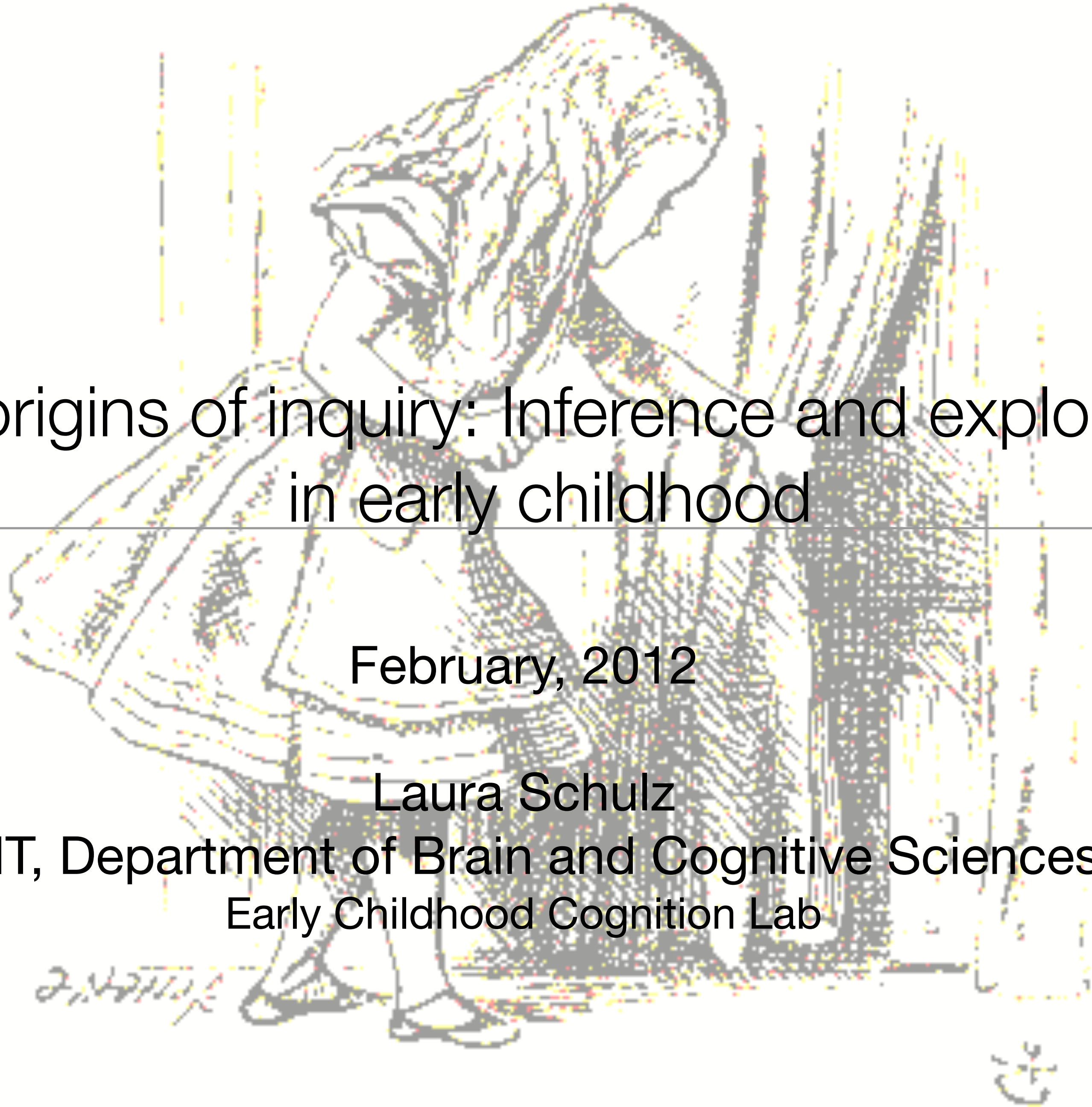


The origins of inquiry: Inference and exploration in early childhood

February, 2012

Laura Schulz
MIT, Department of Brain and Cognitive Sciences
Early Childhood Cognition Lab

Revolutions in our understanding of cognitive development

Developmental Psychology
1987, Vol. 23, No. 5, 655–664

0012-1649/87/$00.75

Object Permanence in 3 ½- and 4 ½-Month-Old Infants

Renée Baillargeon
University of Illinois

Revolutions in our understanding of cognitive development

Psychological Review
1992, Vol. 99, No. 4, 605–632

Copyright 1992 by the American Psychological Association, Inc.
0033-295X/92/$3.00

Origins of Knowledge

Elizabeth S. Spelke, Karen Breinlinger, Janet Macomber, and Kristen Jacobson
Cornell University

Revolutions in our understanding of cognitive development

Nature **358**, 749 - 750 (27 August 1992); doi:10.1038/358749a0

Addition and subtraction by human infants

Karen Wynn

Department of Psychology, University of Arizona, Tucson, Arizona 85721, USA

Revolutions in our understanding of cognitive development

ELSEVIER

Cognition 74 (2000) B1–B11

COGNITION

www.elsevier.com/locate/cognit

Brief article

Large number discrimination in 6-month-old infants

Fei Xu[a,*], Elizabeth S. Spelke[b]

Revolutions in our understanding of cognitive development

ELSEVIER

Cognition 69 (1998) 1–34

COGNITION

Infants selectively encode the goal object of an actor's reach

Amanda L. Woodward*

Revolutions in our understanding of cognitive development

ELSEVIER

Cognition 56 (1995) 165–193

COGNITION

Taking the intentional stance at 12 months of age

György Gergely*, Zoltán Nádasdy, Gergely Csibra, Szilvia Bíró

Revolutions in our understanding of cognitive development

Science AAAS

Do 15-Month-Old Infants Understand False Beliefs?
Kristine H. Onishi, *et al.*
Science **308**, 255 (2005);
DOI: 10.1126/science.1107621

Revolutions in our understanding of cognitive development

PSYCHOLOGICAL SCIENCE

Research Article

ATTRIBUTION OF DISPOSITIONAL STATES BY 12-MONTH-OLDS

Valerie Kuhlmeier, Karen Wynn, and Paul Bloom

Yale University

Revolutions in our understanding of cognitive development

Vol 450|22 November 2007|**doi:10.1038/nature06288** nature

LETTERS

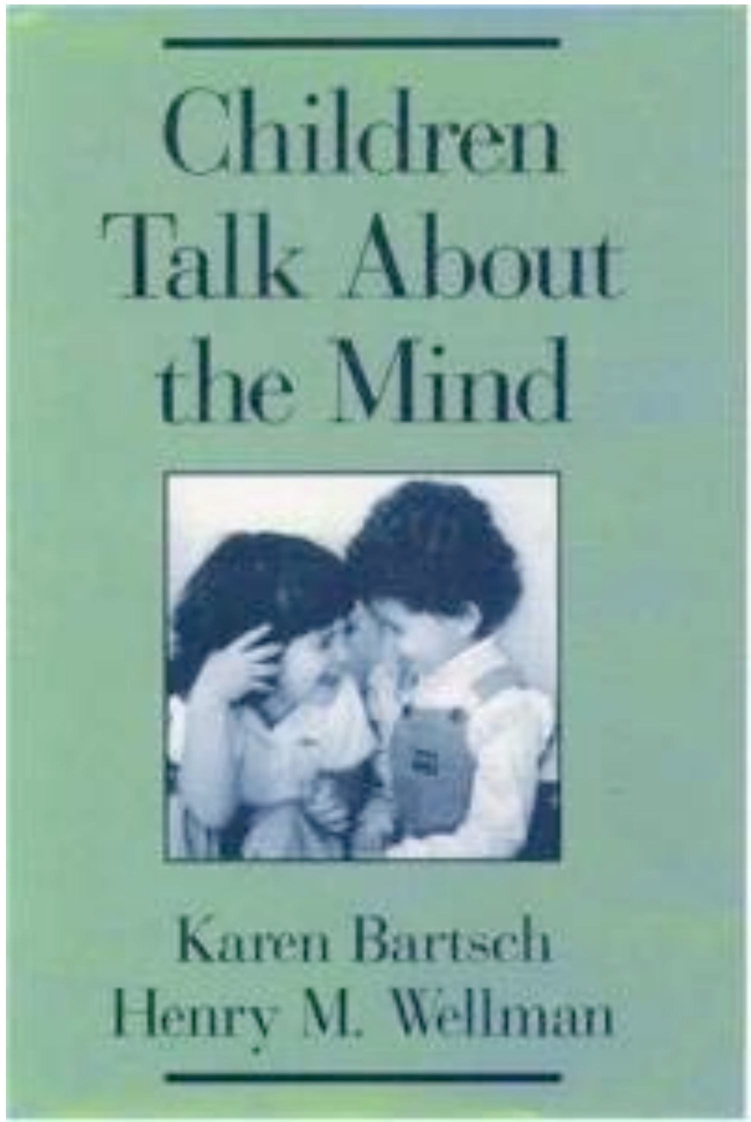
Social evaluation by preverbal infants

J. Kiley Hamlin[1], Karen Wynn[1] & Paul Bloom[1]

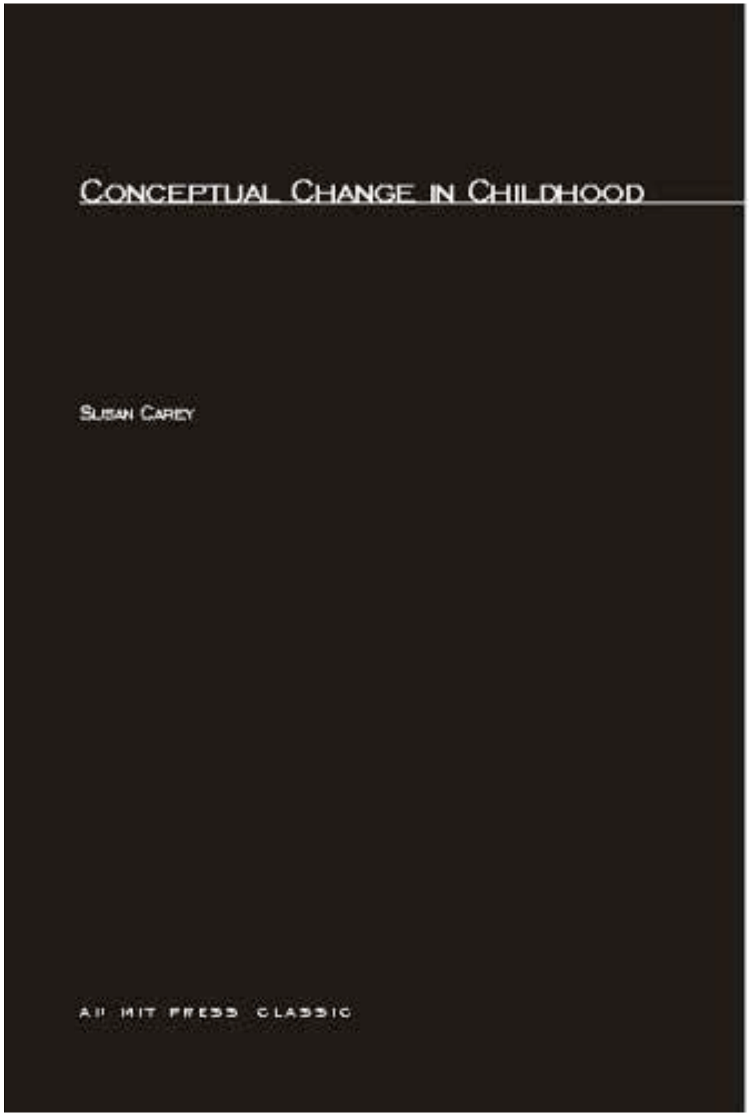
Revolutions in our understanding of cognitive development

The Social Sense: Susceptibility to Others' Beliefs in Human Infants and Adults
Ágnes Melinda Kovács, *et al.*
Science **330**, 1830 (2010);
DOI: 10.1126/science.1190792

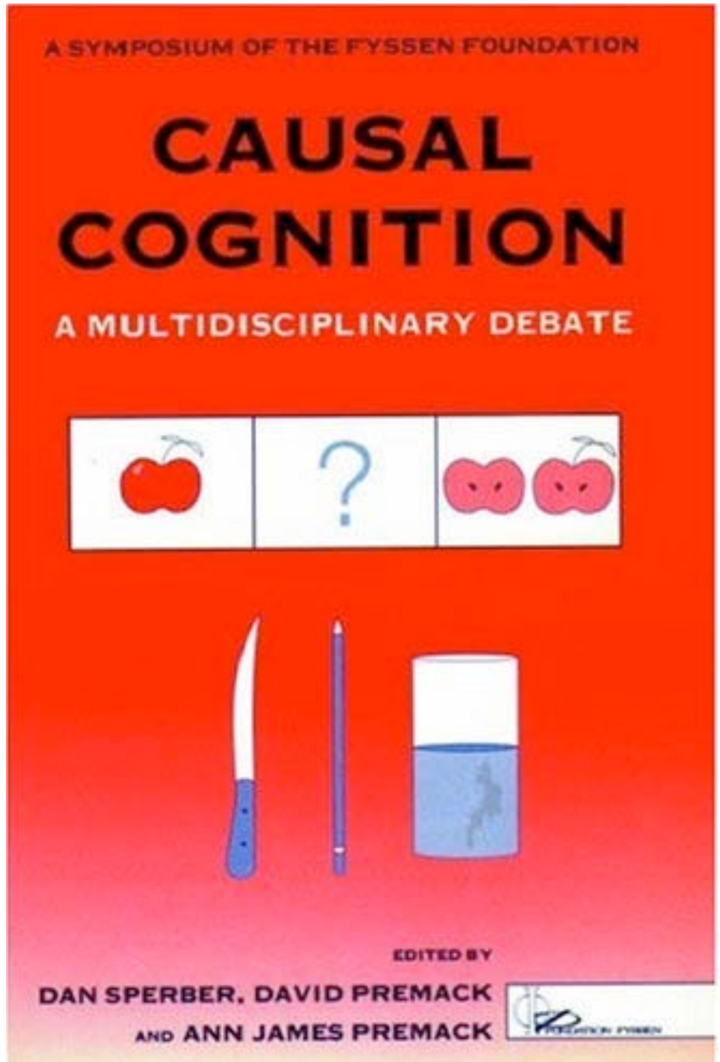
Revolutions in our understanding of cognitive development



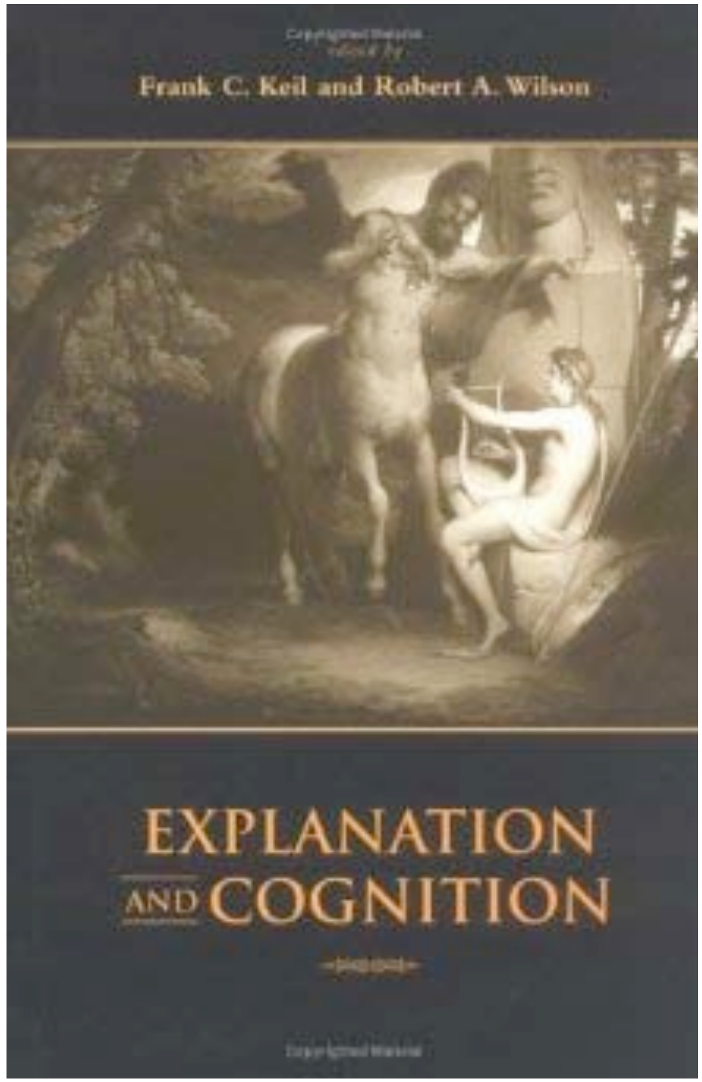
Revolutions in our understanding of cognitive development



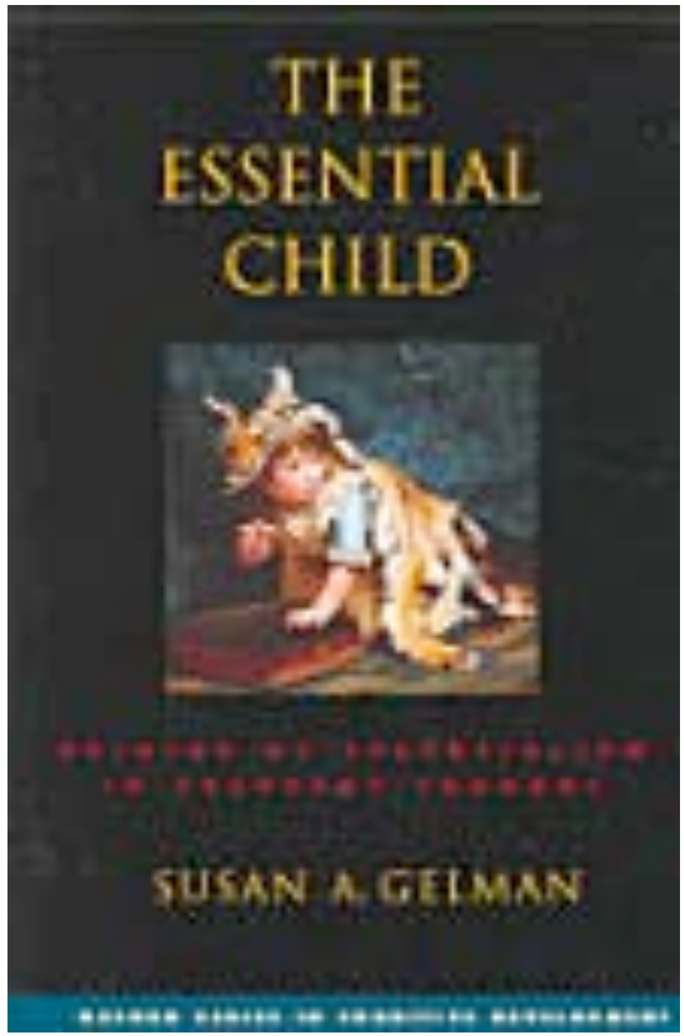
Revolutions in our understanding of cognitive development



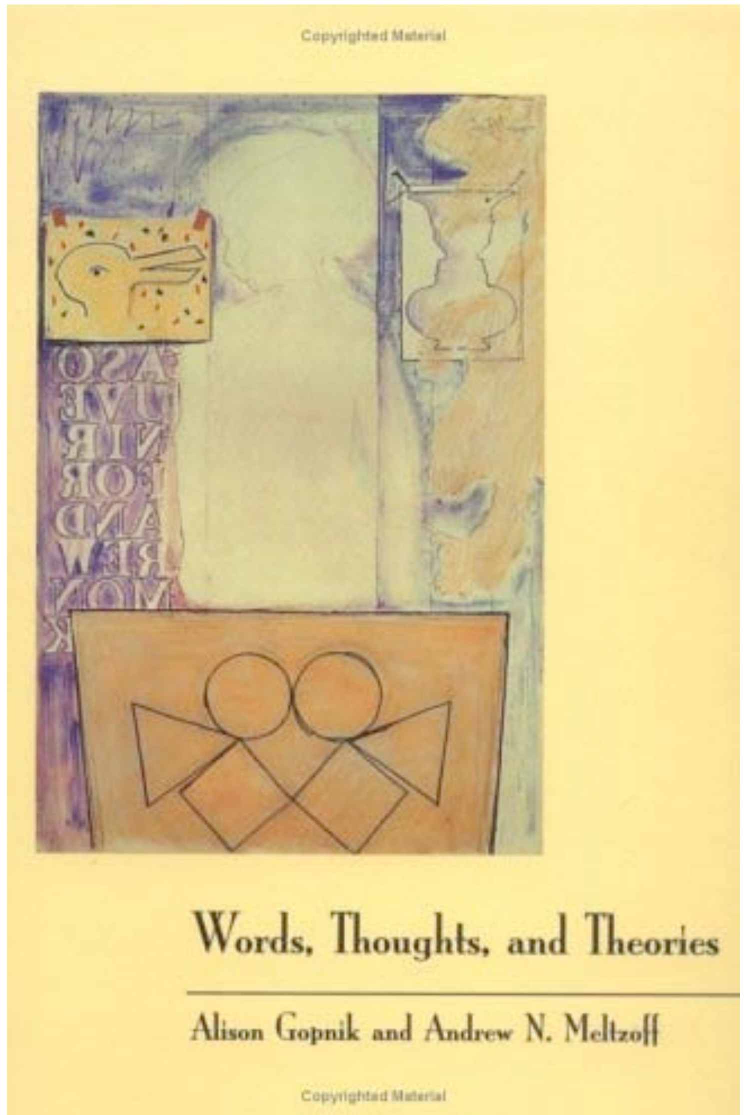
Revolutions in our understanding of cognitive development

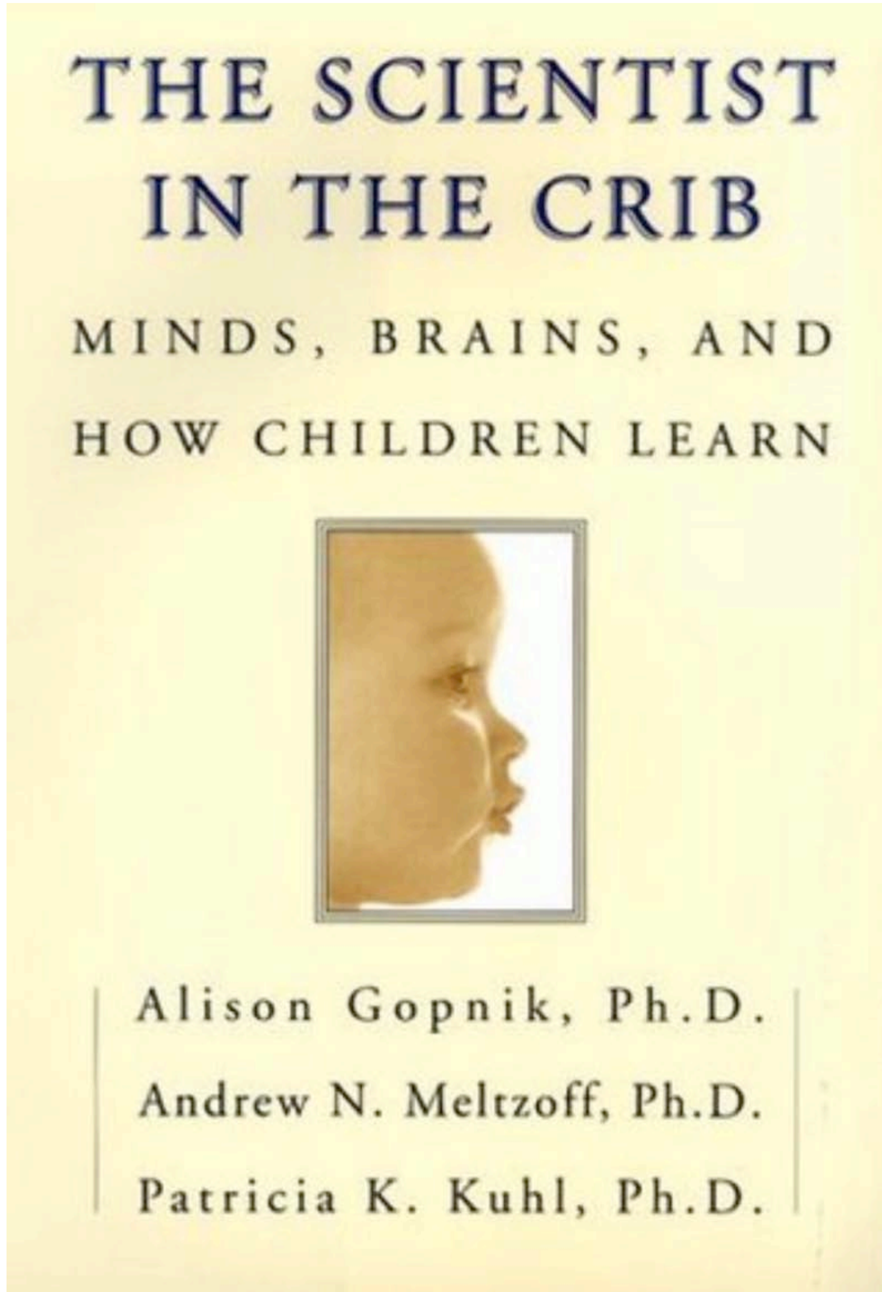


Revolutions in our understanding of cognitive development



Revolutions in our understanding of cognitive development




THE SCIENTIST
IN THE CRIB
MINDS, BRAINS, AND
HOW CHILDREN LEARN
Alison Gopnik, Ph.D.
Andrew N. Meltzoff, Ph.D.
Patricia K. Kuhl, Ph.D.

How do scientists learn?

- Scientists learn from statistical evidence
- Scientists’ beliefs affect their interpretation of statistical evidence
- Scientists distinguish genuine causes from spurious associations
- Scientists selectively explore ambiguous or confounded evidence
- Scientists introduce unobserved variables to explain data otherwise anomalous with respect to their prior beliefs
- Scientists’ generalizations depend on how evidence is sampled
- Scientists infer the relative probability of competing hypotheses and choose interventions most likely to achieve desired outcomes
- Scientists isolate variables to distinguish competing hypotheses
- Scientists evaluate expert knowledge and decide whether to learn from instruction or exploration

How do children learn?

- Children learn from statistical evidence
- Children's beliefs affect their interpretation of statistical evidence
- Children distinguish genuine causes from spurious associations
- Children selectively explore ambiguous or confounded evidence
- Children introduce unobserved variables to explain data otherwise anomalous with respect to their prior beliefs
- Children's generalizations depend on how evidence is sampled
- Children infer the relative probability of competing hypotheses and choose interventions most likely to achieve desired outcomes
- Children isolate variables to distinguish competing hypotheses
- Children evaluate expert knowledge and decide whether to learn from instruction or exploration

Appendix A

Bayesian Model

Bayesian inference provides a rational account of how children should go about combining theory and evidence. It can also be used to make precise quantitative predictions as to whether conclusions are justified by the observation of data, given a set of assumptions about the constraints provided by naive theories. In Experiment 1, children are asked, "Why does [character] have [symptom]? Is it because of [A] or because of [B]?" We model the probability that children choose explanation A as

$$\frac{P(\text{Explanation A}|D)}{P(\text{Explanation A}|D) + P(\text{Explanation B}|D)} \quad (2)$$

This directly contrasts the two possible explanations given the data observed (a similar expression applies in Experiment 2, contrasting the three possible explanations). The probability of each candidate explanation being selected given the data is computed by summing over all possible causal models that are consistent with the explanation. This is formalized as,

$$P(\text{Explanation A}|D) = \sum_{h \in H} P(\text{Explanation A}|h)P(h|D) \quad (3)$$

where h is a hypothesis as to the underlying causal structure, and H is the space of all hypotheses.

We represent hypotheses using causal graphical models (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993), where nodes correspond to variables, arrows from cause to effect represent relationships between variables, and a set of conditional probability distributions captures the probability that each variable takes on a particular value given the values of its causes. We assume that the probability of a cause being selected as an explanation given a particular causal structure h is $1/k$, where k is the size of the set of candidate causes that are present and possess a causal relationship with the effect in h, and where the proposed explanation is a member of this set. The probability of a particular causal structure given the data is obtained via Bayes's rule (Equation 1), using a prior $P(h)$ and likelihood $P(D|h)$ derived from a causal theory.

As proposed by Tenenbaum and Niyogi (2003), Griffiths (2005), and Tenenbaum et al. (2007), we model the framework theory that guides children's inferences as a simple scheme for generating causal graphical models. In this scheme, we allow for different domains. Causal variables have relationships with effect variables; causes are likely to have relationships to effects within their domain, however, there is also a small probability that a cause from one domain can lead to an effect in another domain.

The prior probability associated with each model is simply its probability of being generated by the theory. The process of generating a causal graphical model from this theory breaks down into three steps. First, we identify the nodes (causes and effects) in the model. In our case, the nodes simply correspond to the set of causes and effects that appear in the story. Second, we generate the causal relationships between these nodes. If cause and effects are within domain, then the probability a relationship exists is relatively high and given by p. If the link between two variables crosses domains, then a relationship is unlikely and is given a lower probability, q. With n causes, there are 2^n possible causal models. Assuming that each relationship is generated independently, we can evaluate the prior probability of each of these models by multiplying the probabilities of the existence or nonexistence of the causal relationships involved. The particular values of the probabilities p and q depend on the child's theory. Such theories might change with age and experience; that is, younger children might think cross-domain events are more or less probable than older children. We assume that children think the probability of cross-domain events is low (but not extremely low) by setting $q = .1$, and by setting a higher within-domain probability $p = .4$.

Finally, we specify the conditional probability of the effect given the causes present in the causal model. This allows us to evaluate the probability of a specific model, h, generating the data observed on the mth day, $P(d_m|h)$. These data consist of the values taken on by all variables on that day—the presence or absence of the causes and effects. We assume that the probability of each cause being present or absent is constant across all of the causal models and the only difference is in the probability they assign to the occurrence of the effect on that day. We then take the conditional probability of the effect given the set of causes to be 1 if any cause that influences the effect is present, and ε otherwise, corresponding to a noisy-OR parameterization (Pearl, 2000), where each cause has a strength of 1 and the background has a strength of ε. We assumed that the probability of an effect in the absence of any causes was low, with $\varepsilon = .001$. The probability of the full set of data, D, accumulated over the course of the story is given by

$$P(D|h) = \Pi_m P(d_m|h) \quad (4)$$

where the data observed on each day are assumed to be generated independently.

As can be seen comparing the results predicted by the Bayesian model in Figure A1 with the 4-year-olds' responses in Experiments 1 and 2, our model accurately predicted the responses of the oldest children, with a Pearson product–moment correlation coefficient of $r(9) = .85$. The model gives correct relative weights to the variables at baseline in both the within-domain and cross-domains conditions. Critically, the model predicted the increased A responses after evidence in all conditions, while still capturing the more subtle graded interaction between theory and evidence.

Appendix A

Bayesian Model

Bayesian inference provides a rational account of how children should go about combining theory and evidence. It can also be used to make precise quantitative predictions as to whether conclusions are justified by the observation of data, given a set of assumptions about the constraints provided by naive theories. In Experiment 1, children are asked, "Why does [character] have [symptom]? Is it because of [A] or because of [B]?" We model the probability that children choose explanation A as

$$\frac{P(\text{Explanation A}|D)}{P(\text{Explanation A}|D) + P(\text{Explanation B}|D)} \quad (2)$$

This directly contrasts the two possible explanations given the data observed (a similar expression applies in Experiment 2, contrasting the three possible explanations). The probability of each candidate explanation being selected given the data is computed by summing over all possible causal models that are consistent with the explanation. This is formalized as,

$$P(\text{Explanation A}|D) = \sum_{h \in H} P(\text{Explanation A}|h)P(h|D) \quad (3)$$

where h is a hypothesis as to the underlying causal structure, and H is the space of all hypotheses.

We represent hypotheses using causal graphical models (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993), where nodes correspond to variables, arrows from cause to effect represent relationships between variables, and a set of conditional probability distributions captures the probability that each variable takes on a particular value given the values of its causes. We assume that the probability of a cause being selected as an explanation given a particular causal structure h is $1/k$, where k is the size of the set of candidate causes that are present and possess a causal relationship with the effect in h, and where the proposed explanation is a member of this set. The probability of a particular causal structure given the data is obtained via Bayes's rule (Equation 1), using a prior $P(h)$ and likelihood $P(D|h)$ derived from a causal theory.

As proposed by Tenenbaum and Niyogi (2003), Griffiths (2005), and Tenenbaum et al. (2007), we model the framework theory that guides children's inferences as a simple scheme for generating causal graphical models. In this scheme, we allow for different domains. Causal variables have relationships with effect variables; causes are likely to have relationships to effects within their domain, however, there is also a small probability that a cause from one domain can lead to an effect in another domain.

The prior probability associated with each model is simply its probability of being generated by the theory. The process of generating a causal graphical model from this theory breaks down into three steps. First, we identify the nodes (causes and effects) in the model. In our case, the nodes simply correspond to the set of causes and effects that appear in the story. Second, we generate the causal relationships between these nodes. If cause and effects are within domain, then the probability a relationship exists is relatively high and given by p. If the link between two variables

Kaelbling, 1993; Kang et al., 2009; McClure, Daw, & Read Montague, 2003; Sutton & Barto, 1998). Although many of these studies have focused not on richly structured causal domains, but on arbitrary reinforcement learning problems (e.g., multi-armed bandit tasks, stacked card decks, maze-running tasks with hidden rewards, etc.), computational models of the expected information gain associated with different actions establish an important starting point for the current work.

It is optimal (in a decision-theoretic sense) to maximize the expected information gain of an action (see Oaksford & Chater, 1994); in particular, it is optimal to choose the action that will most decrease the uncertainty over the causal structure that relates the variables of interest. Formally, the information gain from observing data D after taking action A can be represented by:

$$I_g(D, A) = I(P(H|D, A)) - I(P(H)), \quad (1)$$

where $I(p)$ represents the Shannon–Wiener information (Shannon and Weaver, 1948; Wiener, 1948) of distribution p: $I(p) = -\Sigma_i p_i \log_2(p_i)$. The expected information gain from taking action A is then:

$$EI_g(A) = \Sigma_i P(D_i|A) I_g(D_i, A), \quad (2)$$

where the probability of an observation given an action is:

$$P(D_i|A) = \Sigma_k P(D_i|A, H_k) P(H_k). \quad (3)$$

In choosing the action which maximizes the expected information gain there are thus two main contributing factors: $P(H)$ and $P(D|A, H)$. These describe where in the system there is information to be learned and how that information can be gathered. The prior distribution $P(H)$ will capture effects of prior knowledge (such as baserates) and earlier evidence (such as confounded observations). As noted, previous research (Gweon & Schulz, 2008; Schulz & Bonawitz, 2007) has indeed shown that children explore a causal system more when there is information to be gained.

The second term influencing optimal information gain describes how actions give rise to observed data, given the causal structure hypothesis. Typically this is assumed to be a simple, direct function: for each object there is an action (e.g. putting it on the detector) that will lead to data that reflects the causal status of that object (e.g. the detector activates if the object is a 'blicket'). Under this assumption, there is no interesting contribution from the $P(D|A, H)$ term; actions will be taken on the objects about which there is the greatest uncertainty.

However, real world scientific situations, and many situations facing children, have a much more complex relationship between available actions and the causal variables of interest. Designing an experiment is often harder than merely choosing what one should try to learn. For instance, if there is not a one-to-one relationship between actions and objects then $P(D|A, H)$ can reflect an inability to isolate some causal variables. In our first experiment, we set up a situation in which there are four objects of uncertain causal status, but only three actions: two which will lead to evidence about two objects individually, and one which leads to (confounded) evidence about a pair of objects. Thus while the contribution of $P(H)$ to EI_g supports exploration of all four objects equally, the incorporation of $P(D|A, H)$ leads to preferential exploration with the separable objects. We test this prediction in Experiment 1.

Perhaps more than any other claim about causal reasoning in young children, the idea that children might be capable of selectively performing informative interventions puts the analogy between children and scientists to the test. This is the claim we investigate here. We hypothesize that when the probability of information gain is high, preschoolers will exploit available affordances to isolate and test causal variables consistent with their folk theories about plausible mechanisms.

By suggesting that preschoolers selectively perform informative interventions, we mean both something *more* than the idea that children learn through exploratory play and something *less* than the idea that children explicitly understand and apply principles of experimental design. We believe children's spontaneous experimentation can be distinguished from trial and error learning or "mere" exploratory behavior by its selectivity. That is, we predict that children will be more likely to perform actions that isolate relevant variables when the probability of information gain is high than when it is low, and that children will be specifically likely to perform actions that isolate relevant variables (rather than simply acting more in general). Children's spontaneous experimentation can be distinguished from a meta-cognitive understanding of experimental design both by its noisy implementation and its fragility. We do not predict that children will perform only informative experiments, that they will perform informative experiments in preference to other playful actions, or that they will perform informative experiments methodically (without redundancy or interruption). Additionally, we believe that children's ability to generate informative actions can be easily compromised by other task demands (e.g., by increasing the number of variables involved or changing the status of those variables with respect to the children's prior beliefs). We presume that bringing the ability to generate informative interventions to bear on tasks of arbitrary complexity requires formal science education.

Here we look at children's ability to design interventions in a simple toy world. We give children base rate information about candidate causes, showing them either that 4 of 4 beads (the *All Beads* condition) or 2 of 4 beads (the *Some Beads* conditions) activate a toy when the beads are placed, one at a time, on top of the toy. We then show both groups of children two pairs of beads. All children learn that one of the bead pairs can be pulled apart into two individual beads, while the other pair is glued together. Finally, children learn that both bead pairs activate the toy (see Fig. 1).

Although in principle only one bead in each pair might be causally effective, the evidence about the bead pair should be relatively unambiguous for children in the *All Beads* condition; the base rate information strongly supports the hypothesis that both beads in both pairs activate the toy. By contrast, the evidence about the bead pairs is genuinely ambiguous for children in the *Some Beads* conditions; the evidence fails to distinguish which bead works (or whether both do). Put another way, when *All Beads*

Appendix A

Bayesian Model

Bayesian inference provides a rational account of how children should go about combining theory and evidence. It can also be used to make precise quantitative predictions as to whether conclusions are justified by the observation of data, given a set of assumptions about the constraints provided by naive theories. In Experiment 1, children are asked, "Why does [character] have [symptom]? Is it because of [A] or because of [B]?" We model the probability that children choose explanation A as

$$\frac{P(\text{Explanation A}|D)}{P(\text{Explanation A}|D) + P(\text{Explanation B}|D)} \quad (2)$$

This directly contrasts the two possible explanations given the data observed (a similar expression applies in Experiment 2, contrasting the three possible explanations). The probability of each candidate explanation being selected given the data is computed by summing over all possible causal models that are consistent with the explanation. This is formalized as,

$$P(\text{Explanation A}|D) = \sum_{h\in H} P(\text{Explanation A}|h)P(h|D) \quad (3)$$

where h is a hypothesis as to the underlying causal structure, and H is the space of all hypotheses.

We represent hypotheses using causal graphical models (Pearl, 2000; Spirtes, Glymour, & Scheines, 1993), where nodes correspond to variables, arrows from cause to effect represent relationships between variables, and a set of conditional probability distributions captures the probability that each variable takes on a particular value given the values of its causes. We assume that the probability of a cause being selected as an explanation given a particular causal structure h is $1/k$, where k is the size of the set of candidate causes that are present and possess a causal relationship with the effect in h, and where the proposed explanation is a member of this set. The probability of a particular causal structure given the data is obtained via Bayes's rule (Equation 1), using a prior $P(h)$ and likelihood $P(D|h)$ derived from a causal theory.

As proposed by Tenenbaum and Niyogi (2003), Griffiths (2005), and Tenenbaum et al. (2007), we model the framework theory that guides children's inferences as a simple scheme for generating causal graphical models. In this scheme, we allow for different domains. Causal variables have relationships with effect variables; causes are likely to have relationships to effects within their domain, however, there is also a small probability that a cause from one domain can lead to an effect in another domain.

The prior probability associated with each model is simply its probability of being generated by the theory. The process of generating a causal graphical model from this theory breaks down into three steps. First, we identify the nodes (causes and effects) in the model. In our case, the nodes simply correspond to the set of causes and effects that appear in the story. Second, we generate the causal relationships between these nodes. If cause and effects are within domain, then the probability a relationship exists is relatively high and given by p. If the link between two variables

Kaelbling, 1993; Kang et al., 2009; McClure, Daw, & Read Montague, 2003; Sutton & Barto, 1998). Although many of these studies have focused not on richly structured causal domains, but on arbitrary reinforcement learning problems (e.g., multi-armed bandit tasks, stacked card decks, maze-running tasks with hidden rewards, etc.), computational models of the expected information gain associated with different actions establish an important starting point for the current work.

It is optimal (in a decision-theoretic sense) to maximize the expected information gain of an action (see Oaksford & Chater, 1994); in particular, it is optimal to choose the action that will most decrease the uncertainty over the causal structure that relates the variables of interest. Formally, the information gain from observing data D after taking action A can be represented by:

$$I_g(D, A) = I(P(H|D, A)) - I(P(H)), \quad (1)$$

where $I(p)$ represents the Shannon–Wiener information (Shannon and Weaver, 1948; Wi
p: $I(p) = -\Sigma_i p_i \log_2(p_i)$. The expec
taking action A is then:

$$EI_g(A) = \Sigma_i P(D_i|A) I_g(D_i, A),$$

where the probability of an obse

$$P(D_i|A) = \Sigma_k P(D_i|A, H_k)P(H_k).$$

In choosing the action which
information gain there are thus
tors: $P(H)$ and $P(D|A, H)$. These
tem there is information to
information can be gathered. T
will capture effects of prior kno
and earlier evidence (such as co
noted, previous research (Gweo
Bonawitz, 2007) has indeed sho
causal system more when th
gained.

The second term influencing
describes how actions give ris
the causal structure hypothesis.
to be a simple, direct function:
action (e.g. putting it on the dete
that reflects the causal status of
tor activates if the object is a 'bli
tion, there is no interesting cont
term; actions will be taken on
there is the greatest uncertainty

However, real world scientifi
uations facing children, have a
tionship between available
variables of interest. Designing
harder than merely choosing wh
For instance, if there is not a o
tween actions and objects the
inability to isolate some causal v
iment, we set up a situation in w
of uncertain causal status, but
which will lead to evidence abou
and one which leads to (confoun
of objects. Thus while the contr

ports exploration of all four objects equally, the incorporation of $P(D|A, H)$ leads to preferential exploration with the separable objects. We test this prediction in Experiment 1.

Perhaps more than any other claim about causal reasoning in young children, the idea that children might be capable of selectively performing informative interventions puts the analogy between children and scientists to the test. This is the claim we investigate here. We hypothesize that when the probability of information gain is high, preschoolers will exploit available affordances to isolate and test causal variables consistent with their folk theories about plausible mechanisms.

By suggesting that preschoolers selectively perform informative interventions, we mean both something *more* than the idea that children learn through exploratory play and something *less* than the idea that children explicitly understand and apply principles of experimental design. We believe children's spontaneous experimentation can be distinguished from trial and error learning or "mere"

Supporting Information

Gweon et al. 10.1073/pnas.1003095107

Bayesian Model

As discussed in the main text, the joint dependence between S (sampling process) and T (property extension) can be described as a simple Bayesian network (Fig. 1). The learner's goal is to predict Y, which depends directly on T, not S or D. However, inferences about T from D must take into account the different possible values of S; formally, our Bayesian analysis must integrate out S in scoring each value of T. Because the data are inconsistent with hypothesis t_2, only two hypotheses for T are relevant; t_3 predicts that yellow balls squeak whereas t_1 predicts that they do not. Following Tenenbaum and Griffiths (1), the evidence for one of these hypotheses over the other can be measured by the likelihood ratio

$$L = \frac{P(D|t_3)}{P(D|t_1)} = \frac{P(n|t_3, \beta)}{P(n|t_1, \beta)}.$$

We posit that children's exploratory behavior—how much they squeeze the yellow ball, expecting a squeak—will be monotonically related to L (Fig. 3*A*). This analysis makes predictions that are independent of the prior probabilities children assign to t_1 or t_3, removing a degree of freedom that would otherwise need to be measured or fit empirically to their behavior. These likelihoods can be computed by integrating out the sampling process:

$$P(n|t, \beta) = \sum_{s_i \in S} P(n|t, s, \beta)P(s).$$

To evaluate these likelihoods we need the following four conditional probabilities*:

$$\begin{aligned} P(n|t_1, s_1, \beta) &= 1 \\ P(n|t_1, s_2, \beta) &= \beta^n \\ P(n|t_3, s_1, \beta) &= \beta^n \\ P(n|t_3, s_2, \beta) &= \beta^n. \end{aligned}$$

Let α denote the prior probability $P(s_1)$ that the experimenter is sampling from just the squeaky balls: $P(s_2) = 1 - \alpha$. We then have

$$\begin{aligned} P(n|t_1, \beta) &= \sum_{s_i \in S} P(n|t_1, s, \beta)P(s) \\ &= P(n|t_1, s_1, \beta)P(s_1) + P(n|t_1, s_2, \beta)P(s_2) \\ &= \alpha + \beta^n(1-\alpha). \\ P(n|t_3, \beta) &= \sum_{s_i \in S} P(n|t_3, s, \beta)P(s) \\ &= P(n|t_3, s_1, \beta)P(s_1) + P(n|t_3, s_2, \beta)P(s_2) \\ &= \beta^n\alpha + \beta^n(1-\alpha) \\ &= \beta^n. \end{aligned}$$

The likelihood ratio, measuring the evidence in favor of the proposition that yellow balls squeak, is then

$$\begin{aligned} L &= \frac{P(n|t_3, \beta)}{P(n|t_1, \beta)} \\ &= \frac{\beta^n}{\alpha + \beta^n(1-\alpha)}. \end{aligned}$$

By setting the parameter α to 0, we can model the possibility that infants expect that evidence is sampled randomly; by setting the parameter α to 1, we can model the possibility that infants expect that evidence is sampled selectively (Fig. 3 *B* and *C*).

How do children learn?

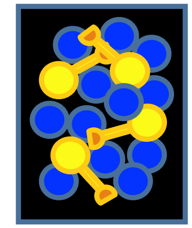
- Children learn from statistical evidence
- Children's beliefs affect their interpretation of statistical evidence
- Children distinguish genuine causes from spurious associations
- Children selectively explore ambiguous or confounded evidence
- Children introduce unobserved variables to explain data otherwise anomalous with respect to their prior beliefs
- Children's generalizations depend on how evidence is sampled
- Children infer the relative probability of competing hypotheses and choose interventions most likely to achieve desired outcomes
- Children isolate variables to distinguish competing hypotheses
- Children evaluate expert knowledge and decide whether to learn from instruction or exploration

How do children learn?

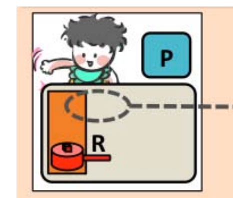
- Children learn from statistical evidence
- Children's beliefs affect their interpretation of statistical evidence
- Children distinguish genuine causes from spurious associations
- Children selectively explore ambiguous or confounded evidence
- Children introduce unobserved variables to explain data otherwise anomalous with respect to their prior beliefs
- Children's generalizations depend on how evidence is sampled
- Children infer the relative probability of competing hypotheses and choose interventions most likely to achieve desired outcomes
- Children isolate variables to distinguish competing hypotheses
- Children evaluate expert knowledge and decide whether to learn from instruction or exploration

Today's talk

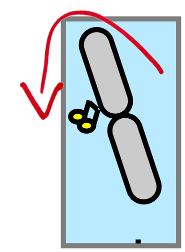
- Children's generalizations depend on how evidence is sampled



- Children infer the relative probability of hypotheses and choose interventions most likely to achieve desired outcomes.



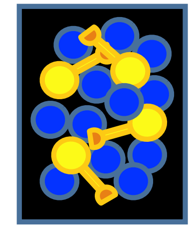
- Children isolate variables to distinguish competing hypotheses



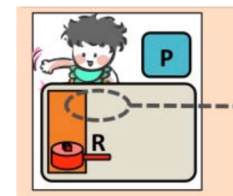
- Children evaluate expert knowledge to decide whether to learn from instruction or exploration

Today's talk

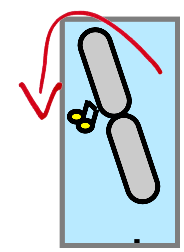
- Children's generalizations depend on how evidence is sampled



- Children infer the relative probability of hypotheses and choose interventions most likely to achieve desired outcomes.



- Children isolate variables to distinguish competing hypotheses



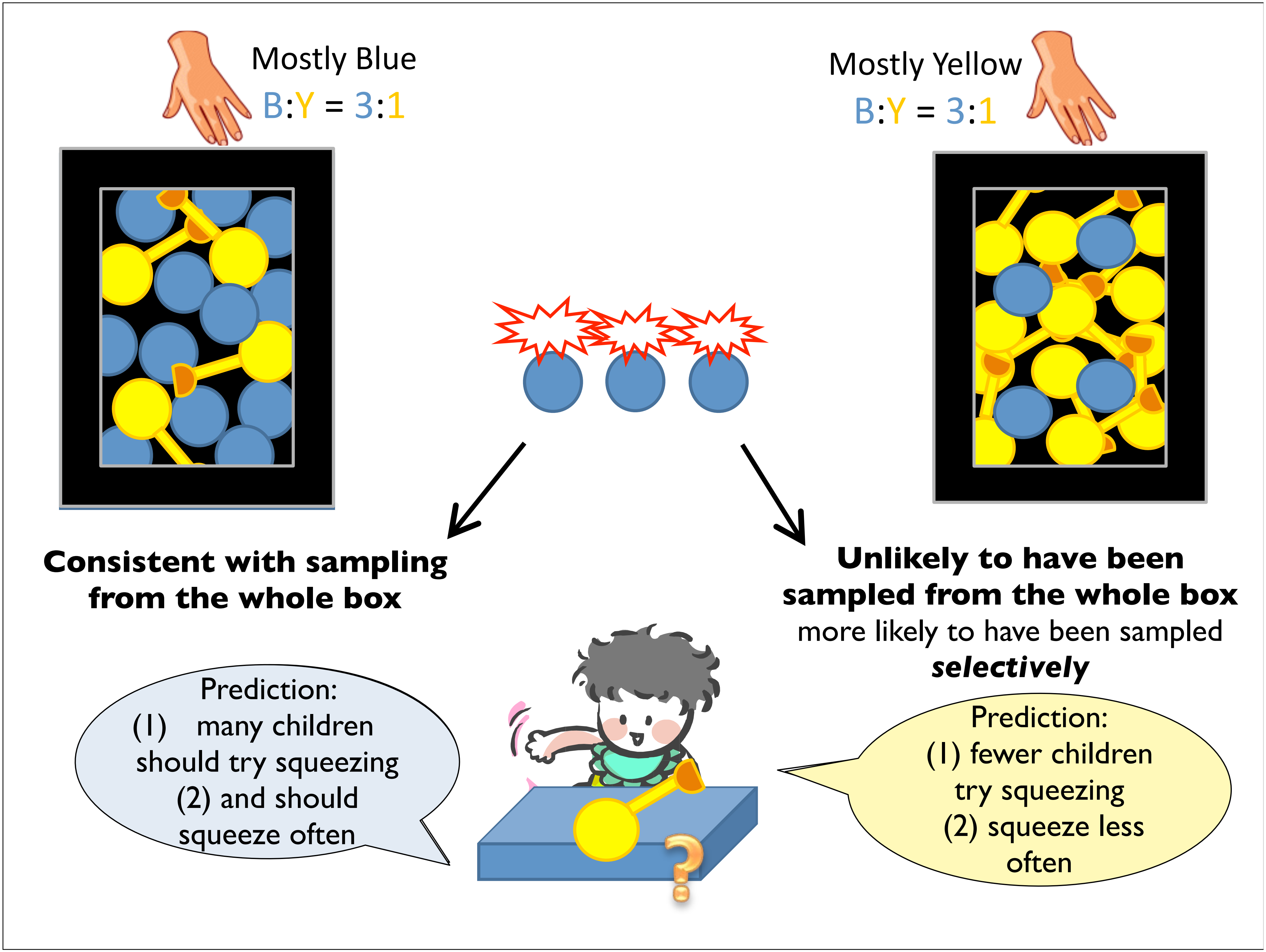
- Children evaluate expert knowledge to decide whether to learn from instruction or exploration

Generalizing from samples

- Science requires generalizing properties from a small sample to a population.
- Can use feature similarity and category membership to infer that things that look alike or belong to the same category will share properties.
- If you know that this sample of Martian rocks has a high concentration of silica, may infer that other Martian rocks have a high concentration of silica.
- If you know that this sample of needles from a Pacific silver fir lie flat on the branch, may infer other Pacific silver fir needles lie flat on the branch.

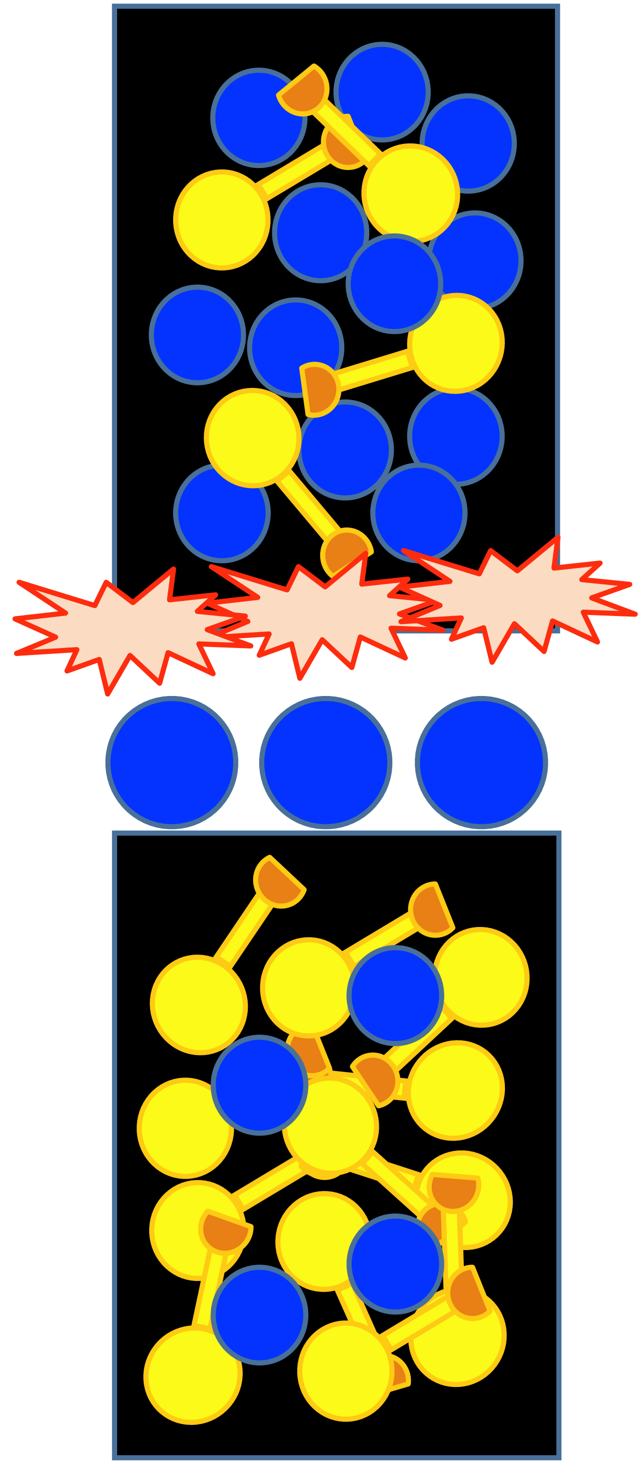
Generalizing from samples

- But as scientists we may know something about the sampling process that affects our inferences.
  - Do all Martian rocks have high concentrations of silica or only dusty rocks on the surface?
  - Do all Pacific silver fir needles lie flat or just those low on the canopy?
- How far we extend our generalizations depends on whether we think the sampling process was random or selective.
- Do infants' generalizations also take the sampling process into account?


Mostly Blue
B:Y = 3:1
Mostly Yellow
B:Y = 3:1
Consistent with sampling from the whole box
Unlikely to have been sampled from the whole box
more likely to have been sampled *selectively*
Prediction:
(1) many children should try squeezing
(2) and should squeeze often
Prediction:
(1) fewer children try squeezing
(2) squeeze less often

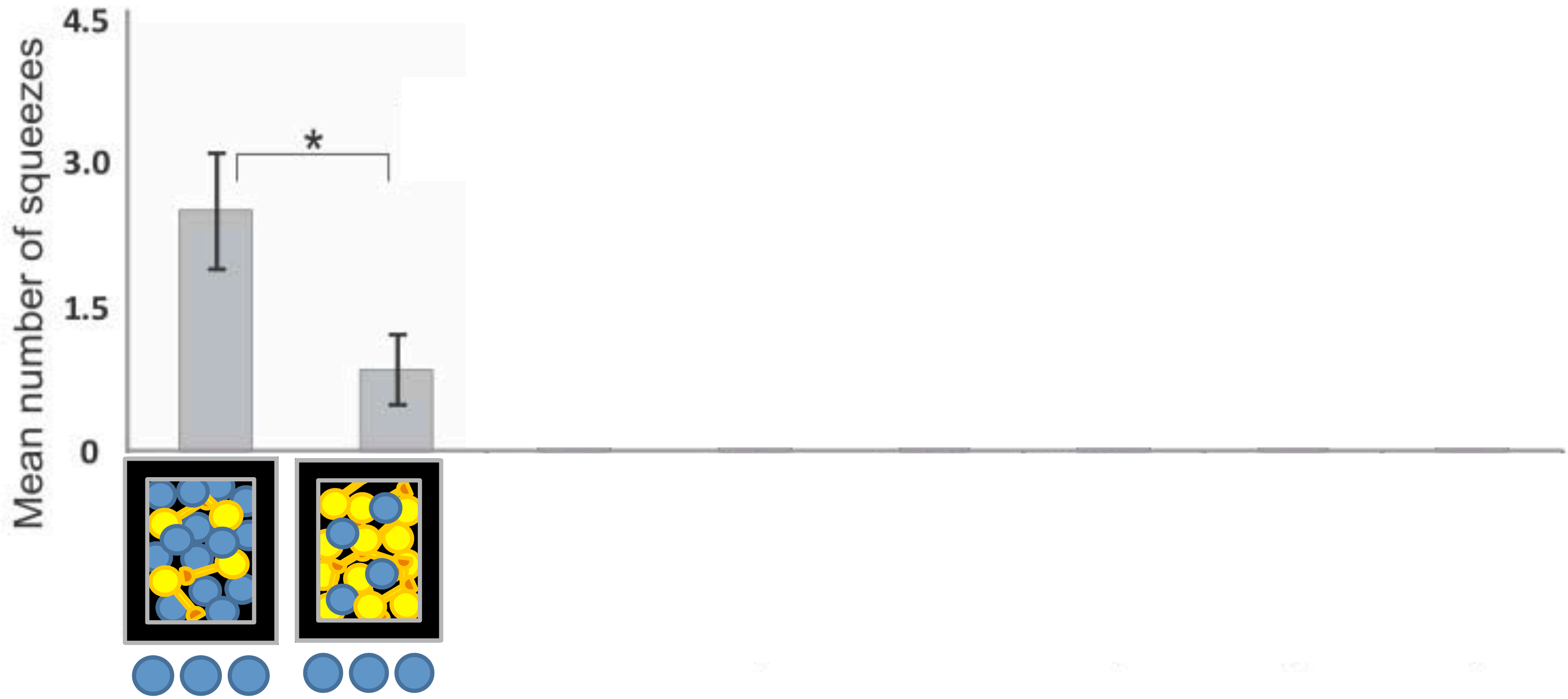
[redacted]

[redacted]



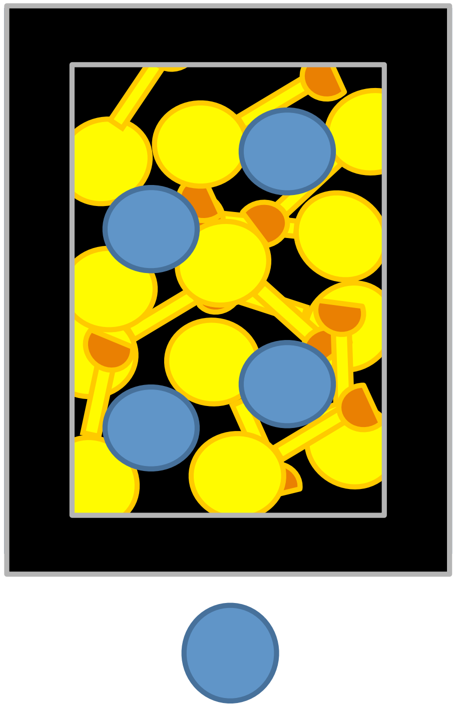
Results

n = 15/condition, mean: 15 months, 15 days, range 13-18 months



*$p < 0.05$

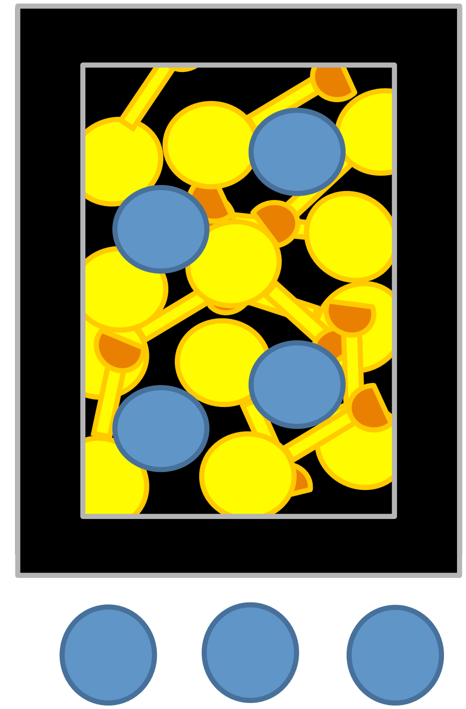
Gweon, Tenenbaum, & Schulz, 2010 *PNAS*



Squeeze once; squeeze 3 x

..not an improbable sample. Could have been generated by sampling randomly from the whole box.

Prediction:
(1) many children should try squeezing
(2) and should squeeze often

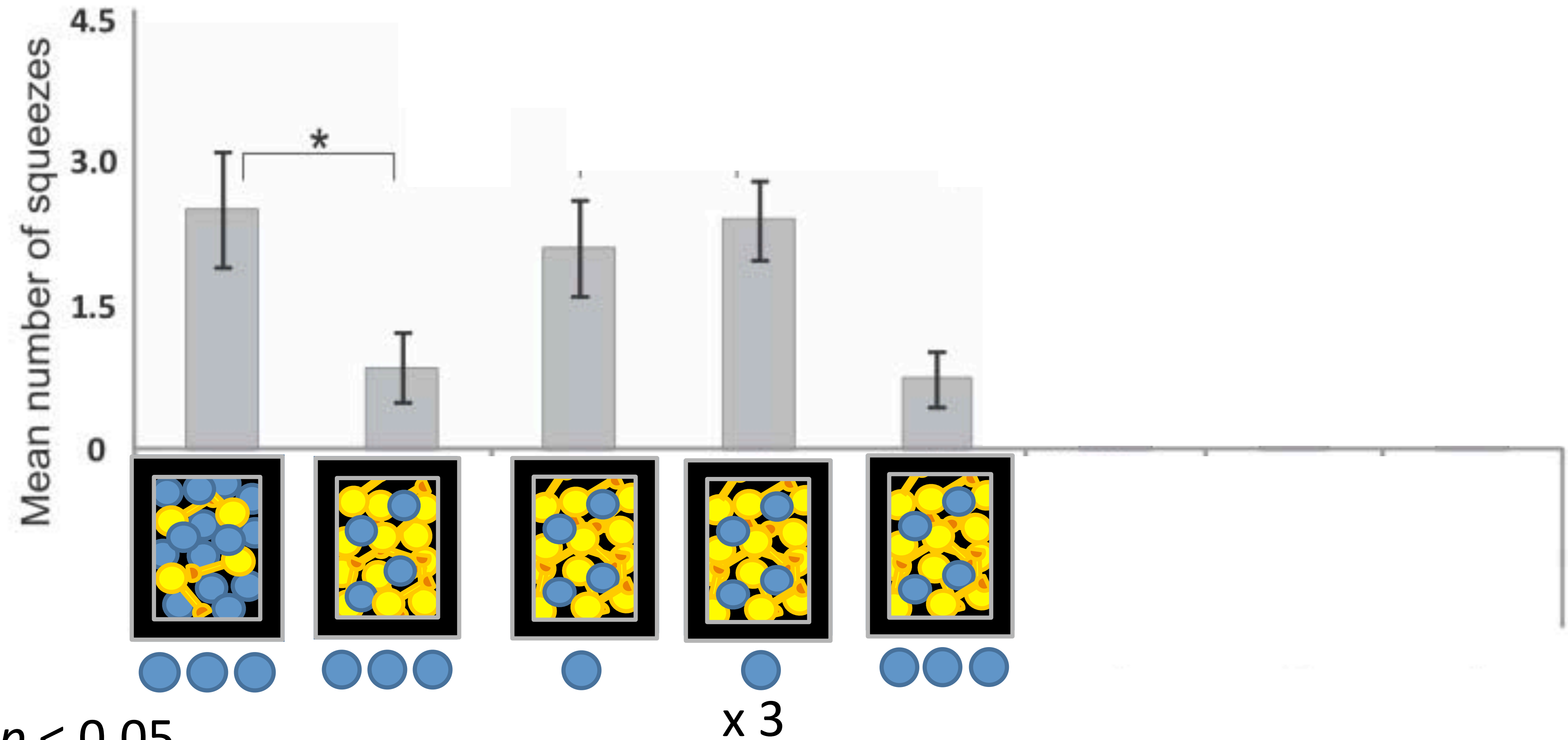


Unlikely to have been sampled from the whole box
more likely to have been sampled ***selectively***

Prediction:
(1) few children try squeezing
(2) squeeze less often

Results

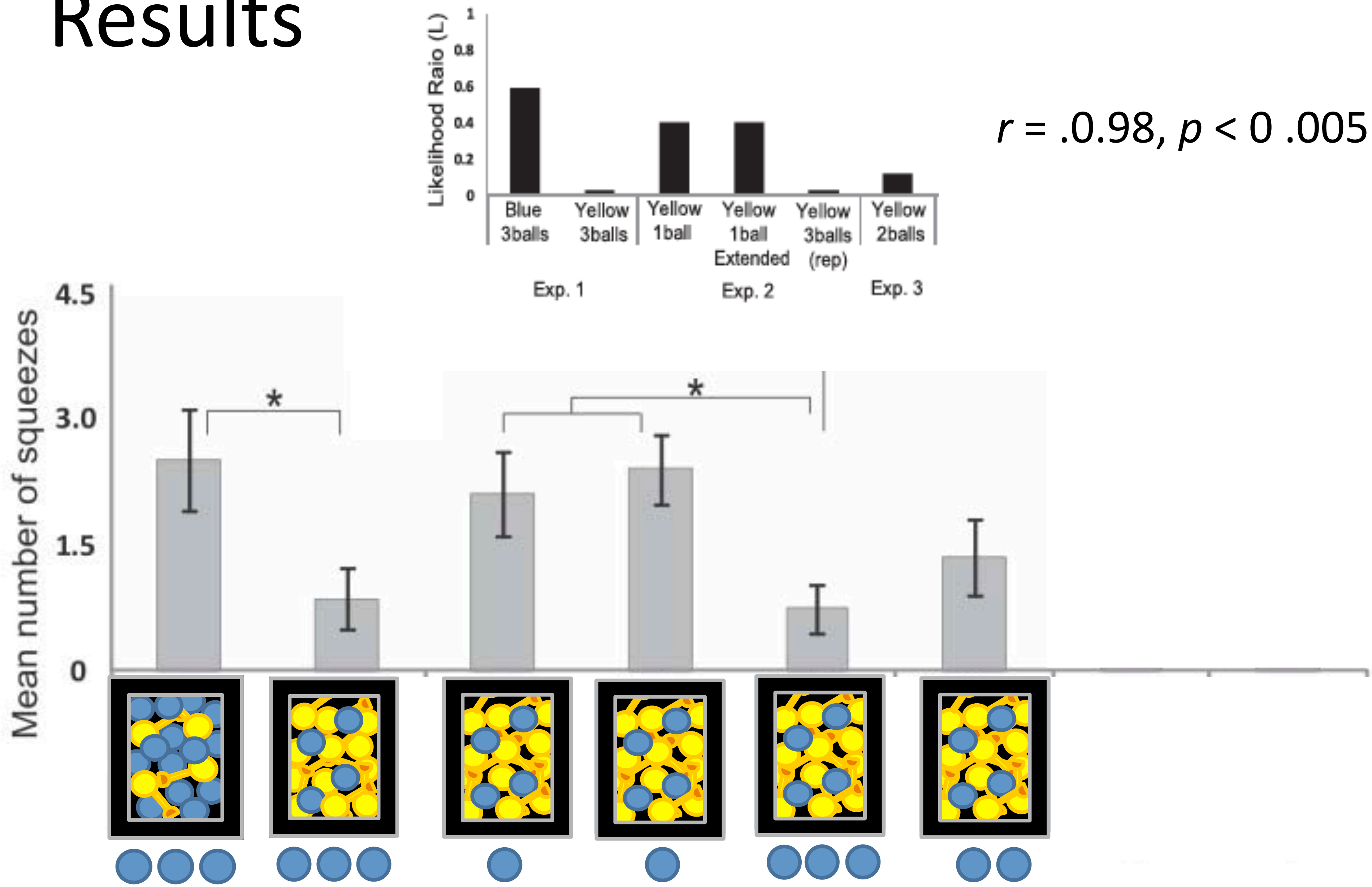
n = 16/condition, mean: 15 months, 15 days, range 13-18 months



*$p < 0.05$

Gweon, Tenenbaum, & Schulz, 2010 *PNAS*

Results

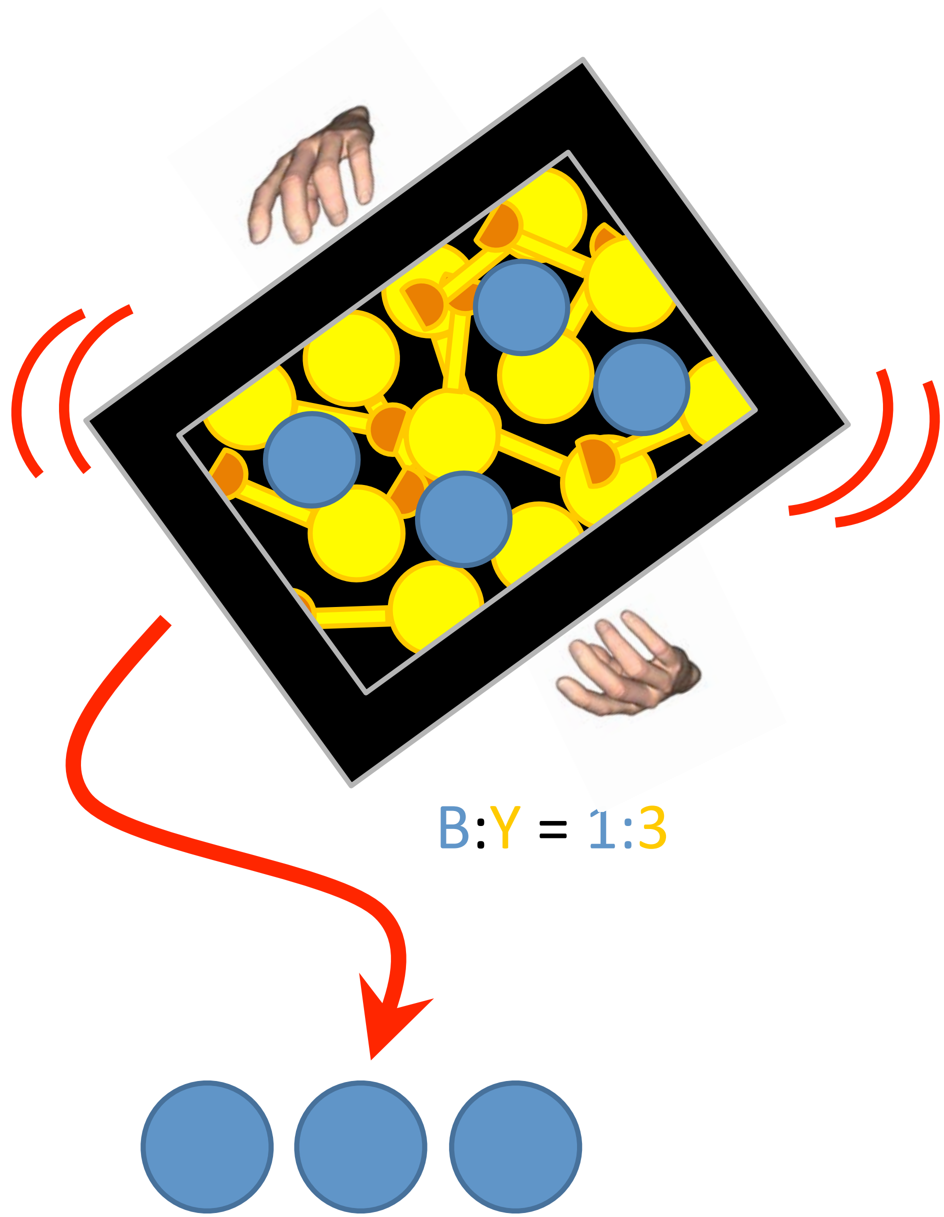


r = .0.98, p < 0 .005

Gweon, Tenenbaum, & Schulz, 2010 *PNAS*

What if…

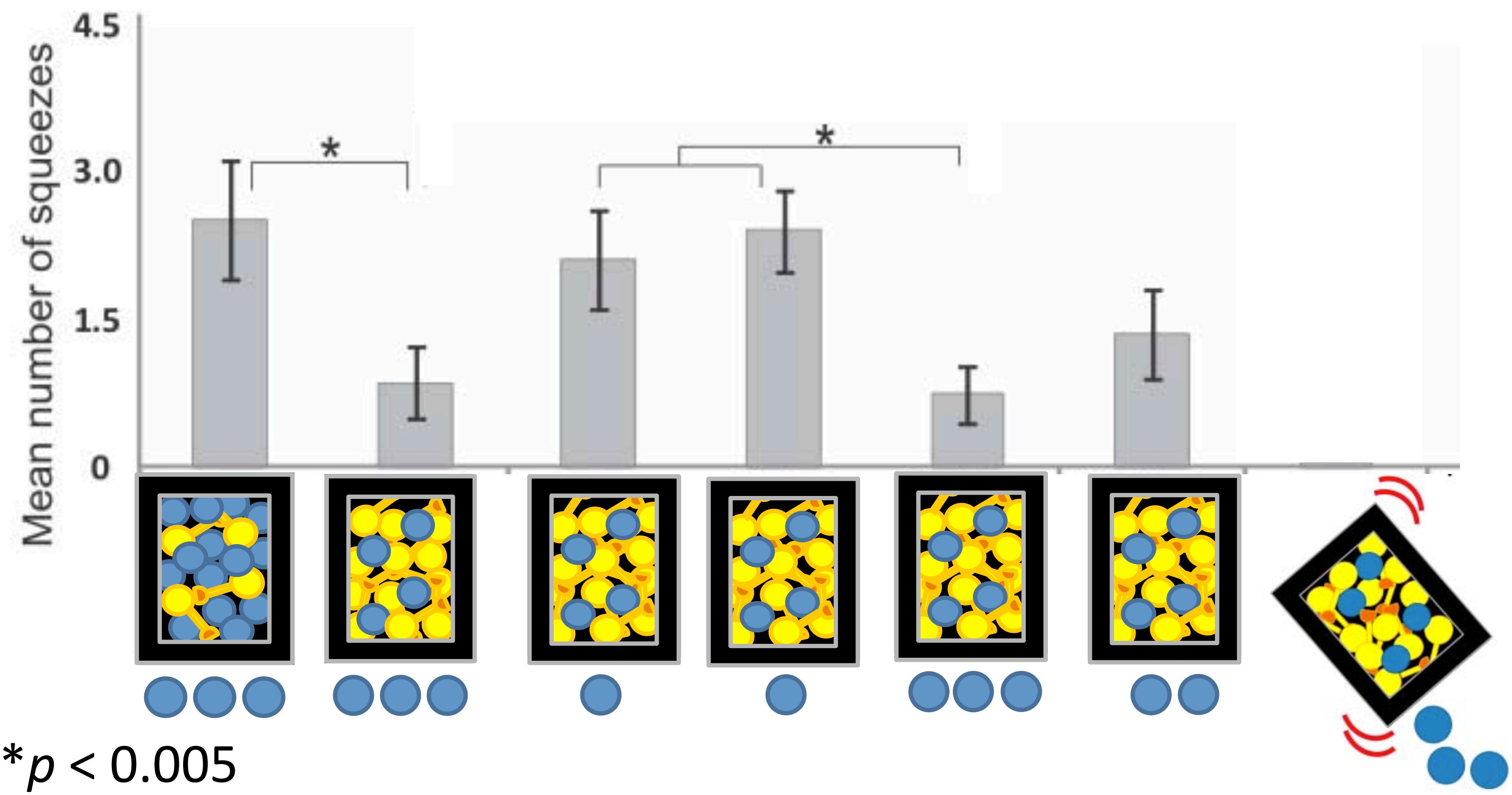
- Explicit cue for random sampling process



Gweon, Tenenbaum, & Schulz, 2010 *PNAS*

Results

n = 16/condition, mean: 15 months, 15 days, range 13-18 months



**$p < 0.005$

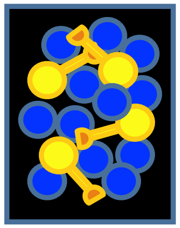
Gweon, Tenenbaum, & Schulz, 2010 *PNAS*

Generalizing from samples

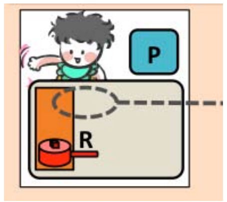
- 15-month-olds' generalizations take into account more than category membership and the perceptual similarity of objects.
- Infants make graded inferences that are sensitive to both the amount of evidence they observe and the process by which the evidence is sampled.

Today’s talk

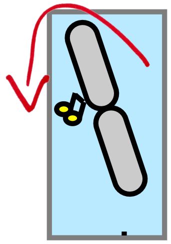
- Children’s generalizations depend on how evidence is sampled



- Children infer the relative probability of hypotheses and choose interventions most likely to achieve desired outcomes.



- Children isolate variables to distinguish competing hypotheses

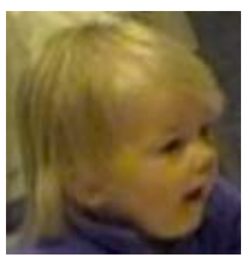


- Children evaluate expert knowledge to decide whether to learn from instruction or exploration

Can 16-month-olds…

- use minimal statistical data to make appropriate causal attributions?
- rationally choose to seek help vs. explore based on these inferences?

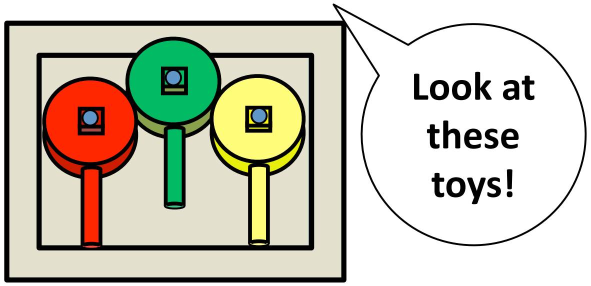
It's **ME**
(agent)



It's **THE WORLD**
(object)

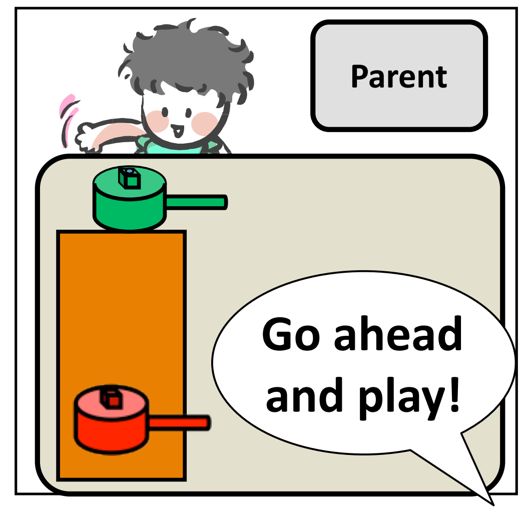
Approach Another Agent

Approach Another Toy



Distribution of successes and failures within & between different agents

DEMONSTRATION	ATTRIBUTION	ACTION
Within Agents E1 succeeds, E1 fails, E2 fails, E2 succeeds	"It's probably **the toy**..."	Change the Object
Between-Agents E1 succeeds, E2 fails, E2 fails, E1 succeeds	"It's probably **me**..."	Change the Agent



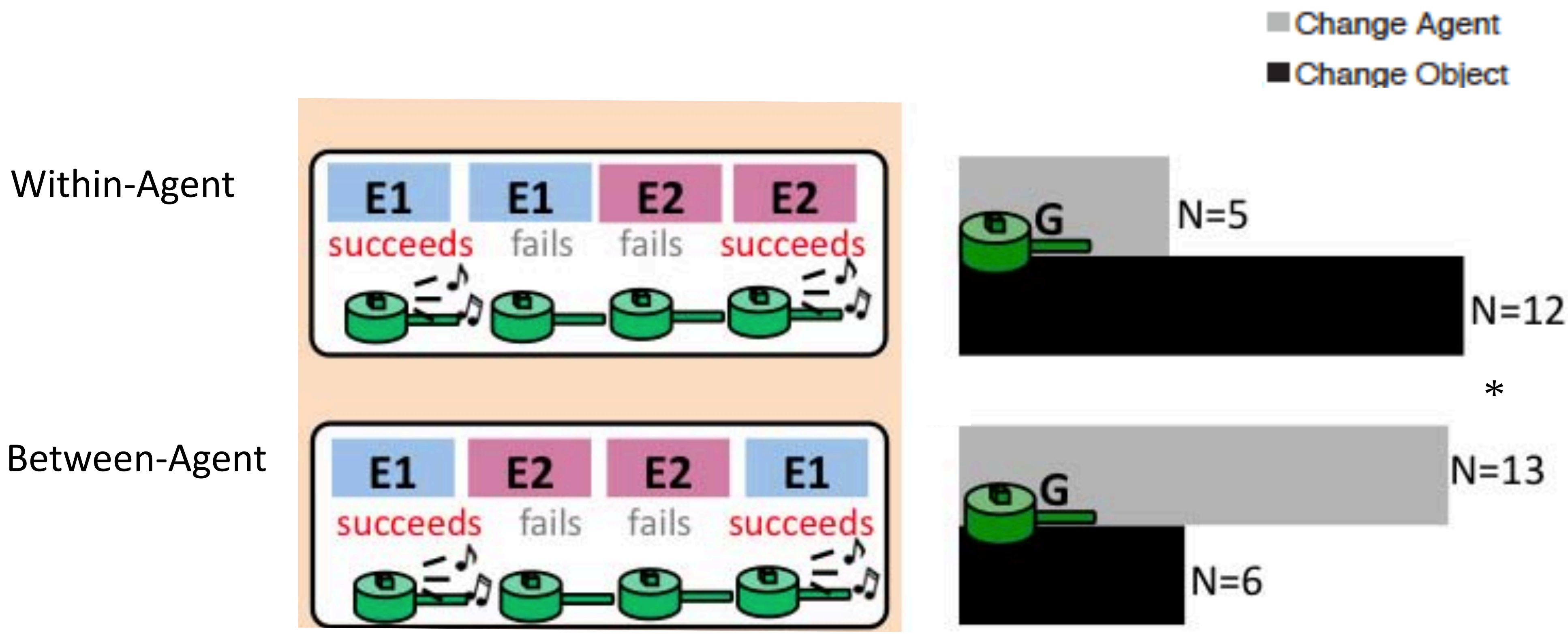
Gweon & Schulz, 2011, *Science*

[redacted]

[redacted]

Results

Histogram showing number of infants performing each action first in each condition
N = 36 infants, mean: 16 months; range: 13-20 months



* difference between conditions,
p < .05 Fisher's exact test

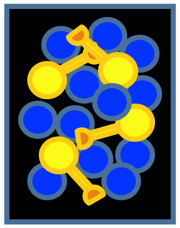
Gweon & Schulz, 2011, *Science*

Rational causal inference in infants

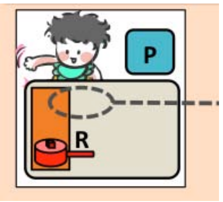
- 16-month-olds…
  – track the statistical dependence between agents, objects, and outcomes
  – can use minimal data to make rational attributions about the cause of failed goal-directed actions
- These distinct explanatory attributions (self vs. world) help them choose between two different strategies for learning
  - seeking instruction from others
  - self-guided exploration

Today's talk

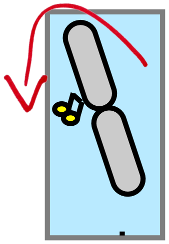
- Children's generalizations depend on how evidence is sampled



- Children infer the relative probability of hypotheses and choose interventions most likely to achieve desired outcomes.

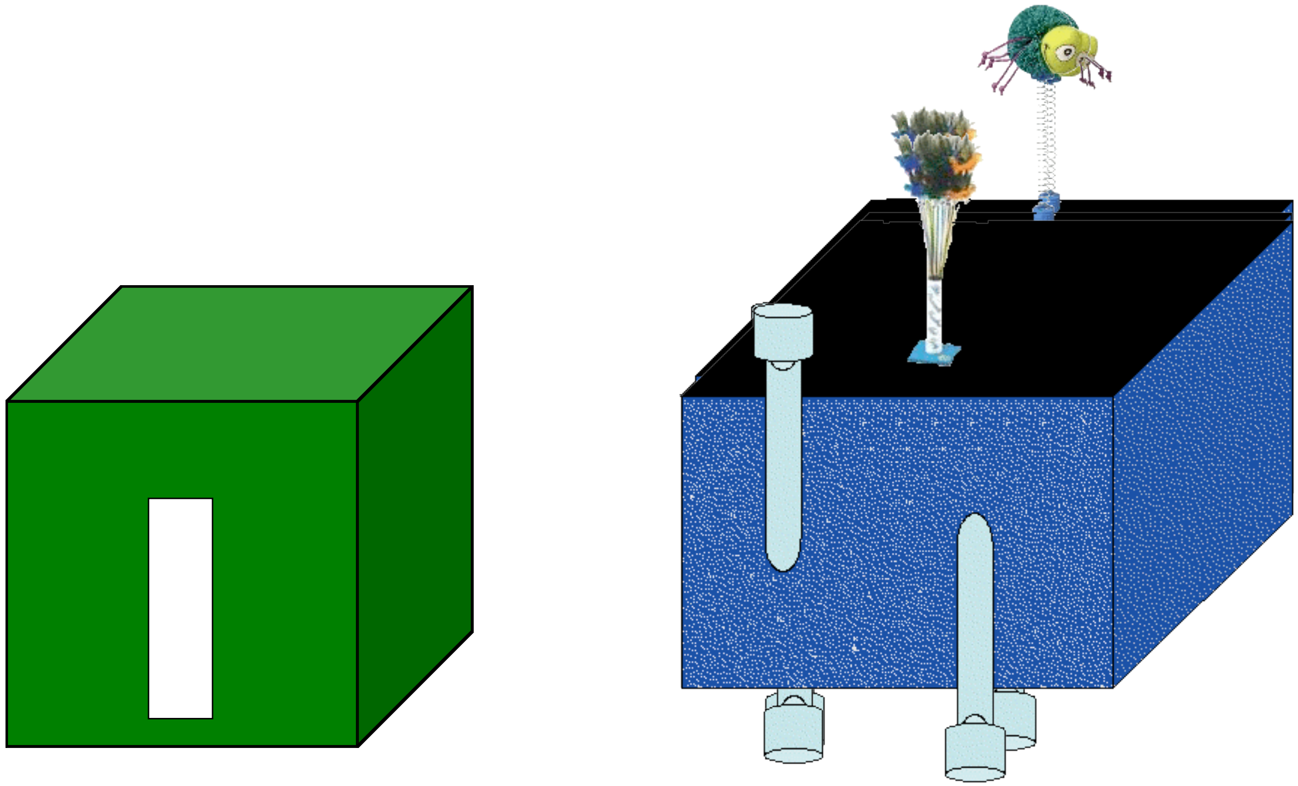


- Children isolate variables to distinguish competing hypotheses



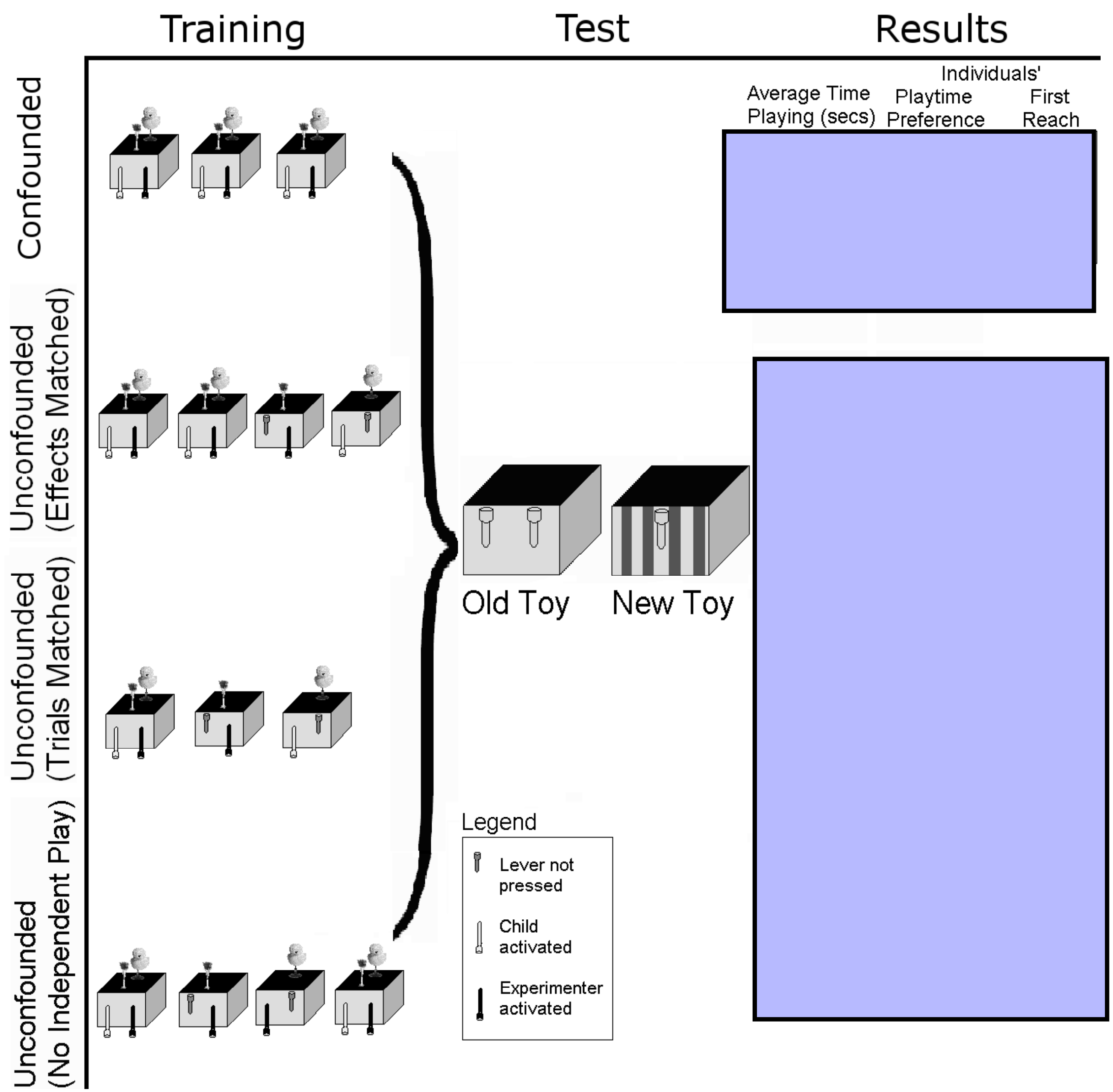
- Children evaluate expert knowledge to decide whether to learn from instruction or exploration

Children selectively engage in exploratory play when evidence fails to distinguish competing hypotheses (e.g., when evidence is confounded)



Schulz & Bonawitz, 2007, *Developmental Psychology*

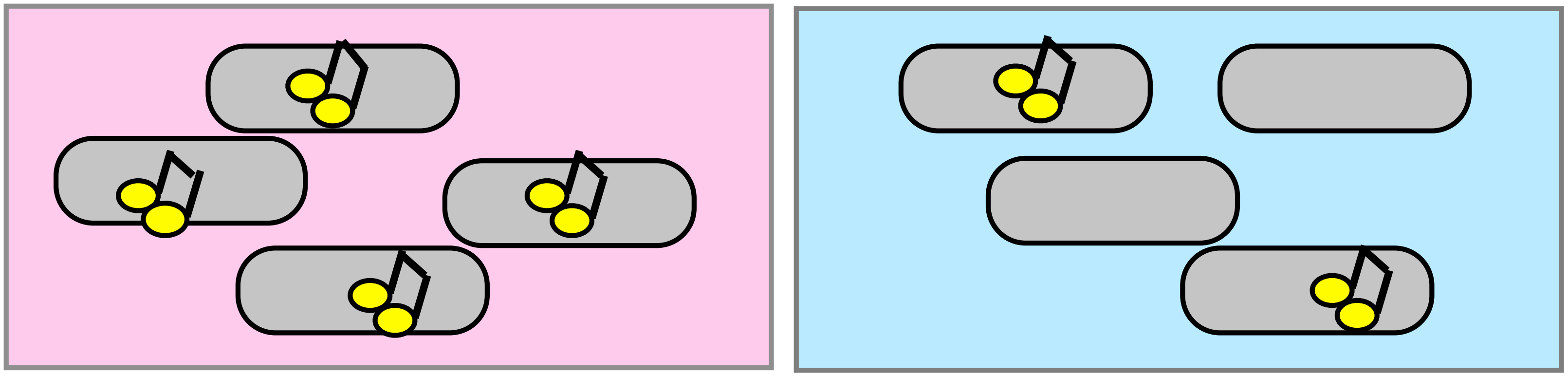
n = 16/condition
four & five-year-olds
mean: 57 months



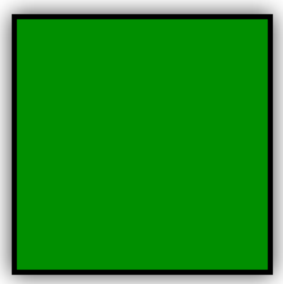
Study 1

Schulz & Baraff Bonawitz, 2007, Developmental Psychology

Children assigned to one of two training conditions

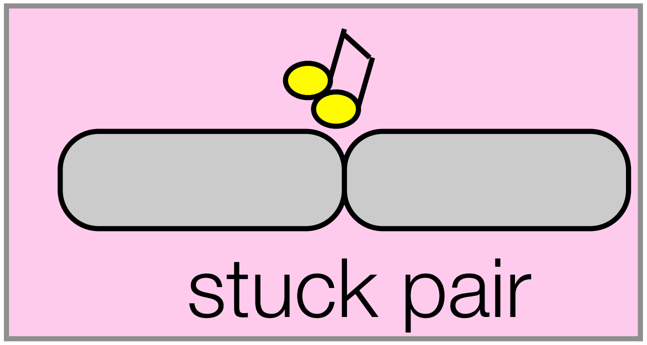


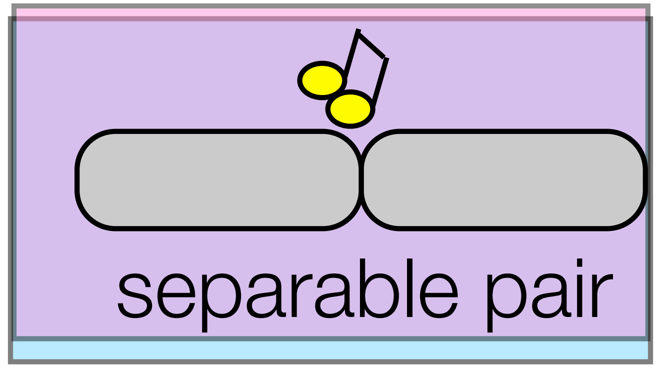
all beads condition some beads condition



Cook, Goodman, & Schulz, 2011, *Cognition*

All children given the same test condition



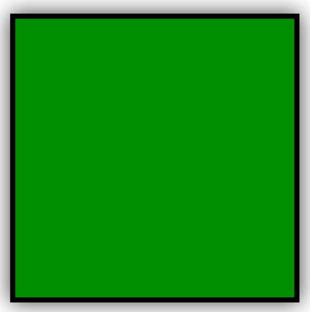


some beads condition

all beads condition

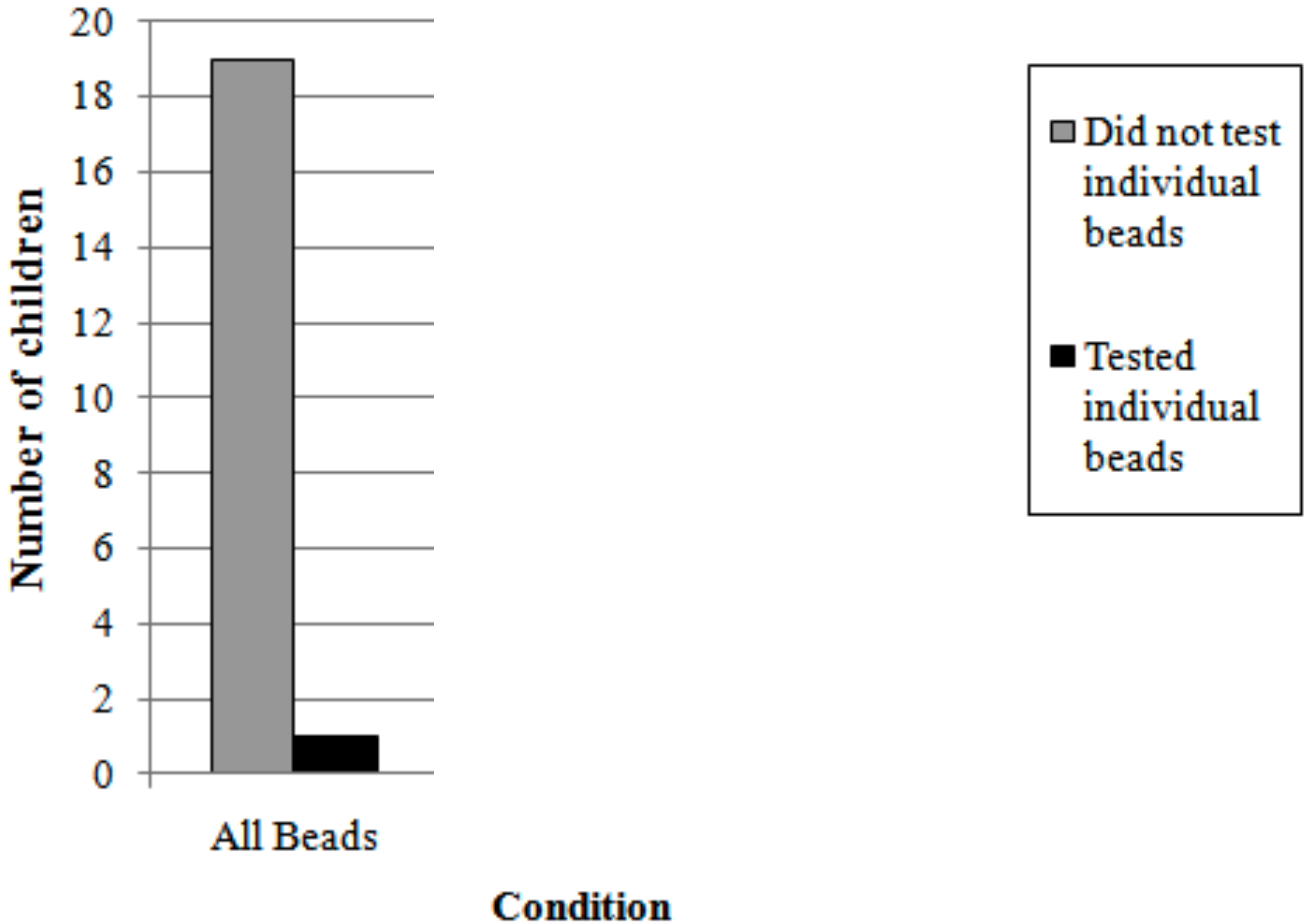
Some Beads B

Some Beads AB



Cook, Goodman, & Schulz, 2011, *Cognition*

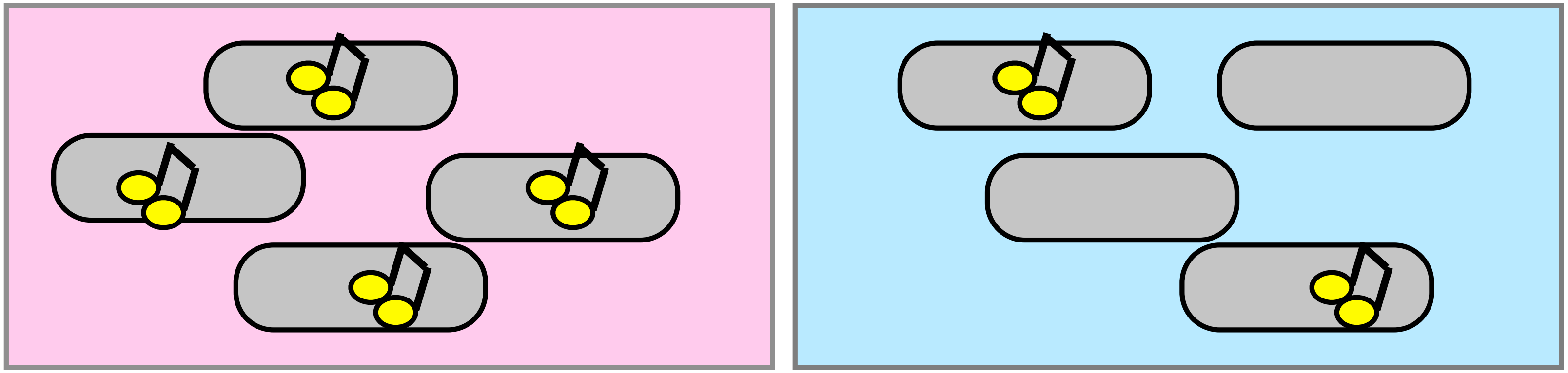
Histogram showing number of children generating each action in each condition
n = 20/condition, mean: 53 months; range: 46-63 months



** difference in responding between All and Some Beads conditions,
p < .01 Fisher’s exact test

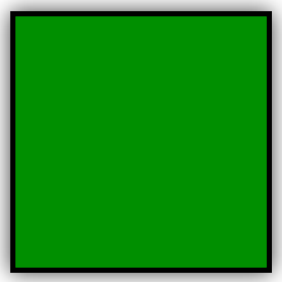
Cook, Goodman, & Schulz, 2011, *Cognition*

Children assigned to one of two training conditions



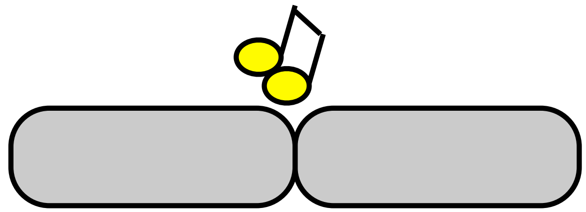
all beads condition

some beads condition

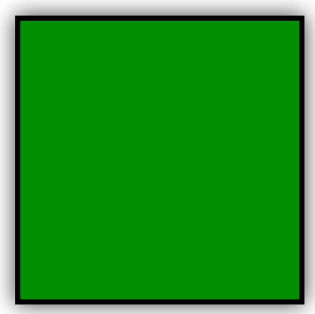


Cook, Goodman, & Schulz, 2011, *Cognition*

All children given the same test condition

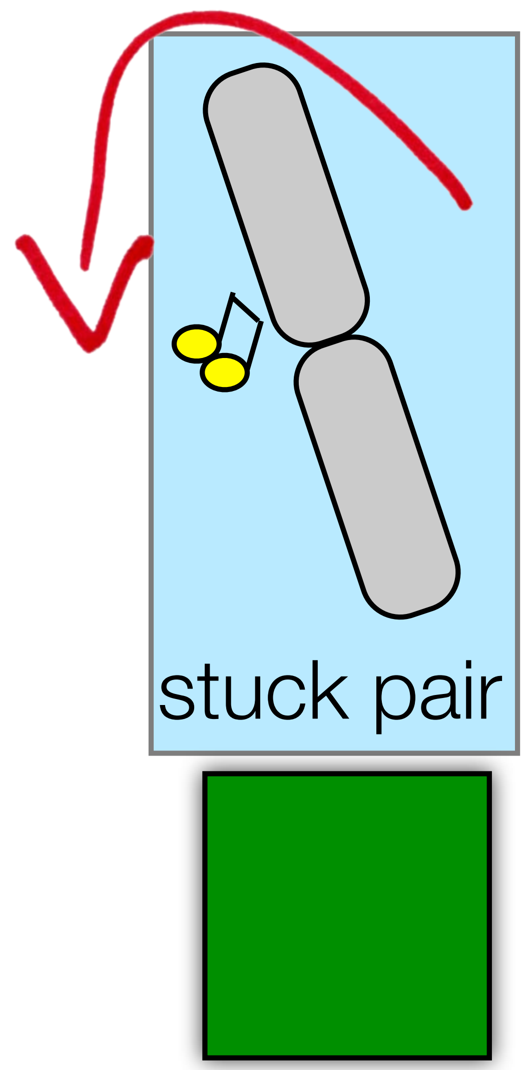


stuck pair



Cook, Goodman, & Schulz, 2011, *Cognition*

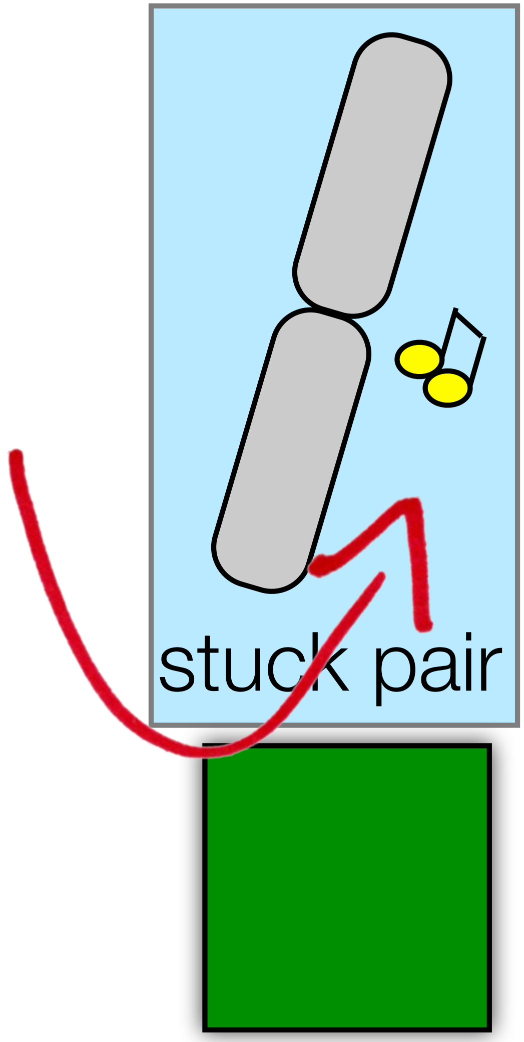
All children given the same test condition



some beads condition

Cook, Goodman, & Schulz, 2011, *Cognition*

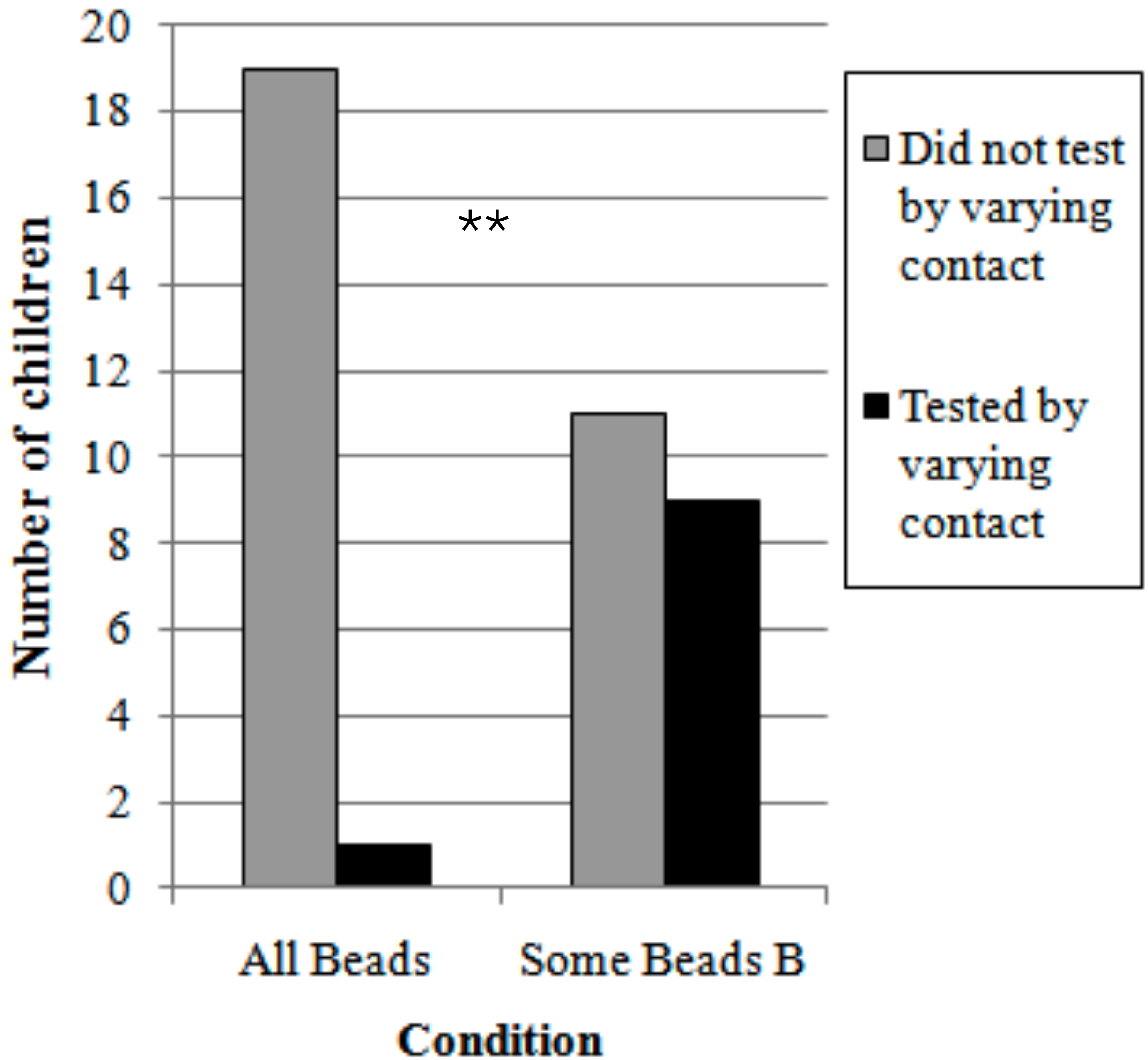
All children given the same test condition



some beads condition

Cook, Goodman, & Schulz, 2011, *Cognition*

Histogram showing number of children generating each outcome
n = 20/condition, mean: 54 months; range: 46-64 months



** difference in responding between All and Some Beads conditions, $p < .01$ Fisher's exact test

Cook, Goodman, & Schulz, 2011, *Cognition*

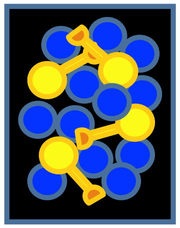
Conclusions

- Preschoolers can use information about the base rate of candidate causes to distinguish the relative ambiguity of evidence.

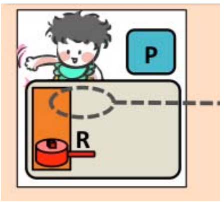
- Given ambiguous evidence, children select -- and design -- potentially informative interventions that isolate relevant causal variables.

Today's talk

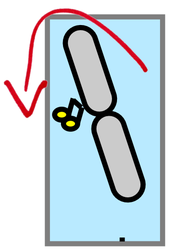
- Children's generalizations depend on how evidence is sampled



- Children infer the relative probability of hypotheses and choose interventions most likely to achieve desired outcomes.



- Children isolate variables to distinguish competing hypotheses



- Children evaluate expert knowledge to decide whether to learn from instruction or exploration

Rational pedagogy

- If you assume that an adult is helpful and knowledgeable …
  - Can assume that evidence they show you is not only true
  - But helps distinguish the target hypothesis from other hypotheses.

Bonawitz, Shafto, Gweon, Spelke, & Schulz, 2011, *Cognition*

Rational pedagogy

- Thus for instance, if a knowledgeable teacher shows you n properties of a toy, should assume that there are not $n + 1$.
- If the same evidence is demonstrated by a naïve learner (or discovered by the child herself), should be much less likely to make this assumption (could well be more than n).
- Pedagogy strengthens the inference that absence of evidence is evidence of absence.

Bonawitz, Shafto, Gweon, Spelke, & Schulz, 2011, *Cognition*

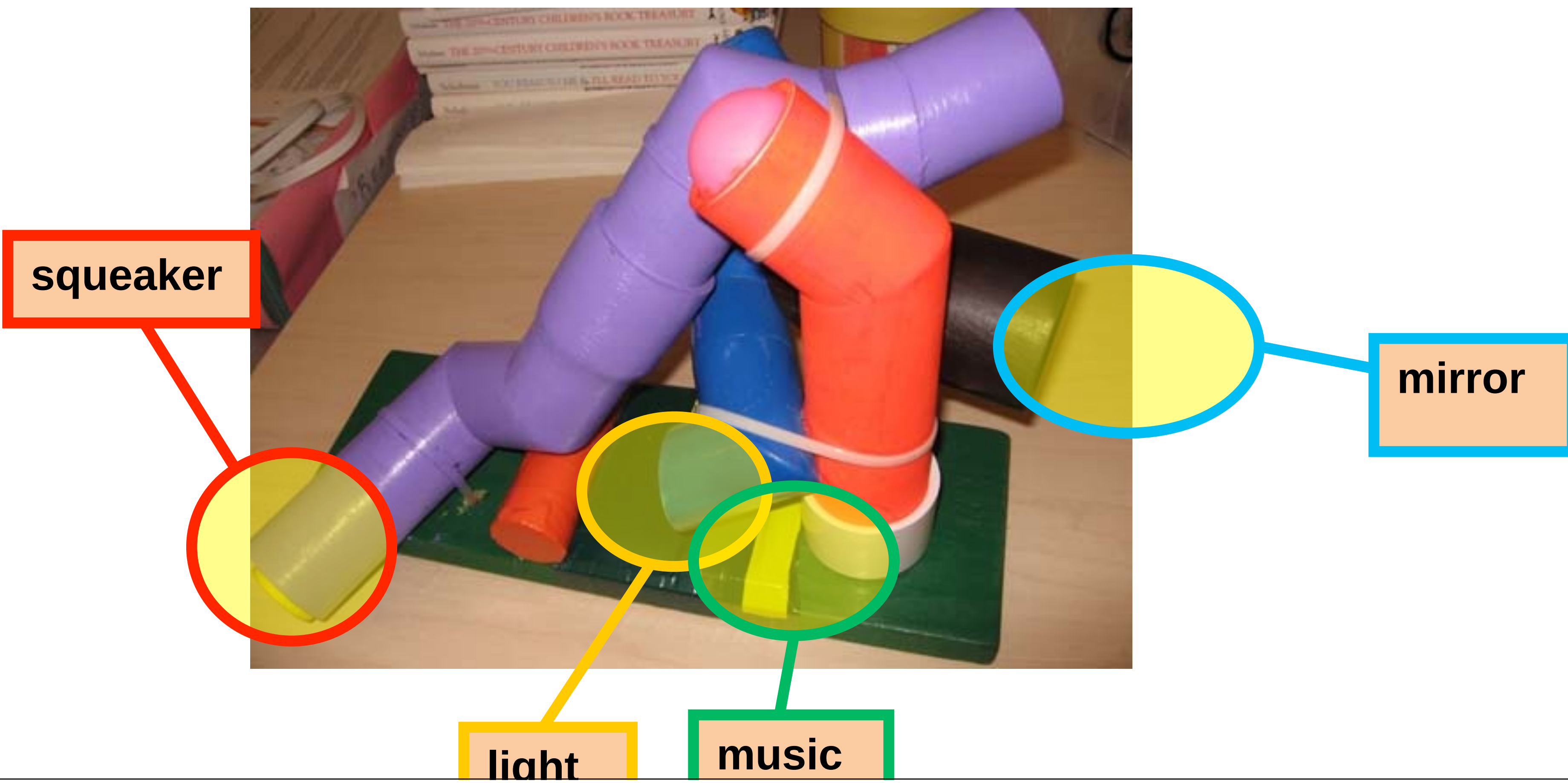
Learning from instruction and exploration

- If a knowledgeable teacher demonstrates properties of a toy, children should not engage in additional exploration.
- If a naïve learner demonstrates the same properties, children should make no such assumption and should explore broadly.

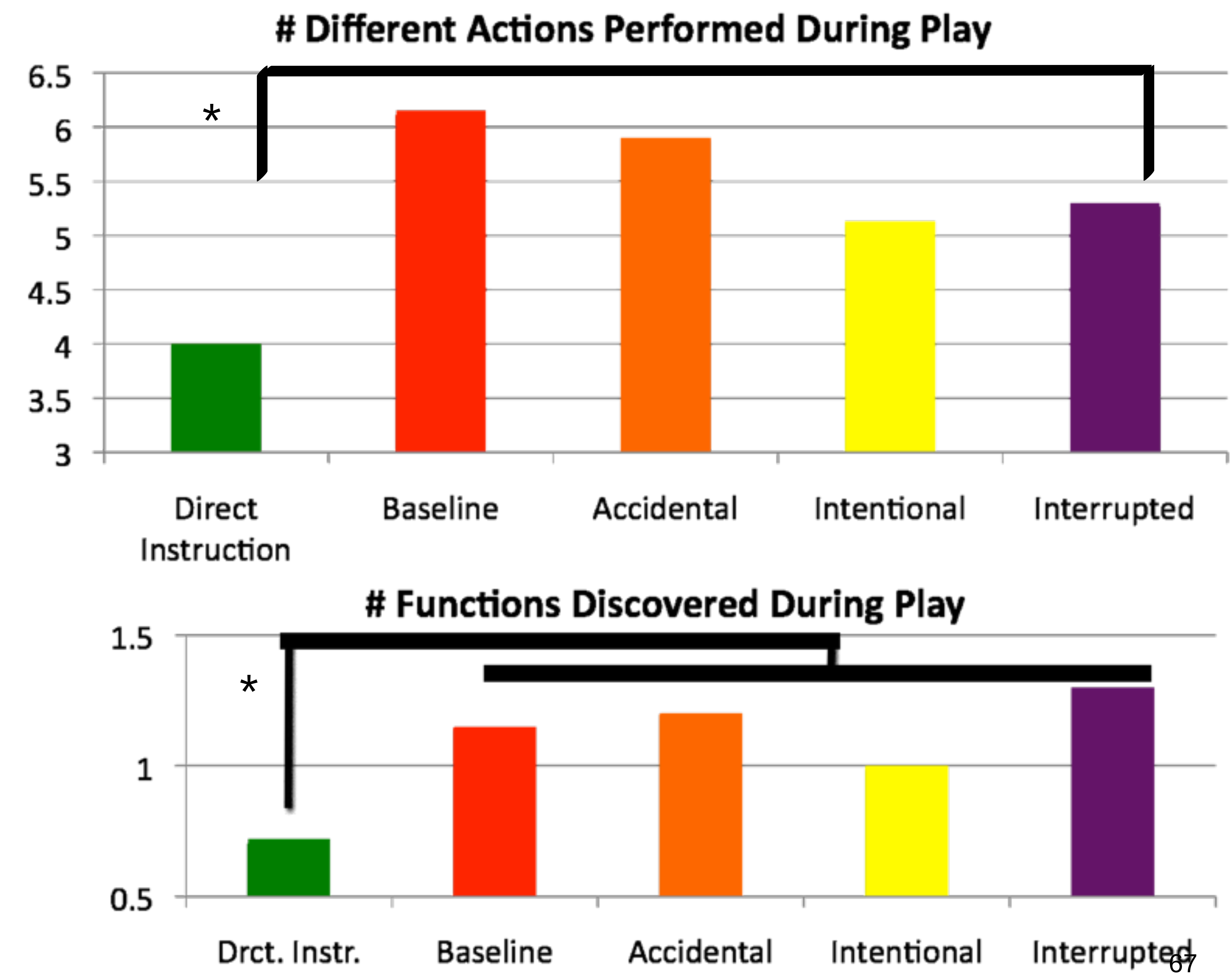
Bonawitz, Shafto, Gweon, Spelke, & Schulz, 2011, *Cognition*

PEDAGOGICAL	ACCIDENTAL	NO DEMO	INTERRUPTED
"Watch this, I'm going to show you my toy." [intentionally pull tube] "Wow, see that?"	"Look at this neat toy I found here." [accidentally pull tube] "Wow, see that?"	"Look at this neat toy that I have." [rotate toy for child] "Wow, see that?"	Identical to Pedagogical except interrupted immediately after "Wow, see that?"

- Four interesting properties



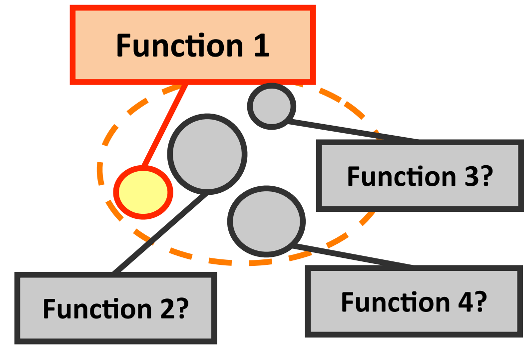
Pedagogical Condition



Different Actions Performed During Play
6.5
6
5.5
5
4.5
4
3.5
3
*
Direct Instruction
Baseline
Accidental
Intentional
Interrupted
Functions Discovered During Play
1.5
1
0.5
*
Drct. Instr.
Baseline
Accidental
Intentional
Interrupted

* $p < .05$

Knowing when learning from others isn't enough...



- When a teacher provides

  insufficient information ....

  – do children recognize "sins of omission" as failures?
  – do sins of omission affect children's judgements of teachers?
  – do children modulate their behavior depending on their evaluation of the teacher?

- Caveat: there are many good reasons for a teacher to provide insufficient information ...
  – the evidence supports generalization
  – exhaustive evidence is too complicated or too extensive for the learner to handle.

- Providing limited information is not always a sin of omission. Here we are going to focus on cases where it is and ask whether children recognize it as such.

Exp 1. Design & Procedure

1. Explore
2. Observe
3. Rate

Teach 1 of 1

Teach 1 of 4

N = 40, 6 – 7 yrs (M = 6.94)

Gweon, Pelton, & Schulz, 2011, *Cog Sci*, and in prep

Exp 1. Design & Procedure

1. Explore
2. Observe
3. Rate

Teach 1 of 1

One-Function Toy

wind-up mechanism

Teach 1 of 4

Four-Function Toy

wind-up mechanism

This is how my toy works! (teach the wind-up mechanism)

1. TOY TEACHER

STUDENT

N = 40, 6 – 7 yrs (M = 6.94)

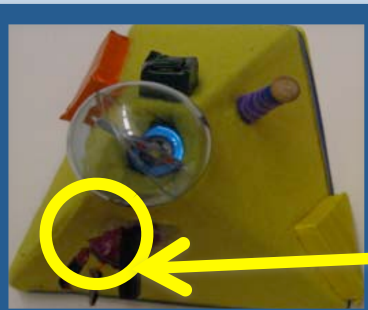
Gweon, Pelton, & Schulz, 2011, *Cog Sci*, and in prep

Exp 1. Design & Procedure

1. Explore
2. Observe
3. Rate

Teach 1 of 1

One-Function Toy



wind-up mechanism

Teach 1 of 4

Four-Function Toy

wind-up mechanism

N = 40, 6 – 7 yrs (M = 6.94)

This is how my toy works! (teach the wind-up mechanism)

1. TOY TEACHER

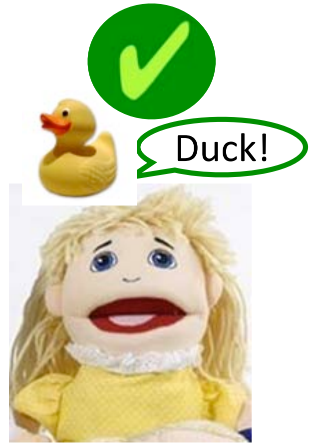
STUDENT

"how helpful was the teacher?"

Gweon, Pelton, & Schulz, 2011, *Cog Sci*, and in prep

- Exclusion criteria

"how helpful was the teacher?"



CORRECT TEACHER

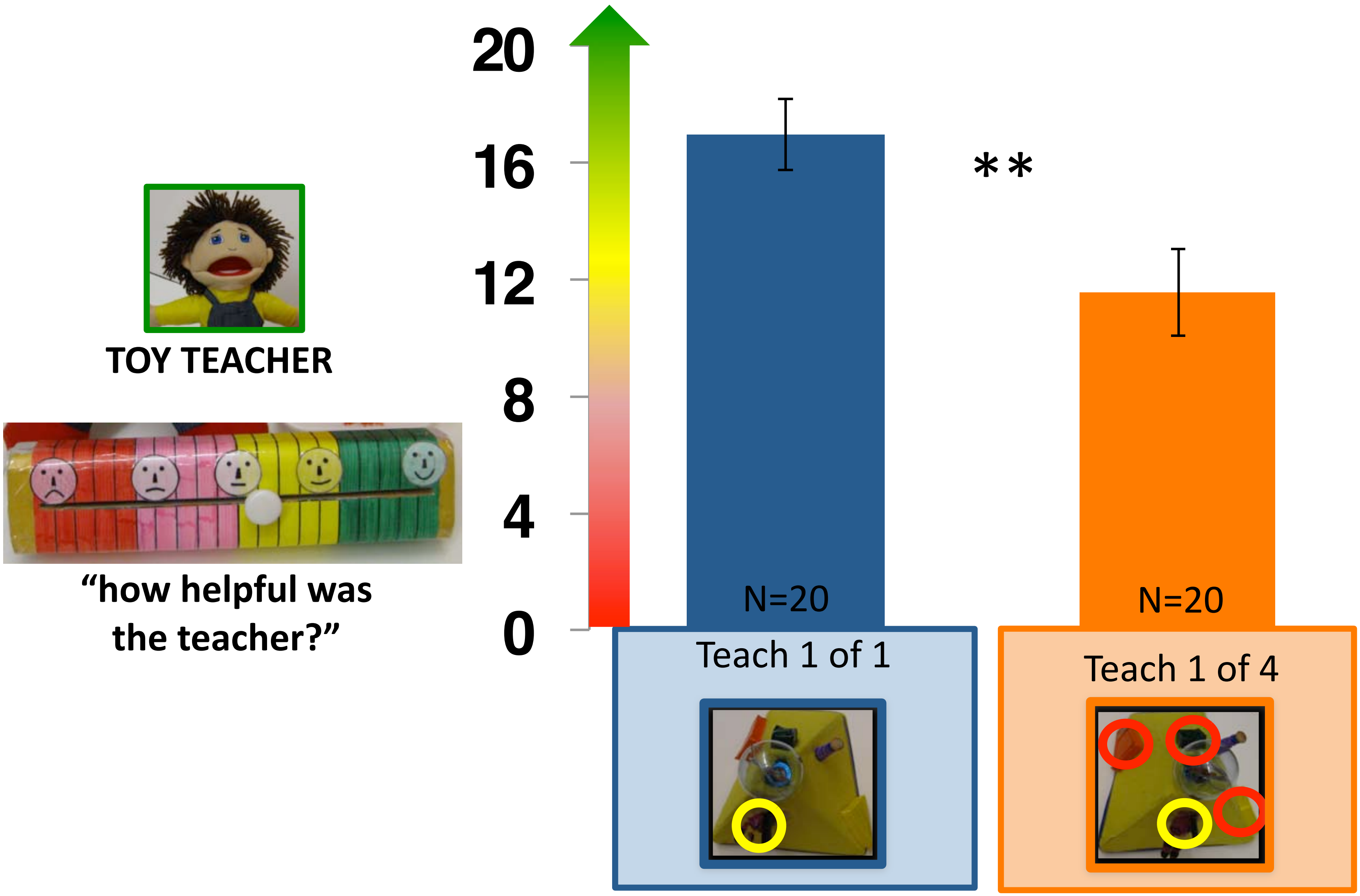
INCORRECT TEACHER

  – Correct Teacher < Incorrect Teacher (N=4)
  – No significant difference between conditions; mean: 14.9 (Correct) vs. 3.4 (Incorrect)

- Prediction
  – Although the teachers provide identical demonstrations in both conditions, children should rate the teacher lower in the "Teach 1 of 4" condition

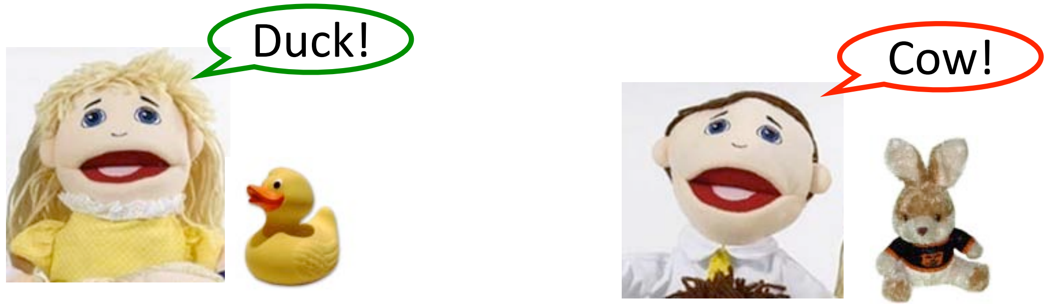
Gweon, Pelton, & Schulz, 2011, *Cog Sci*, and in prep

Ratings for Toy Teacher



**: p <0.01 (Mann-Whitney U)

We want the truth….



…and the *whole* truth.

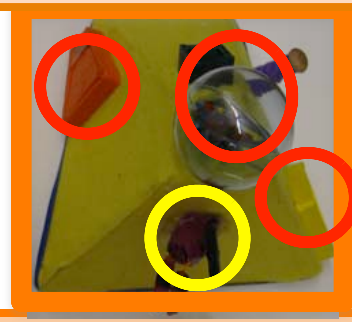
Does children's evaluation of teachers affect ***how*** they learn from them?

Exp 2. Design & Procedure

Teach 1 of 1

One-Function Toy

Teach 1 of 4



Four-Function Toy

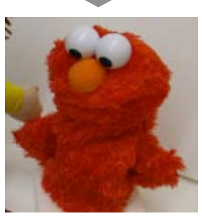
Teach 4 of 4

Four-Function Toy

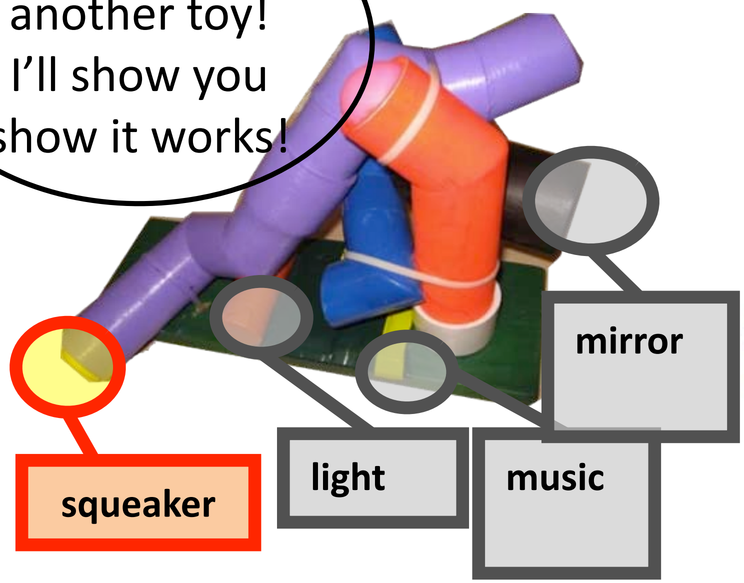
This is how my toy works!

Now, here's another toy! I'll show you show it works!

TOY TEACHER



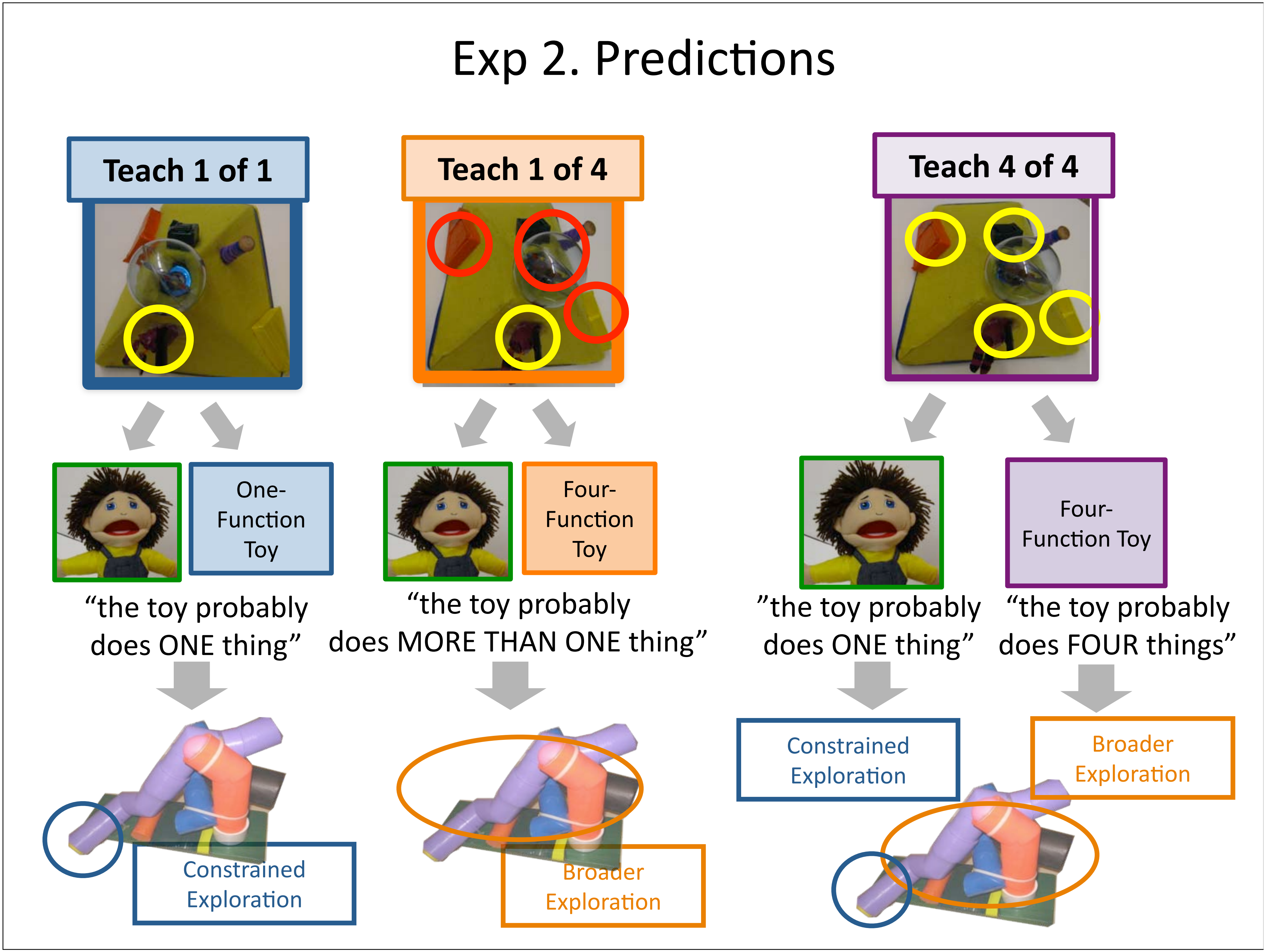
STUDENT



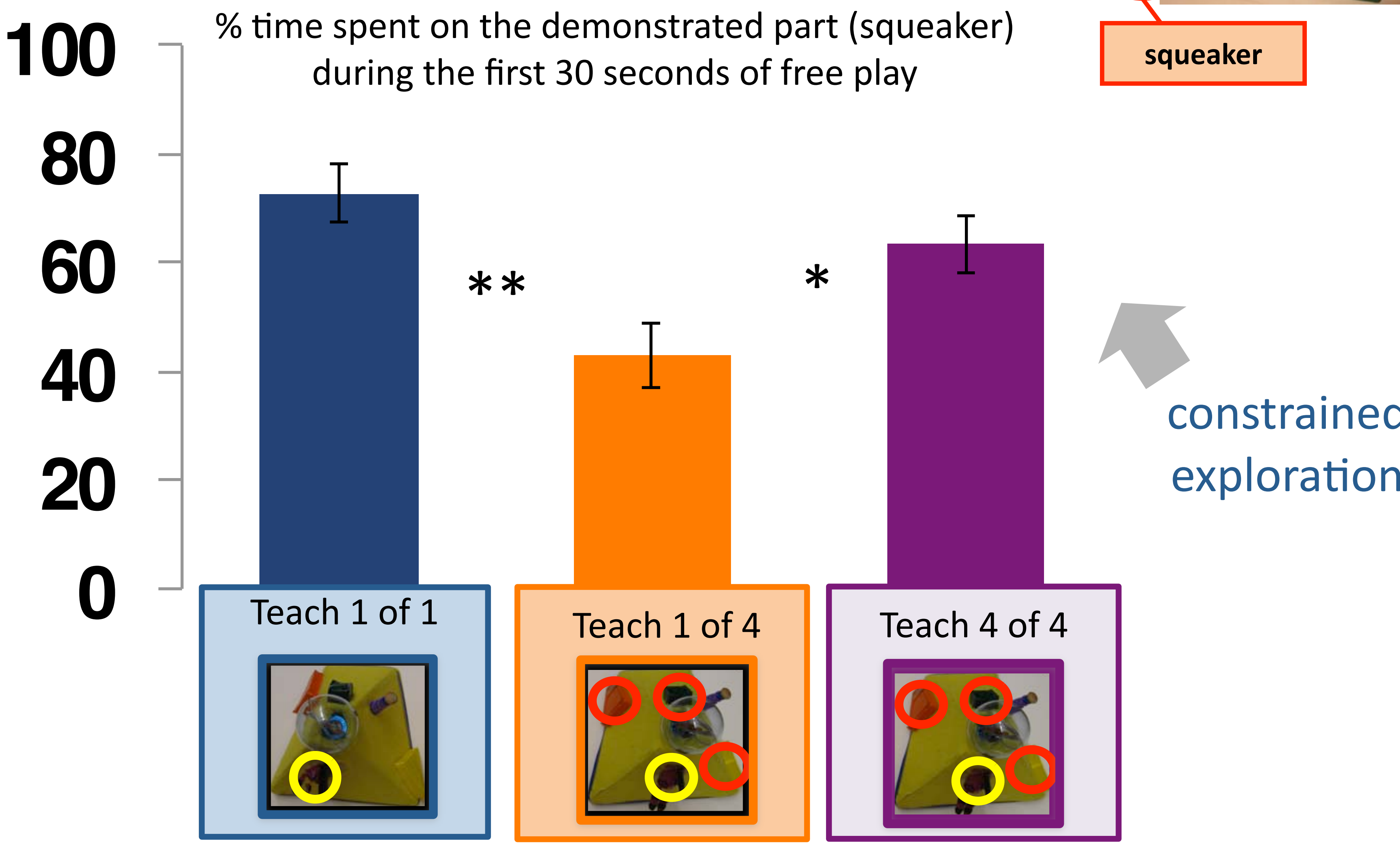
Free Play

6-year-olds (M = 6.45 yrs) N=20 per condition

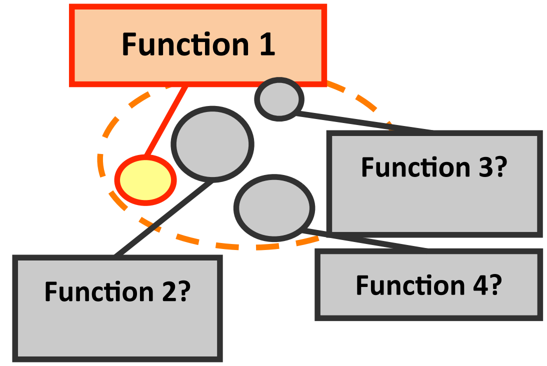
Gweon, Pelton, & Schulz, 2011, *Cog Sci*, and in prep


Exp 2. Predictions
Teach 1 of 1
Teach 1 of 4
Teach 4 of 4
One-Function Toy
Four-Function Toy
Four-Function Toy
“the toy probably does ONE thing”
“the toy probably does MORE THAN ONE thing”
”the toy probably does ONE thing”
“the toy probably does FOUR things”
Constrained Exploration
Broader Exploration
Constrained Exploration
Broader Exploration

Free Play with the 2nd Toy



Pairwise comparisons
* $p < 0.05$, ** $p < 0.005$ Mann-Whitney U

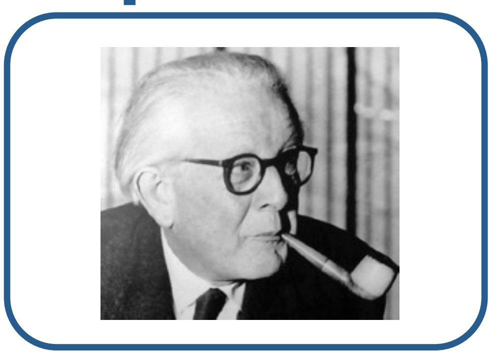


- When the teacher provides

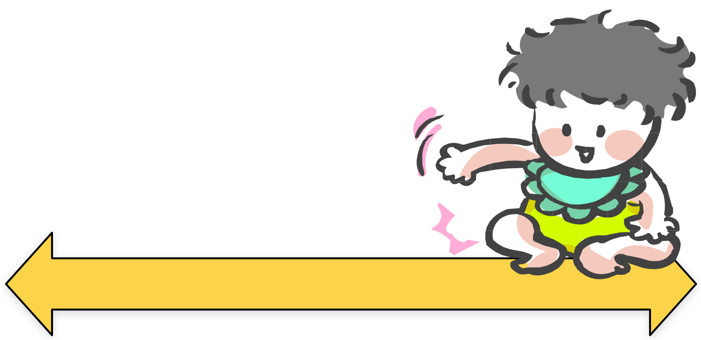
  "Not Enough Information" ….

  - ✓ do children recognize "sins of omission" as failures?
  - ✓ evaluate others accordingly?
  - ✓ modulate their learning based on such evaluations?
  - i.e. more self-guided exploration when in doubt

Learn from Exploration



Piaget (1971)

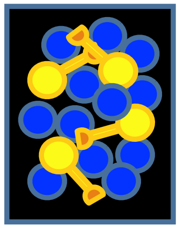


Learn from Others

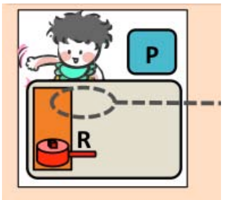
Vygotsky (1978)

Today's talk

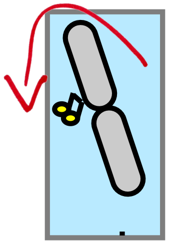
- Children's generalizations depend on how evidence is sampled



- Children infer the relative probability of hypotheses and choose interventions most likely to achieve desired outcomes.



- Children isolate variables to distinguish competing hypotheses

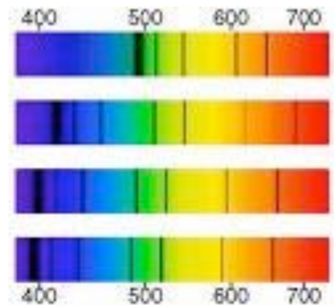


- Children evaluate expert knowledge to decide whether to learn from instruction or exploration

How do children learn?

- Children learn from statistical evidence
- Children's beliefs affect their interpretation of statistical evidence
- Children distinguish genuine causes from spurious associations
- Children selectively explore ambiguous or confounded evidence
- Children introduce unobserved variables to explain data otherwise anomalous with respect to their prior beliefs
- Children's generalizations depend on how evidence is sampled
- Children infer the relative probability of competing hypotheses and choose interventions most likely to change target outcomes
- Children isolate variables to distinguish competing hypotheses
- Children rely on expert knowledge and trade-off instruction and exploration

“There is something fascinating about science, one gets such wholesale returns of conjecture out of such a trifling investment in fact” (Mark Twain, 1883)



Thanks to co-authors Elizabeth Bonawitz, Claire Cook, Noah Goodman, Hyowon Gweon, Hannah Pelton, Pat Shafto, Elizabeth Spelke, and Josh Tenenbaum








members of the ECCL, many outstanding UROP's and participating parents and children at the Museum of Science, Boston and the Boston Children's Museum

Research supported by grants from the James S. McDonnell foundation, the John Merck Scholars Foundation, the National Science Foundation, and the John Templeton Foundation