ndg@mit.edu

Ideal observers in social cognition: A parable of rational analysis

Noah D. Goodman

Computational Cognitive Science Group,
MIT

Mind Brain Behavior Conference,
16 February, 2007

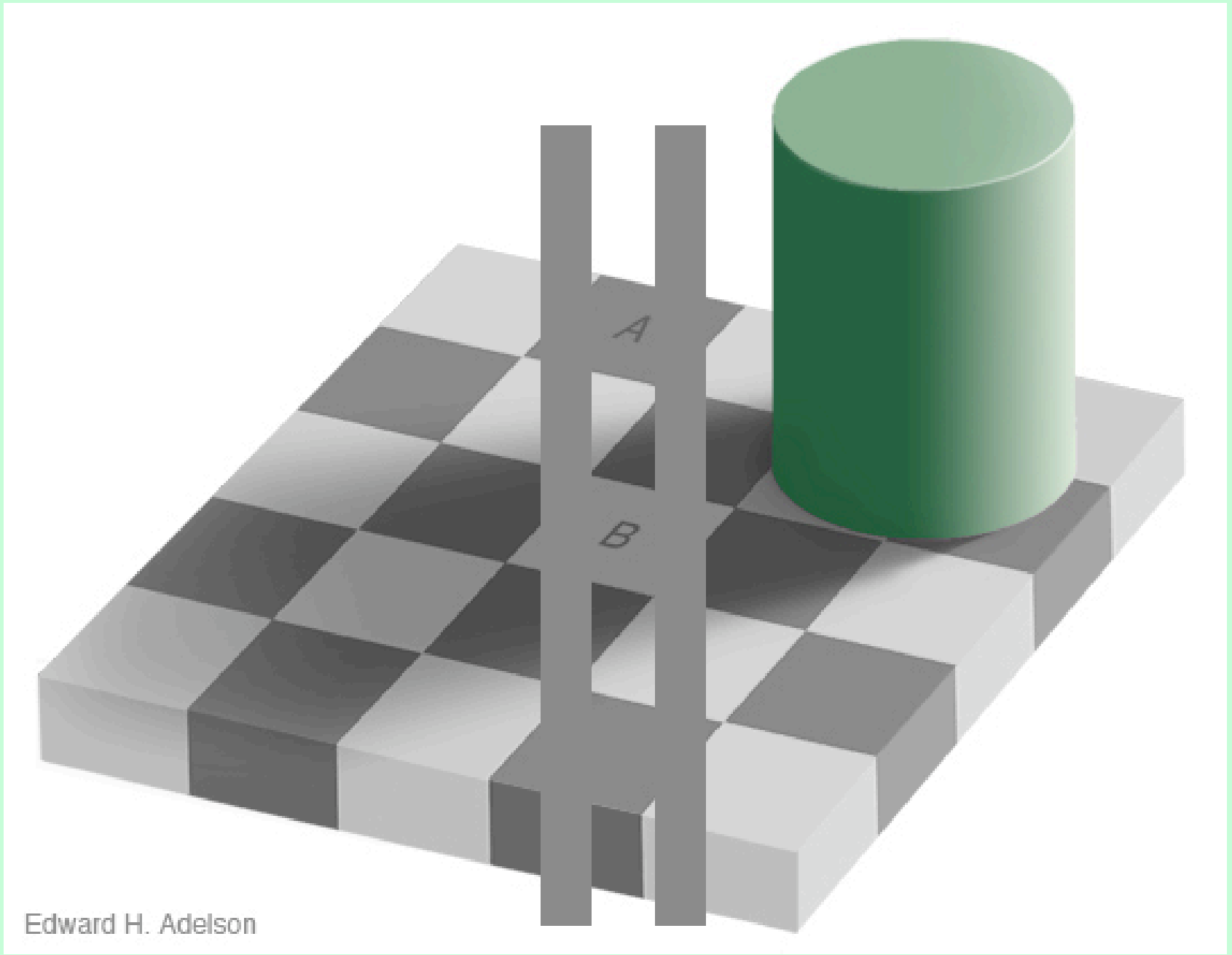
Thanks to:

- My coauthors:
  Chris Baker,
  Liz Baraff Bonawitz,
  Vikash Mansinghka,
  Alison Gopnik,
  Henry Wellman,
  Laura Schulz,
  Josh Tenenbaum.
- Rebecca Saxe.
- This research was supported by the J. S. McDonnell Foundation Causal Learning Collaborative Initiative

Tammy is always polite to her coworkers and goes out of her way to do nice things for them.

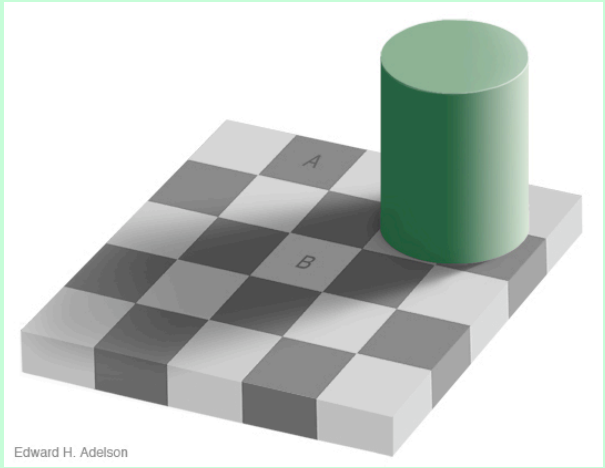
Last month Tammy made coffee for her boss. She put lots of milk in the coffee, just the way he likes it. There had been a power outage the night before, though, and the milk had gone bad. Tammy's boss got food poisoning and was sick for a week.

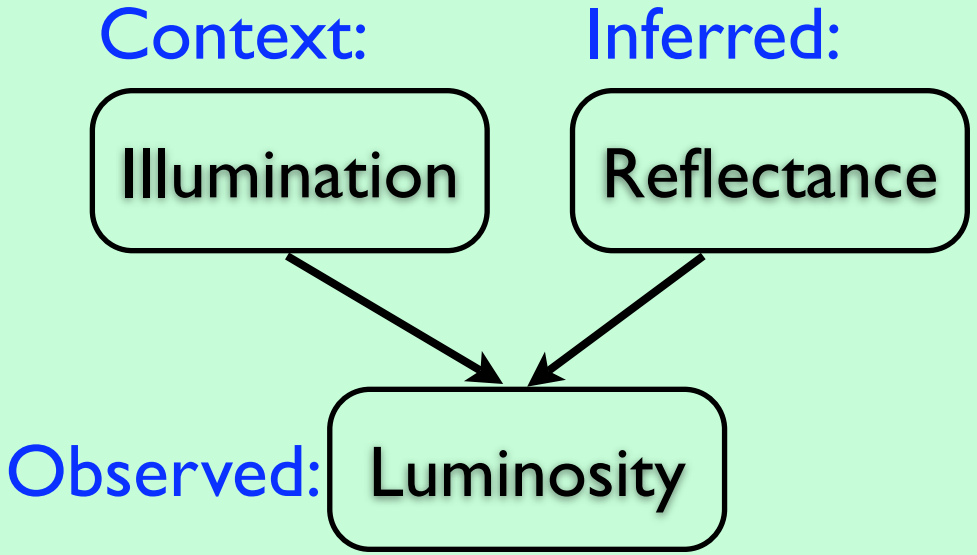
In her journal Tammy wrote, “Well, he didn’t die -- maybe next time I’ll use poison!”


A
B
Edward H. Adelson

Under-constrained inverse problems

Perception:



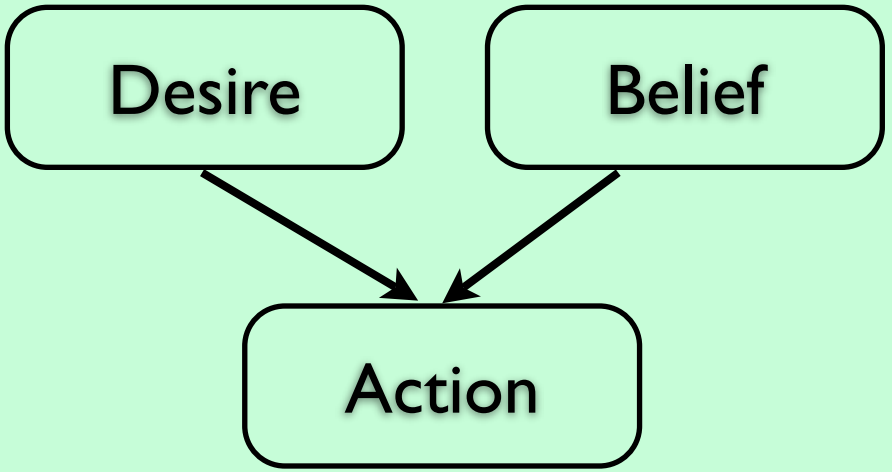


Social cognition:

"Tammy is kind..."

"Tammy made coffee for her boss with spoiled milk..."

Did Tammy believe the milk was spoiled?



Rational analysis

- Thesis: cognition is an adaptive response -- an approximately optimal solution to computational problems posed by the environment. (See Shepard, Marr, Anderson, Tenenbaum, etc.)
  - In perception, one compares human performance to the Ideal Observer -- this analysis has been very successful at predicting and explaining human perception (e.g. the checkershadow illusion). (See Adelson, Geisler.)
- Is it coherent/useful to analyze social cognition as ideal? (See Kelley.)

Questions about Ideal Social Observers

- Would an ideal social observer attribute hidden mental states to other agents?
- What would an ideal social observer believe about the mental states of other agents? How would attributions depend on evidence?
- Learning: if it's so ideal, why does it change?

Questions about Ideal Social Observers

- Would an ideal social observer attribute hidden mental states to other agents?
- What would an ideal social observer believe about the mental states of other agents? How would attributions depend on evidence?
- Learning: if it's so ideal, why does it change?

False belief: the Sally-Anne task

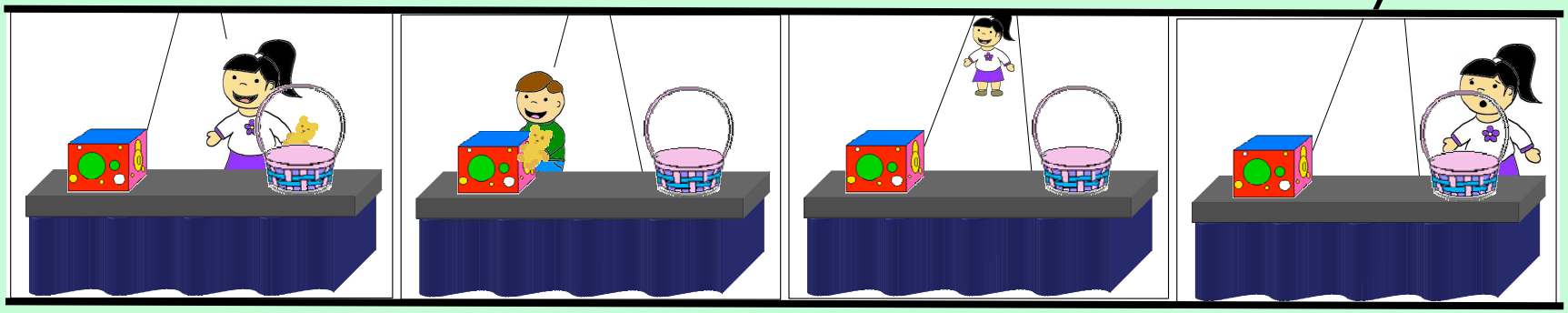
(Wimmer and Perner, 1983)

Sally puts her toy in the basket...

Anne moves it to the box...

Sally returns...

Where will sally look for her toy?



Sally Believes her toy is in the basket.

But she's wrong.

This is analogous to the "Tammy poisons her boss" example.

False belief

- There is ample* evidence that children undergo a *false belief transition* at age 3-4yrs: before this change they incorrectly expect Sally to go where her toy actually is.
- If adult false-belief attributions are ideal, why do children initially fail to make them?

*Ample but not uncontested.

Strong and weak optimality in cognitive development

- There are two positions consistent with an Ideal Social Observer analysis of adult social cognition:
  - Weak: adults are ideal, children aren't. That is, only in the end state of development is social cognition ideal.
  - Strong: children make ideal predictions with respect to their *best guess* about the structure of other people, and this structure ('theory') is rationally updated in response to accumulating evidence.

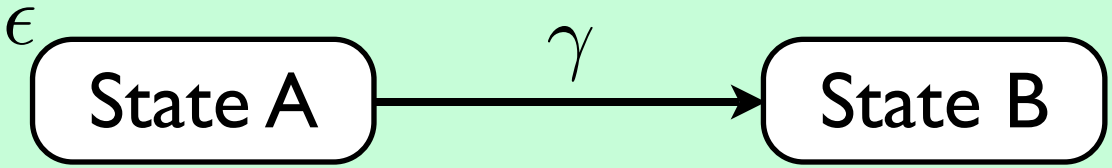
Strong optimality for false belief

- Could a delayed false-belief transition ever result from rational updating of a child's 'theory of people'?
- If so, do children make inferences that are ideal with respect to these best guess theories, even before the false-belief transition?
- We can begin to answer these questions by examining two simple 'theories' in the Sally-Anne setting...

Causal Bayes' net models

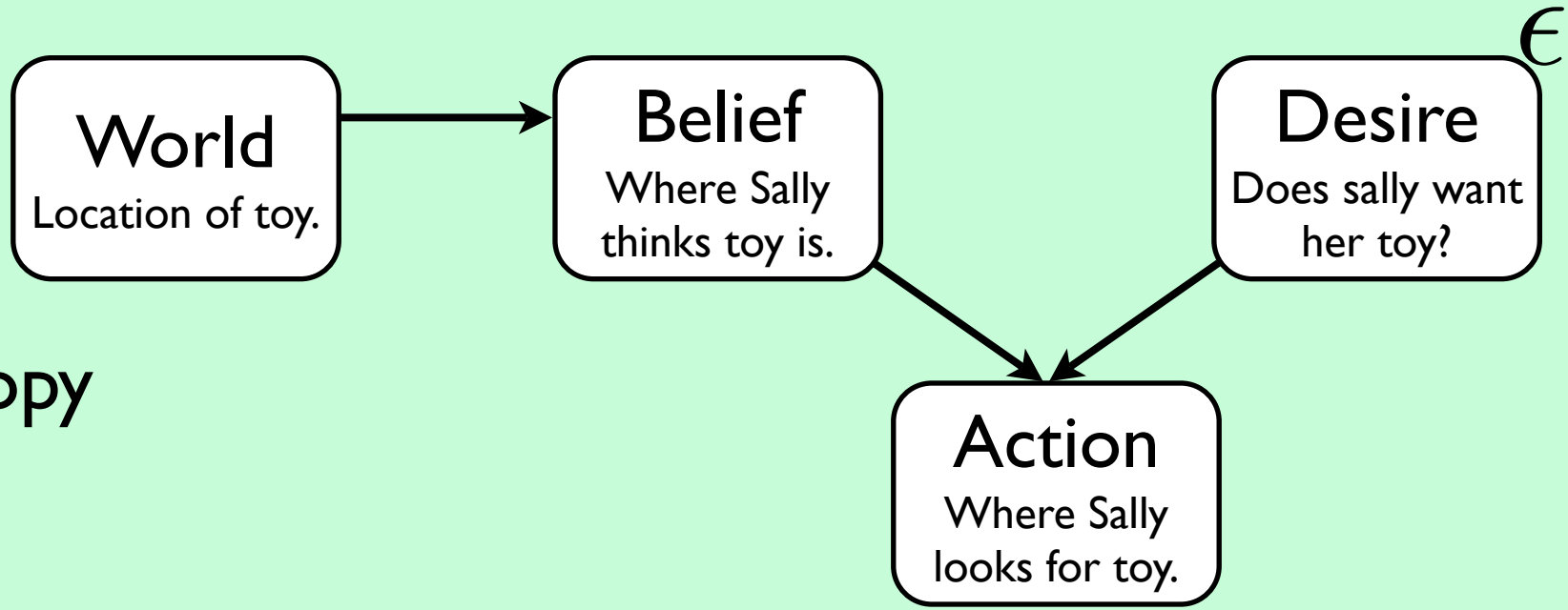
- We represent each 'theory' by a causal Bayesian network: a graphical representation of probabilistic causal structure.
- Ideal use of a theory can then be described by Bayesian inference, and
- The optimal degree of belief, and rational updating, is prescribed by Bayesian model selection methods.

Notation:

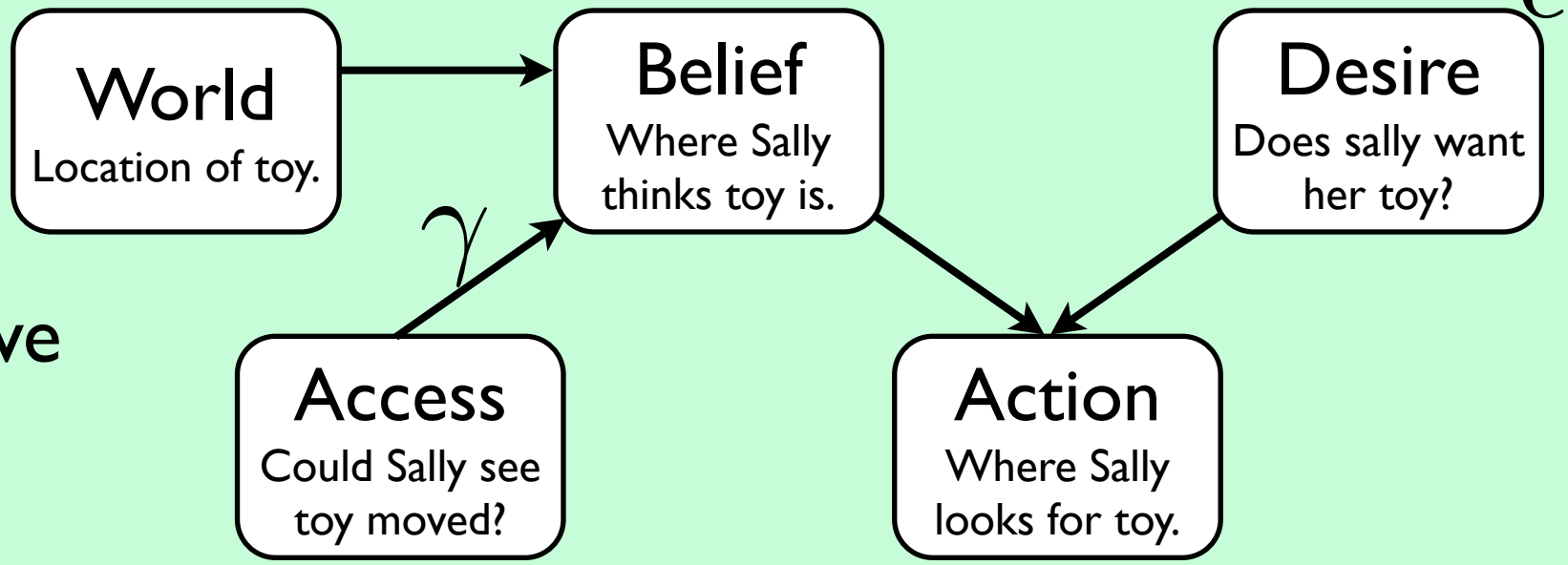


Two simple Bayes' net models

CT:
(Simpler "Copy Theory")



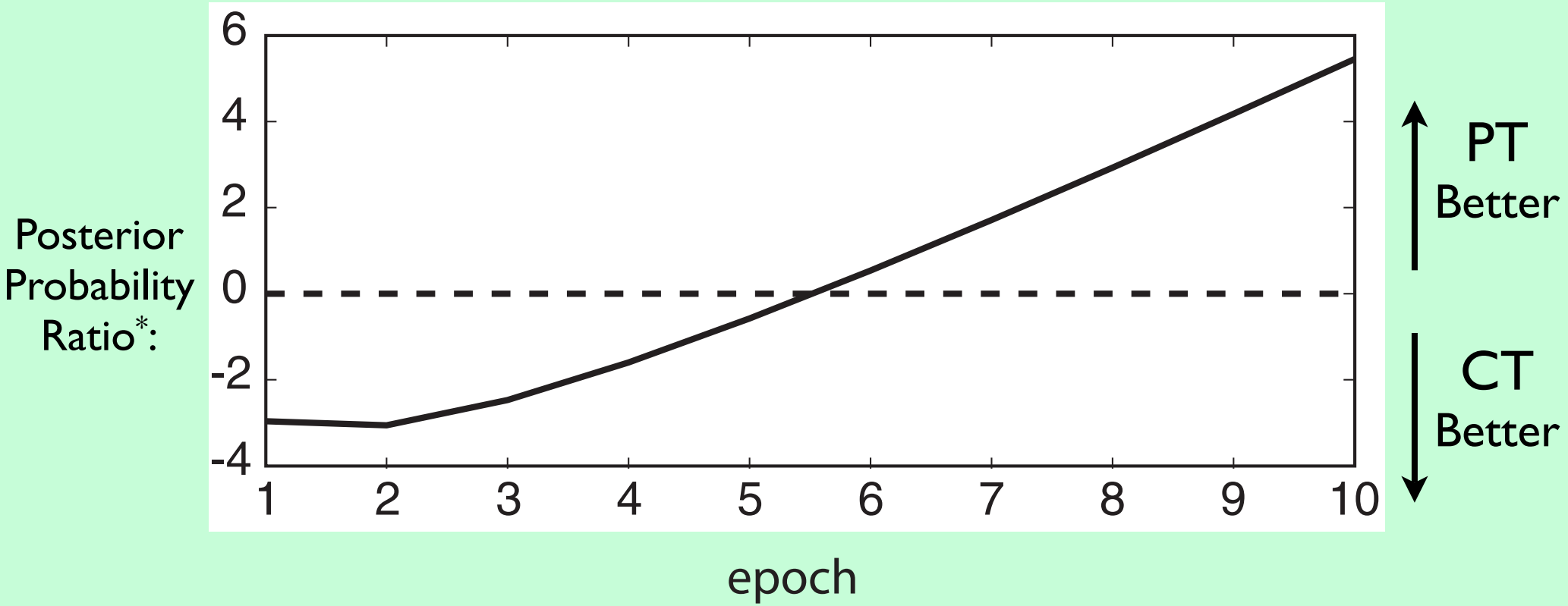
PT:
("Perspective Theory")



- Could a delayed false-belief transition ever result from rational updating of a child's 'theory of people'?
  - Specifically, is it ever optimal to maintain the earlier (CT) theory, in the face of counter-evidence, but to eventually switch to the later (PT) theory?

Model results: rational revision

- Yes!



* $W_{\mathrm{PT/CT}} = -\log(P(\mathrm{PT}|X)/P(\mathrm{CT}|X))$,

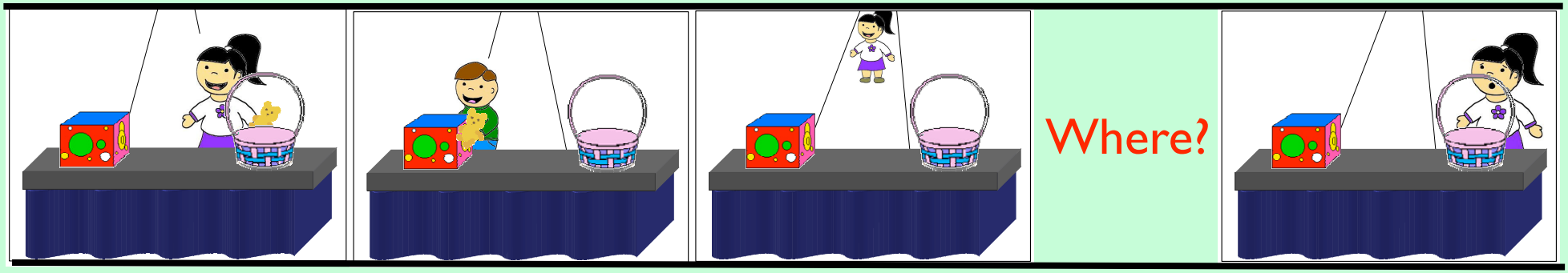
Rational revision

- Why this delayed transition?
  - The CT theory is simpler, hence *a priori* more likely*,
  - but the PT theory is more flexible: it is able to explain surprising evidence by *false belief* explanations as well as *unexpected desire* explanations.
  - This is a characteristic tradeoff of Bayesian model selection.

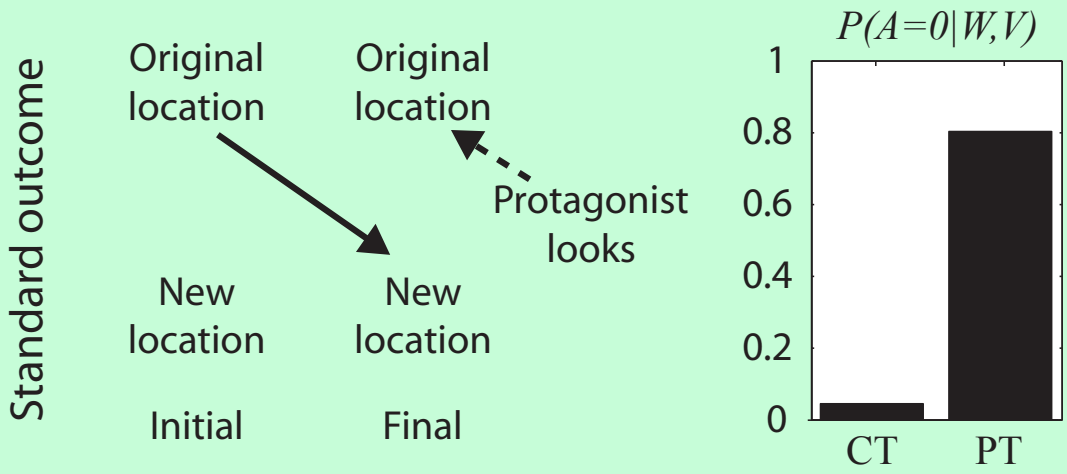
*This simplicity bias is realized by a Bayesian Occam's razor.

- Do children make inferences that are ideal with respect to their 'best guess' theory, even before the false-belief transition?
  - In particular, do children explain unexpected outcomes in ways consistent with their theory (CT or PT)?

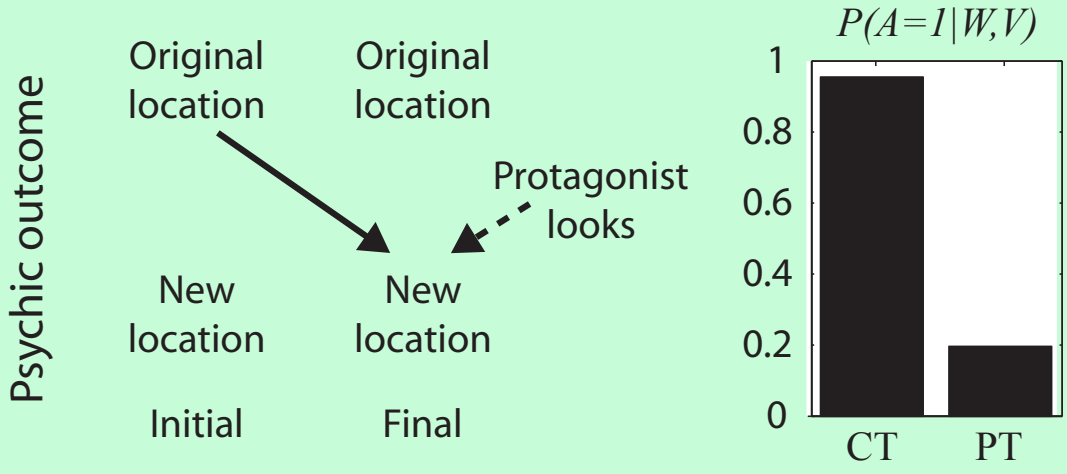
Two outcome conditions



- ‘Standard’ outcome is surprising to CT:



- ‘Psychic’ outcome is surprising to PT:



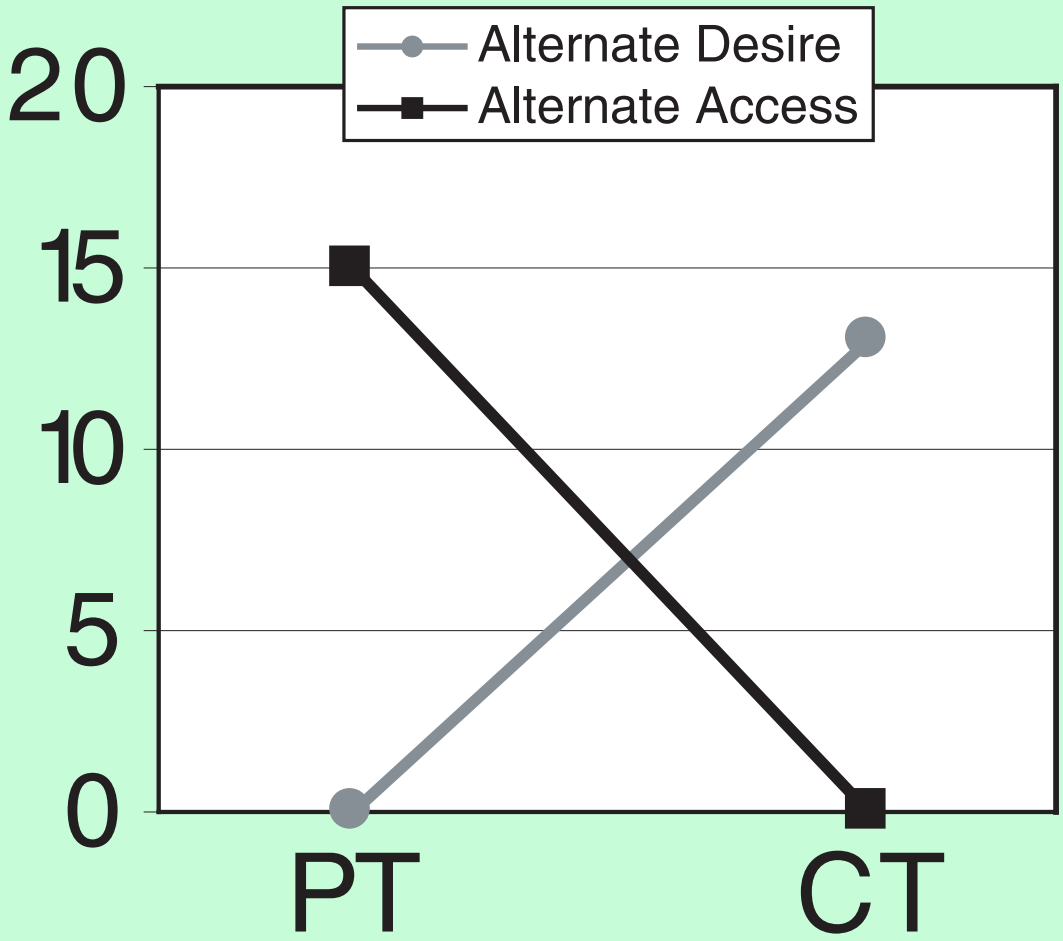
Experiment

- N=49 children (R=3;0-4;11, M=3;11).
- Two 'Sally-Anne'-style storybooks (Standard and Psychic), order counterbalanced.
- Children were asked to predict where Sally would look for her toy. After seeing where she did look, they were asked for an explanation.

Experimental results

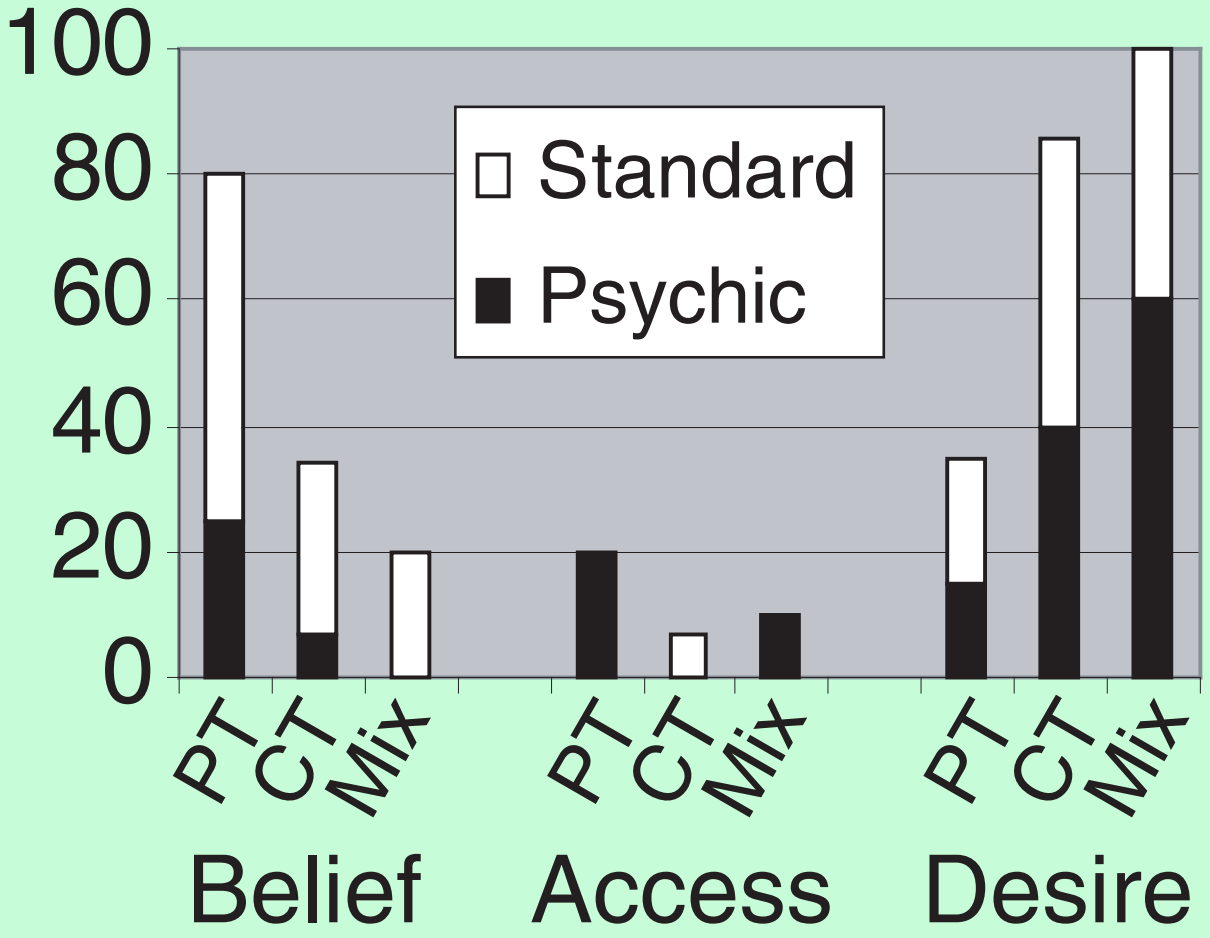
- Three groups of children were identified based on their predictions: 15 CT, 20 PT, and 10 Mixed predictors. (4 failed memory test.)
- There was a strong interaction between prediction type and order of storybooks:
  - This is predicted by the model.
  - May be a marker of children near the false belief transition.

Children's Explanations



- The two groups appeal to different explanations of surprising outcomes:
  - CT appeal to unusual desires. (E.g. "she wanted to move the basket.")
  - PT appeal to unusual belief/access. ("She heard her brother move it.")
    - Note the toy was very desirable....

Children's Explanations



- The overall pattern of explanations was different between groups, and consistent with predictions from respective 'theories'.

Conclusion

- An 'ideal use and ideal learning' analysis of social cognition is consistent with the false-belief transition and with children's explanations, both before and after the transition.
- Ideal Social Observer analysis offers a useful vantage point on social cognition and its development.
- It remains to understand the Ideal Social Observer in richer and more flexible settings.