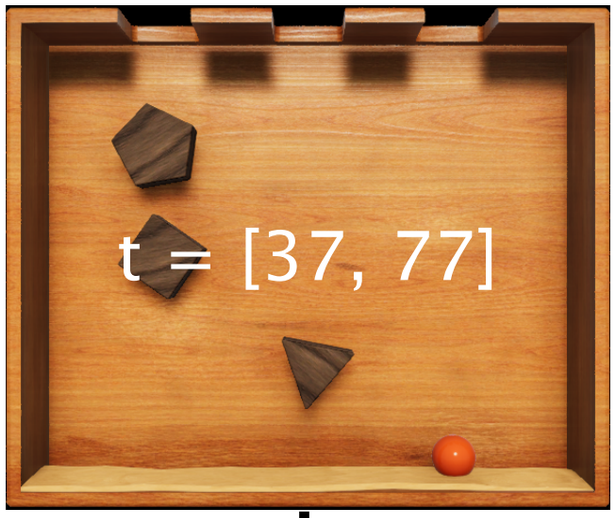
t_1 = [45]

Inference: In which hole was the ball dropped?

distance between ball's true x position and x position in sample

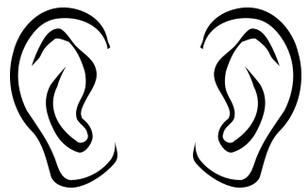
$$\exp(-\frac{d(\text{ball\_x}_{\text{final}}, \text{ball\_x}_{\text{hole}})}{2\sigma^2})$$

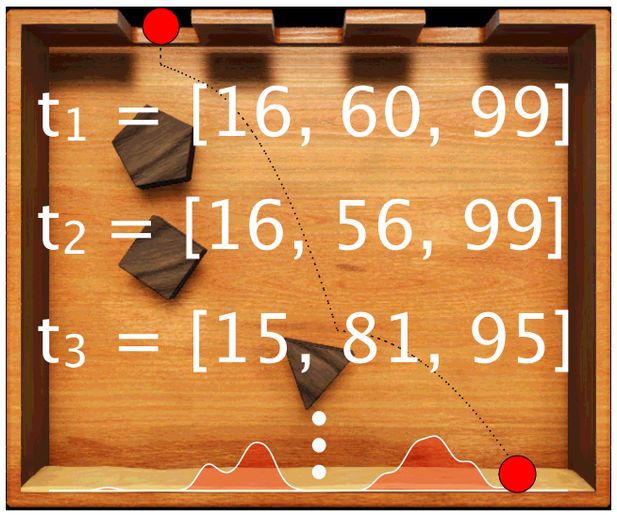


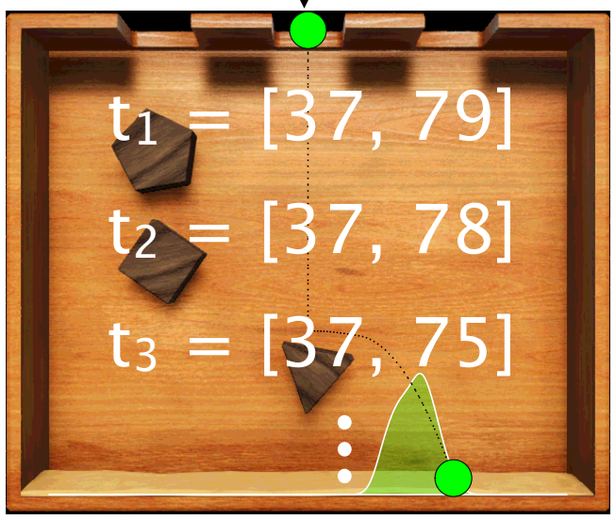


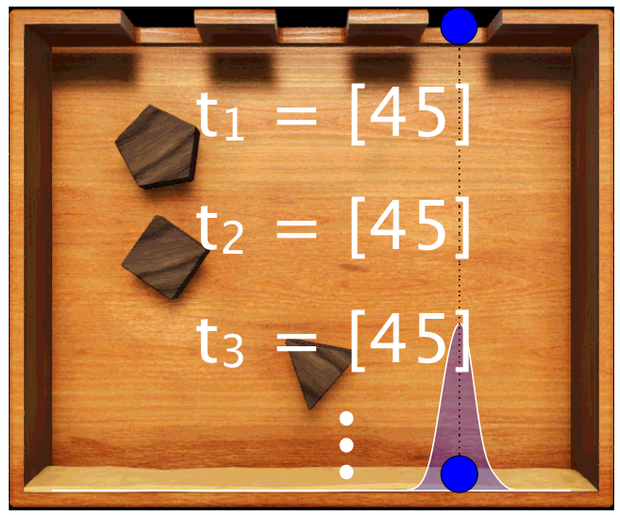
average temporal distance between time points

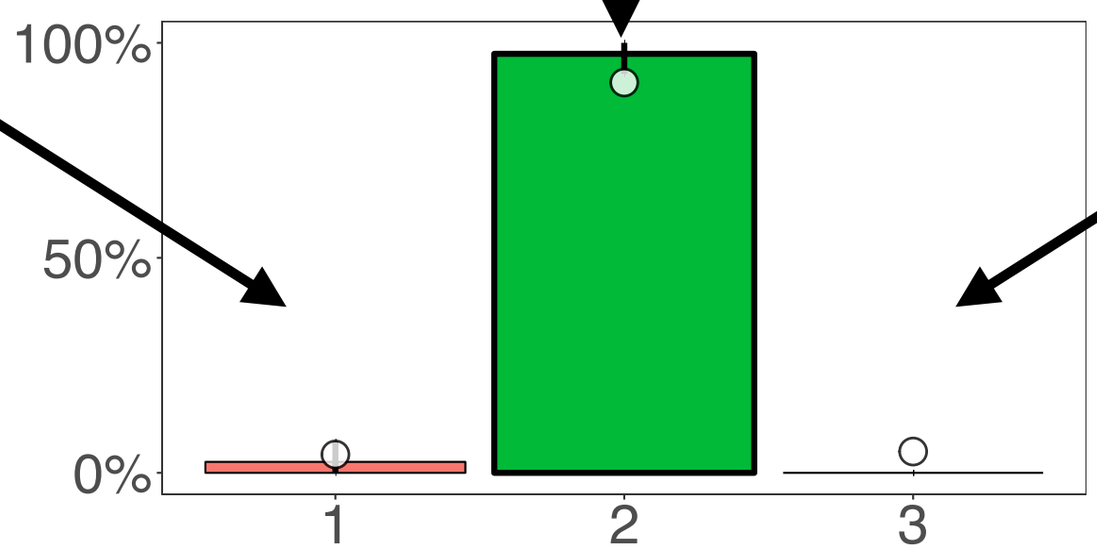
$$\frac{\sum_i^N \exp(-\frac{d(\text{sound\_true}_i, \text{sound\_simulation}_i)}{2\sigma^2})}{N}$$



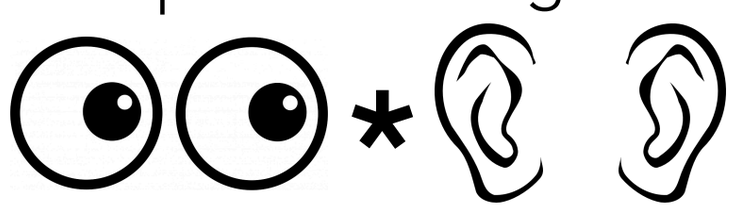




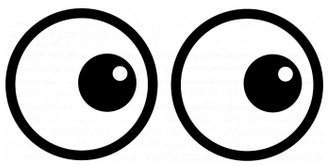


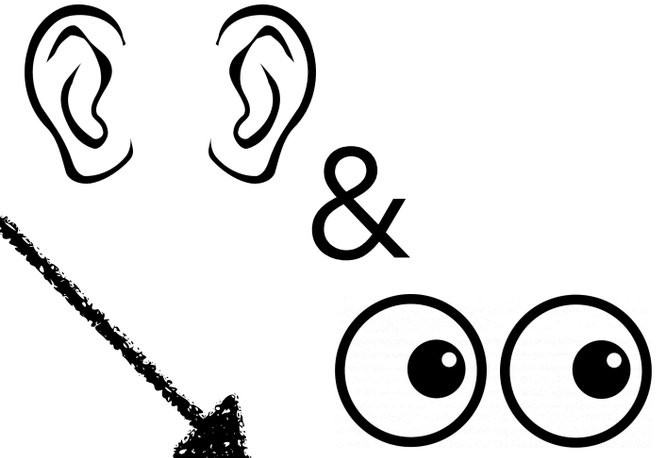


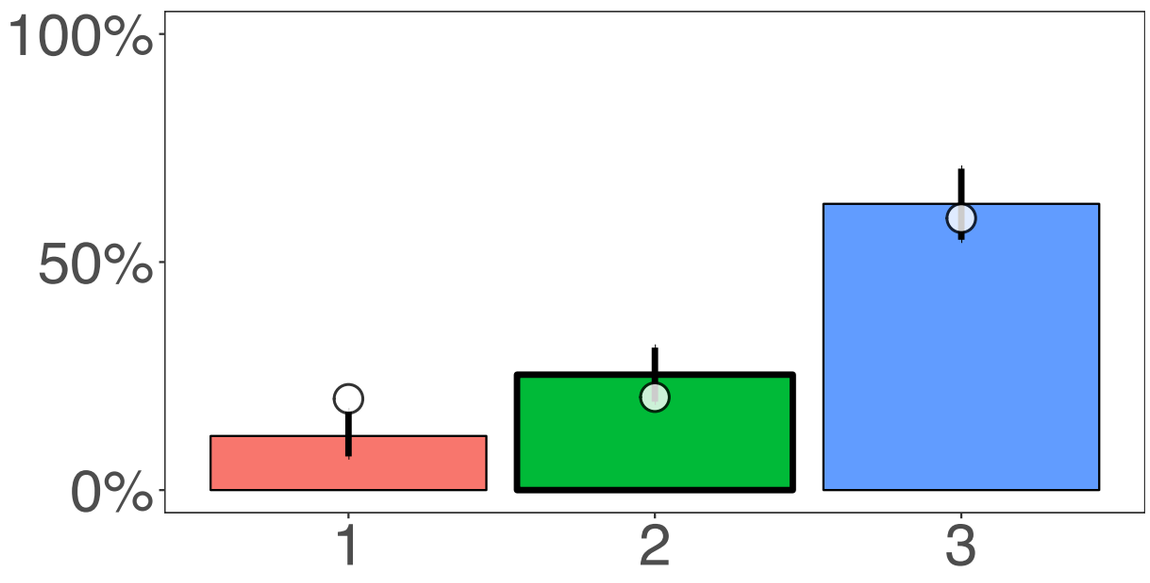
multiplicative integration

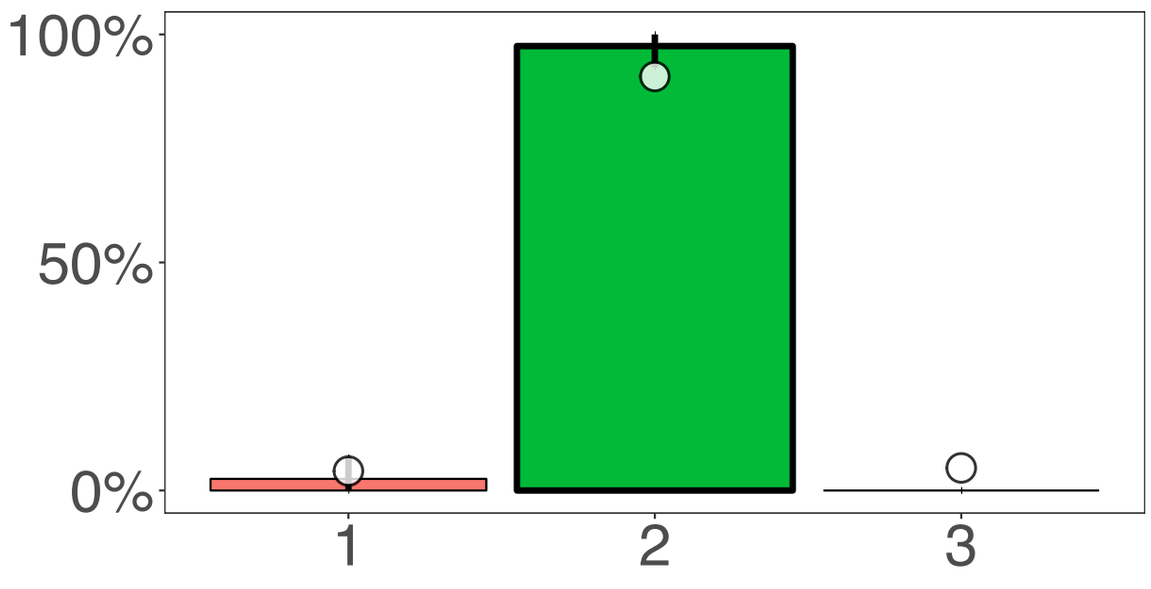


Inference: In which hole was the ball dropped?









Conclusion

- we build **rich** mental models of the world
- we simulate these models to:
  - predict the future
  - infer the past
  - evaluate counterfactuals
- together, these capabilities allow us to understand **why** something happened

Gerstenberg & Tenenbaum (2017) Intuitive Theories. Oxford Handbook of Causal Reasoning

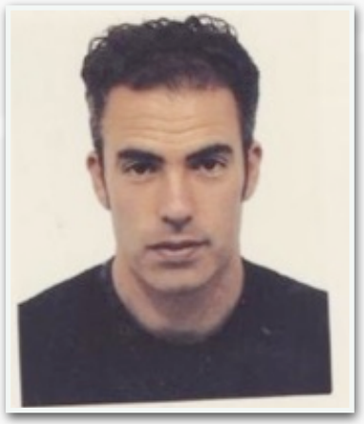
Goodman, Tenenbaum, & Gerstenberg (2015) Concepts in a probabilistic language of thought. The Conceptual Mind: New Directions in the Study of Concepts

Lake, Ullman, Tenenbaum, & Gershman (2016) Building machines that learn and think like people. Behavioral and Brain Sciences

C i C Causality in Cognition

Thanks!

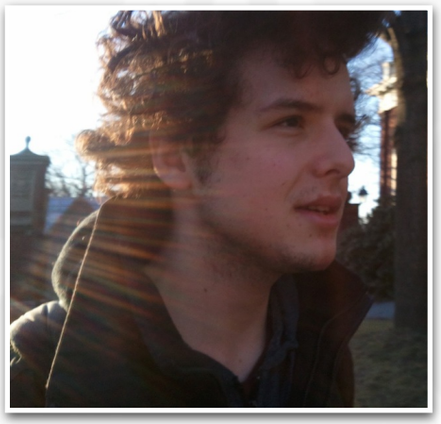
Josh Tenenbaum



David Lagnado

Noah Goodman

Matt Peterson



Max Siegel

Gerstenberg, Peterson, Goodman, Lagnado, & Tenenbaum (2017) Eye-tracking causality. Psychological Science

Goodman, N. D., Tenenbaum, J. B., & Gerstenberg, T. (2015). Concepts in a probabilistic language of thought. In The Conceptual Mind: New Directions in the Study of Concepts. MIT Press.

Gerstenberg, T. & Tenenbaum, J. B. (2017). Intuitive Theories. In Oxford Handbook of Causal Reasoning. Oxford University Press.

@tobigerstenberg

http://cicl.stanford.edu/