

# Active Learning Model Results

## The Model

The model we use is a hierarchical model with three levels: the *theory* level, the *hypotheses* level, and the *data* level, and is depicted in figure 1. At the *theory* level, there is a single variable  $t$  indexing the different theories. At the *hypotheses* level, there is one variable  $h_i$  indexing hypotheses for each machine  $i \in \{1, 2, 3\}$ . Finally, *data*  $\mathbf{d}_i$  is split according to which machine it corresponds to, and the bold notation denotes many single datapoints or *actions* (placing a toy against a machine to see whether it lights up or not). Finally, toys and machines have two properties, color and shape, of arbitrary dimensions  $n_c$  and  $n_s$ .

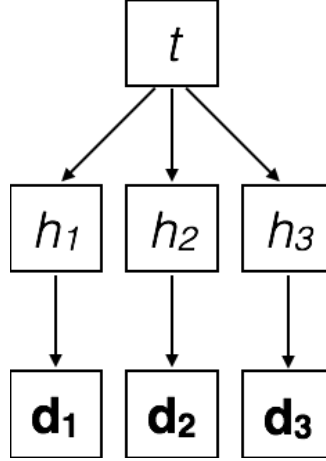


Figure 1: Model structure

## The *theory* level

We consider the following 12 theories:

t	theory
0	No block works
1	Any block works
2	Matching color
3	Matching shape
4	Match color & shape
5	Match color or shape
6	Any specific color
7	Any specific shape
8	Specific color and shape
9	Nonmatching color
10	Nonmatching shape
11	Independent

We take a flat prior for the theories,  $p(t) = 1/12$ .

## The *hypotheses* level

For each machine  $i$ , we consider the following hypotheses, sorted into different kinds:

kind	amount	$h_i$	hypotheses
0	1	0	No block works
1	1	1	Any block works
2	$n_c$	$[2, 2+n_c)$	Specific color
3	$n_s$	$[2+n_c, 2+n_c+n_s)$	Specific shape
4	$n_c n_s$	$[2+n_c+n_s, 2+n_c+n_s+n_c n_s)$	Specific color & shape
5	$n_c n_s$	$[2+n_c+n_s+n_c n_s, 2+n_c+n_s+2n_c n_s)$	Specific color or specific shape
6	$\binom{n_c}{2}$	$[2+n_c+n_s+2n_c n_s, 2+n_c+n_s+2n_c n_s+\binom{n_c}{2})$	Two specific colors
7	$\binom{n_s}{2}$	$[2+n_c+n_s+2n_c n_s+\binom{n_c}{2}, 2+n_c+n_s+2n_c n_s+\binom{n_c}{2}+\binom{n_s}{2})$	Two specific shapes

There are a total of  $n_h = 2 + n_c + n_s + 2n_c n_s + \binom{n_c}{2} + \binom{n_s}{2}$  hypotheses per machine.

## The theory-hypothesis likelihood

Theories are linked to hypotheses for each machine through the following likelihood,  $p(h_i|t)$  (omitting zeros for clarity):

$t/h_i$ kind	0	1	2	3	4	5	6	7
0	1							
1		1						
2			1					
3				1				
4					1			
5						1		
6			$1/n_c$					
7				$1/n_s$				
8					$1/n_c n_s$			
9			$\frac{1}{n_c-1}$				$\frac{1}{\binom{n_c}{2}-(n_c-1)}$	
10				$\frac{1}{n_s-1}$				$\frac{1}{\binom{n_s}{2}-(n_s-1)}$
11	$1/n_h$	$1/n_h$	$1/n_h$	$1/n_h$	$1/n_h$	$1/n_h$	$1/n_h$	$1/n_h$

## The *data* level and the data-hypothesis likelihood

Single datapoints are  $(toy, machine)$  pairs, together with a  $result \in \{\text{True}, \text{False}\}$ , indicating whether the machine lit up or not. Data are thus split according to their corresponding machine, and each machine's data are independent from the rests', conditioned on the corresponding hypothesis. The likelihood then factorizes as follows:

$$p(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3 | h_1, h_2, h_3) = p(\mathbf{d}_1 | h_1) p(\mathbf{d}_2 | h_2) p(\mathbf{d}_3 | h_3). \quad (1)$$

Each machine's data  $\mathbf{d}_i$  in turn a collection of datapoints  $d_i$  corresponding to that machine. These different datapoints are again conditionally independent given the hypothesis, such that we only need look at the likelihood for a single datapoint  $d_i$ :

$$p(d_i | h_i) = \begin{cases} 1 & \text{if } d_i \text{ is compatible with } h_i \\ \epsilon & \text{if it is not,} \end{cases} \quad (2)$$

where  $\epsilon$  is a free noise parameter. Compatibility here means that the toy in  $d_i$  satisfies the criteria specified by  $h_i$  for the machine to activate, if the state of the machine in  $d_i$  is active, or that the toy

fails to satisfy the criteria in  $h_i$ , and the state of the machine is inactive. As an example, a machine being activated with a blue toy is compatible with the *blue toys* hypothesis, as is a machine not being activated by a red toy.

## Full model

With these ingredients, the full joint distribution factors as:

$$p(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, h_1, h_2, h_3, t) = \left[ \prod_{i=1}^3 p(\mathbf{d}_i | h_i) p(h_i | t) \right] p(t) \quad (3)$$

## Active learning

We implement three different active learner models, all based on the model above, but each seeking to maximize information gain at a different level. Although the three learners vary in which distribution they focus on, the general expression for the expected information gain is the same:

$$\text{EIG}(a) = - \sum_{d^{new}} \left( H[p(\dots | d^{new}, \mathbf{d}^{old})] - H[p(\dots | \mathbf{d}^{old})] \right) p(d^{new} | a, \mathbf{d}^{old}), \quad (4)$$

where the sum is over all possible data  $d^{new}$  consistent with action  $a$  (that is, the machine lighting or not), and  $\mathbf{d}^{old}$  is all accumulated data prior to the action choice. The dots in the probability distributions stand for the arguments each learner is concerned with:  $t$  for the *theory* learner,  $h_1, h_2, h_3$  for the *hypotheses* learner, and  $t, h_1, h_2, h_3$  for the *joint* theory-hypotheses learner. Finally, we note that the second term in the difference does not depend on the action choice, yielding a constant term that we can omit, leaving the learners with the problem of minimizing the expected final entropy:

$$\text{EFE}(a) = \sum_{d^{new}} H[p(\dots | d^{new}, \mathbf{d}^{old})] p(d^{new} | a, \mathbf{d}^{old}). \quad (5)$$

Working with this quantity is equivalent to using the EIG when taking differences, since the constants cancel out.

The  $p(d^{new} | a, \mathbf{d}^{old})$  in the equations above is the probability of finding each piece of data  $d^{new}$  given an action  $a$  and prior experience  $\mathbf{d}^{old}$ , and is the same for all models:

$$\begin{aligned} p(d^{new} | a, \mathbf{d}^{old}) &= \sum_{h_i} p(d_i^{new} | a, h_i, \mathbf{d}^{old}) p(h_i | a, \mathbf{d}^{old}) = \\ &= \sum_{h_i} p(d_i^{new} | a, h_i) \sum_t p(h_i | \mathbf{d}^{old}, t) p(t | \mathbf{d}^{old}) = \\ &= \sum_{h_i, t} p(d_i^{new} | a, h_i) p(h_i | \mathbf{d}_i^{old}, t) p(t | \mathbf{d}^{old}) = \\ &= \sum_{h_i, t} p(d_i^{new} | a, h_i) \frac{p(\mathbf{d}_i^{old} | h_i) p(h_i | t)}{p(\mathbf{d}_i^{old} | t)} p(t | \mathbf{d}^{old}), \end{aligned} \quad (6)$$

where some index magic took place. First, we converted  $d^{new}$  to  $d_i^{new}$ , which only amounts to saying that a single datapoint can only correspond to an individual machine, labeled  $i$ , which is the one whose hypotheses we sum over on the RHS. Second, from the second to the third line, we see  $\mathbf{d}^{old}$  reduced to  $\mathbf{d}_i^{old}$ . This is because of hypothesis  $h_i$  being independent of data in other machines given the theory. Finally, we used Bayes' theorem and the fact that data are independent of the theory given the hypothesis in the corresponding machine to go from the third to the last line. This leaves us with

two more things to compute in (6): the theory posterior  $p(t|\mathbf{d})$ , and  $p(\mathbf{d}_i|t)$ . This last probability is given by:

$$p(\mathbf{d}_i|t) = \sum_{h_i} p(\mathbf{d}_i|h_i)p(h_i|t), \quad (7)$$

whereas the expression for the theory posterior is shown below.

### Theory learner

The theory learner seeks to minimize the EFE given in equation (5) for the theory posterior:

$$p(t|\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) \propto p(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3|t)p(t) \quad (8)$$

with

$$p(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3|t) = \prod_i \sum_{h_i} p(\mathbf{d}_i|h_i)p(h_i|t). \quad (9)$$

### Hypotheses learner

The hypotheses learner seeks to minimize the EFE given in equation (5) for the hypotheses posterior:

$$p(h_1, h_2, h_3|\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) \propto p(\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3|h_1, h_2, h_3) = \sum_t \left( \prod_i p(\mathbf{d}_i|h_i)p(h_i|t) \right) p(t). \quad (10)$$

### Joint learner

The joint theory/hypotheses learner seeks to minimize the EFE given in equation (5) for the joint posterior:

$$p(h_1, h_2, h_3, t|\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3) \propto \left( \prod_i p(\mathbf{d}_i|h_i)p(h_i|t) \right) p(t). \quad (11)$$

## Results

Summary of results so far. For all learners, we will evaluate the difference in EIG between the kids and the optimal model learner (red in the histograms, y axes in scatter plots), as well as the difference between a random choice model and the optimal learner (blue and x axes). We study these differences for each action in the playing sequence, feeding the model learners the actual kids' gameplay up to the given choice. All random results are averaged over 20 realizations, and we set a value of  $\epsilon = 0.001$ .

### Theory learner

The results for this learner are presented in figures 2 and 3.

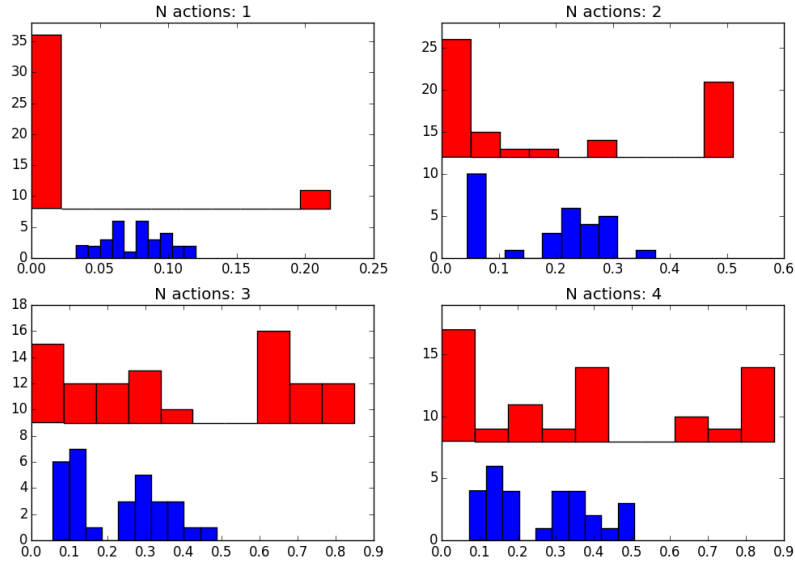


Figure 2: Theory model EIG histograms. Kids - optimal (red) and random - optimal (blue)

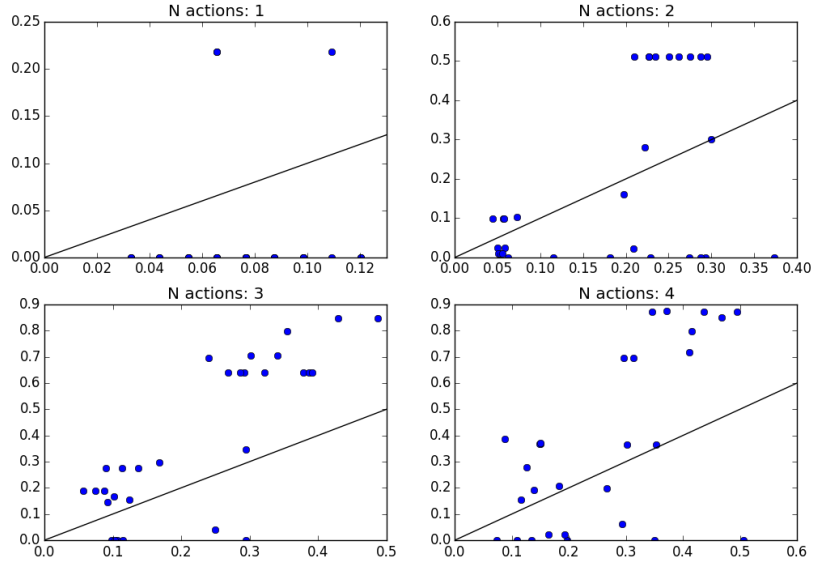


Figure 3: Theory model EIG scatterplots. Kids - optimal (y axes) vs random - optimal (x axes). Kids below the identity line perform better than random.

## Hypotheses learner

The results for this learner are shown in figures 4 and 5.

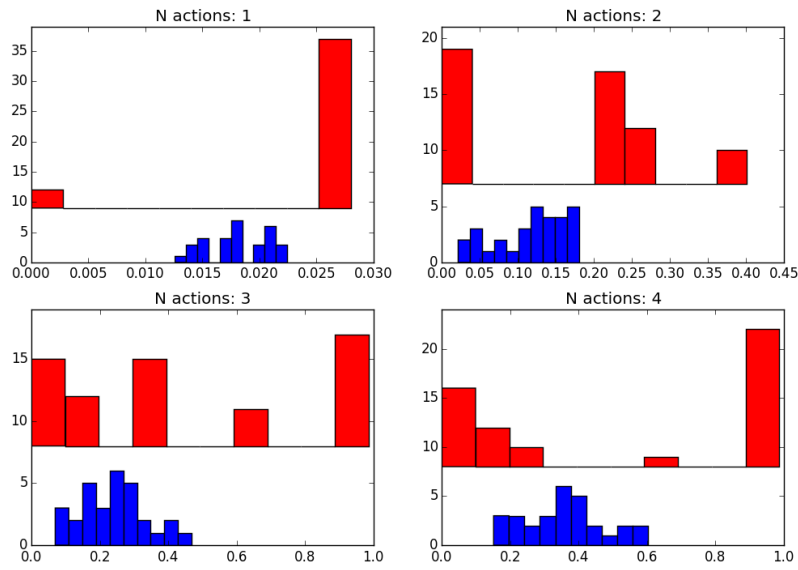


Figure 4: Theory model EIG histograms. Kids - optimal (red) and random - optimal (blue)

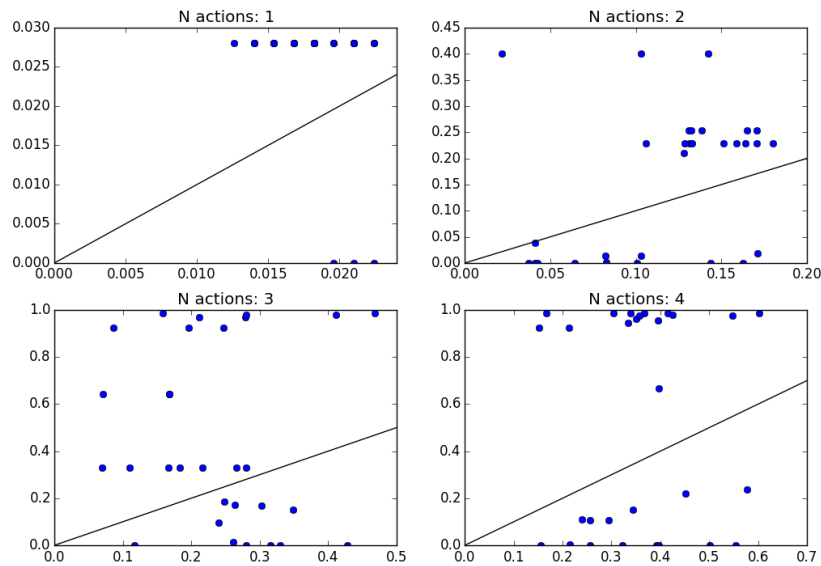


Figure 5: Hypotheses model EIG scatterplots. Kids - optimal (y axes) vs random - optimal (x axes). Kids below the identity line perform better than random.

## Joint learner

The results for this learner are presented in figures 6 and 7.

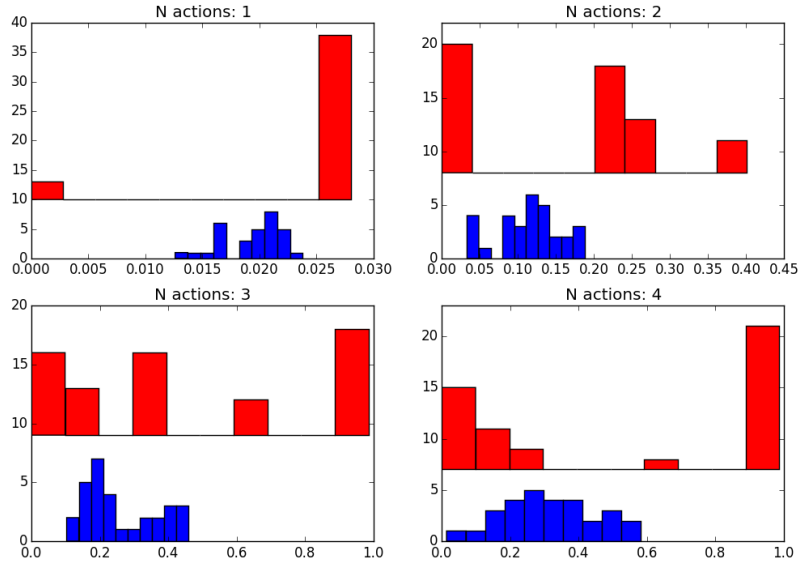


Figure 6: Joint model EIG histograms. Kids - optimal (red) and random - optimal (blue)

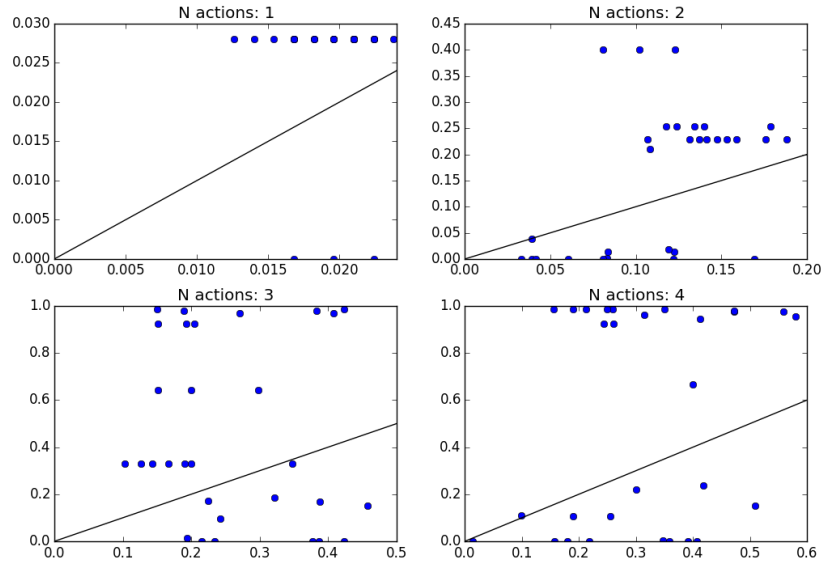


Figure 7: Joint model EIG scatterplots. Kids - optimal (y axes) vs random - optimal (x axes). Kids below the identity line perform better than random.

## All learners

A summary of the results through paired  $t$ -test for the different actions and different models is shown in the following table. Positive  $t$  means kids better than random.

model	theory		hypotheses		joint	
# actions	$t$	$p$	$t$	$p$	$t$	$p$
1	4.473	0.000	-4.089	0.000	-3.718	0.001
2	-0.301	0.766	-1.845	0.075	-2.064	0.048
3	-4.292	0.000	-2.939	0.006	-2.635	0.013
4	-2.060	0.049	-1.925	0.064	-2.602	0.015

## Varying $\epsilon$

We varied the parameter  $\epsilon$  for the theory model to analyze how the predictions changed. The results are shown in figure 8. In the figure the  $t$  statistics (green) and  $p$  values (purple) for the test in the previous section are shown as a function of  $\epsilon$ , for different action numbers.  $t$  values below 0 (black dotted line) indicate kids doing better than chance, with significance assessed by the corresponding  $p$  value. Kids are better than chance in their first action choice for all values of  $\epsilon$ . For other action numbers, they are always worse than chance or similar to it apart from the second action, in which they are better for  $\epsilon = 0.25$ .

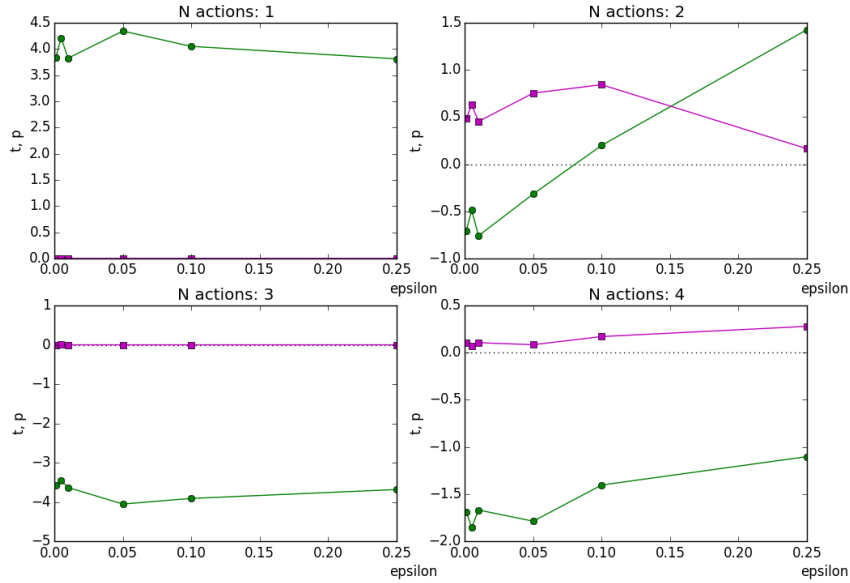


Figure 8: Theory model  $\epsilon$  dependence.  $t$  statistic (green) and  $p$  value (purple) for tests in previous section as a function of  $\epsilon$ , for the different action numbers (panels).

## Scrutinizing (second) action choices

We look at kids' choice of second action, depending on their first choice. We categorize the first choice as successful matches (a toy/machine match with the correct property according to the specific condition,  $a_1 = +$ ), unsuccessful matches (a toy/machine match with the opposite property,  $a_1 = -$ )



and nonmatches (toy and machine different in color and shape, machine always inactive,  $a_1 = N$ ). The second choice can be split according to 6 categories as follows (I use numbers for notation). Same property as first action, different toy and machine:  $a_2 = 0$ . Same machine as  $a_1$ , but matching the opposite property,  $a_2 = 1$ . Same toy as  $a_1$ , opposite property  $a_2 = 2$ . Same machine as  $a_1$ , no match between toy and machine,  $a_2 = 3$ . Same toy as  $a_1$ , no match between toy and machine,  $a_2 = 4$ . Different toy and machine than  $a_1$ , nonmatching either,  $a_2 = 5$ . Note that categories  $a_2 = 3$  and  $a_2 = 4$  are unsatisfiable for  $a_1 = n$ , and that the number of actions that feeds each category varies between 1 and 2, what must be taken into account when considering random choices.

In the following table, we sort kids' second choice of action as a function of their first one. Entries are number of kids in each condition. 30 kids are considered in total, dropping one that only made one action:

$a_2/a_1$	+	-	N
0	10	3	1
1	5	3	2
2	0	1	0
3	1	2	0
4	1	0	0
5	0	1	0

The theory model gives the following expected final entropies for these actions:

$a_2/a_1$	+	-	N
0	1.843	2.084	2.731
1	<b>1.331</b>	1.995	<b>2.570</b>
2	1.354	<b>1.986</b>	2.731
3	1.612	2.011	—
4	1.632	2.089	—
5	1.632	2.089	2.906

where we marked in bold the correct second action given the first. The model essentially tells us to try the property opposite to that already tried, whether we were successful or not on the first try. Only 8 kids out of 30 get the perfect choice, and 12 out of 30 if we include marginally suboptimal choices. The remaining kids make bad choices. In particular, the 10 kids with  $a_1 = +$  and  $a_2 = 0$  choose to repeat their successful property match, but either changing the toy or the machine, which is the worst possible action to take according to the model.

Setting  $\epsilon = 0.25$ , the best fitting value for the second action choice, we get for the entropies:

$a_2/a_1$	+	-	N
0	2.547	2.563	3.052
1	<b>2.506</b>	2.543	<b>3.031</b>
2	2.508	<b>2.541</b>	3.052
3	2.572	2.573	—
4	2.571	2.581	—
5	2.571	2.581	3.091

We see that the model still suggests the same actions, which speaks about its robustness. Other action combinations vary in value, which explains how children can in this setting be better than chance.

The hypotheses model predicts, for the extreme values of  $\epsilon$  considered:

	$\epsilon = 0.001$			$\epsilon = 0.25$		
$a_2/a_1$	+	-	N	+	-	N
0	4.332	5.792	5.186	6.023	6.571	6.509
1	4.355	<b>5.391</b>	<b>5.147</b>	6.057	<b>6.496</b>	<b>6.501</b>
2	4.313	<b>5.391</b>	5.186	6.052	6.497	6.509
3	<b>4.102</b>	5.405	–	<b>6.013</b>	6.499	–
4	4.120	5.392	–	6.017	<b>6.496</b>	–
5	4.120	5.392	5.225	6.017	<b>6.496</b>	<b>6.501</b>

The choices are essentially the same for the two values of  $\epsilon$ , since the numerical differences are minimal, which shows again the robustness of the model. Most choices again are to the theory model, with the notable exception of  $a_1 = +$ . This model suggests choosing a nonmatching toy with the same machine, a pattern that only a single kid follows. The action that most kids after  $a_1 = +$  is now not the worst, but clearly worse than the optimal and other alternatives.

Finally, the joint model predicts, for the extreme values of  $\epsilon$ :

	$\epsilon = 0.001$			$\epsilon = 0.25$		
$a_2/a_1$	+	-	N	+	-	N
0	4.407	5.913	5.324	6.111	6.685	6.633
1	4.431	<b>5.512</b>	<b>5.286</b>	6.145	6.611	6.625
2	4.388	<b>5.512</b>	5.324	6.139	6.611	6.633
3	<b>4.177</b>	5.526	–	<b>6.101</b>	6.613	–
4	4.195	5.513	–	6.105	<b>6.610</b>	–
5	4.195	5.513	5.364	6.105	<b>6.610</b>	<b>6.624</b>

Again, the choices are similar for the two values of  $\epsilon$ . The choices again generally repeat the pattern of the theory model, apart from the case of  $a_1 = +$ . The joint model, as the theory one, considers the action that most kids take after  $a_1 = +$  as the worst possible one.

## Phase 2

We now look at the data from the second phase of the experiment, in which kids are tested for specific choices. This second phase is divided in two parts: a) with one of the prior machines and novel objects, with shapes and colors previously seen in the active learning phase and b) with novel objects and a novel machine, all of new shapes and colors.

### Effect of active learning

We first consider how responses in phase 2 are affected by the active learning phase. For this, we perform logistic regressions for the responses in phases 2 a) and b) with the EIG achieved by each kid in the active learning phase, as before, split in the different first four actions. The results are shown in figures 9, 10 and 11 for the three models considered.

As before, the relative EIG is close to 0 when the kids are performing close to the corresponding optimal model. A positive effect of the active learning phase in the test phase would be seen in these graphs as a decreasing relation, that is, more correct responses for smaller values of EIG. The tendency in the majority of the cases is however the opposite: performance shows a slight increase with increasing relative EIG. These figures are for  $\epsilon = 0.001$ , a value of  $\epsilon = 0.25$ , at the other extreme of the range explored, gives similar results.

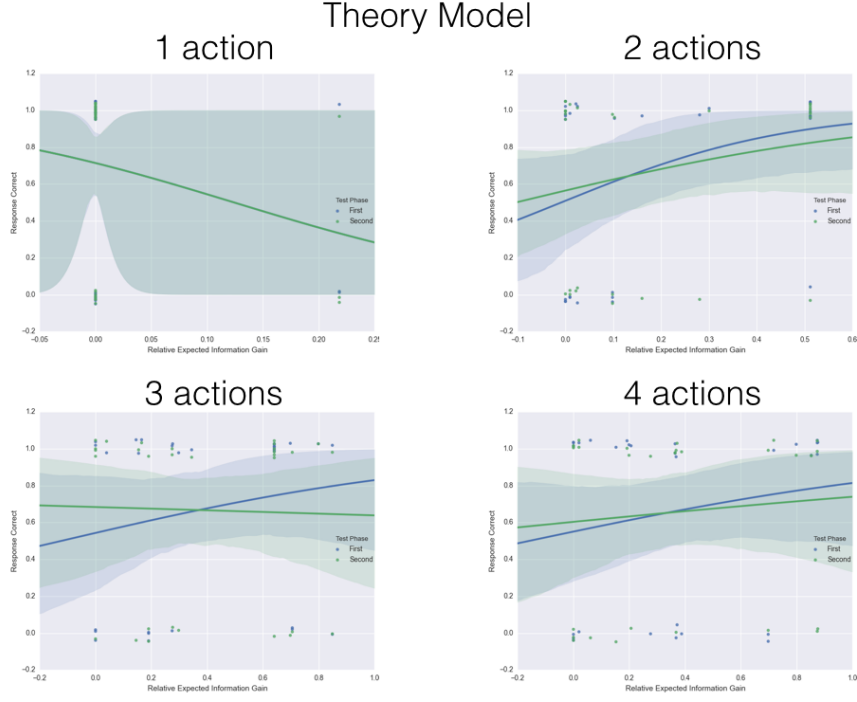


Figure 9: Correctness of response in phases 2a (blue) and 2b (green) as a function of EIG relative to theory model in the active learning phase, for the four first actions in this phase.

## Predictive posterior

The results in the second phase can also be analyzed in terms of the models' predictions, taking into account the experience gathered in the active learning phase. Here, we take the models' choices as those that maximize the probability of activation among the options available to children. Apart from a couple of isolated cases, the experience of the active learning phase allows the models to make perfect predictions in the second phase, so that a direct comparison with children is moot. For phase 2b, the huge dimensionality of the model makes it impossible to run in practice.

## Success Probability

We now go back to the first phase of the experiment, and consider it in terms of children not maximizing an information gain but instead maximizing the probability of machine activation. We keep filtering identical consecutive actions, which are the maximizing choice, both in the data and in the models, which amounts to including some amount of exploration in the kids' gameplay.

The probability of machine activation (success) for a specific action is given by equation (6), and is the same for all models. In fact, these probabilities are fixed for the available set of actions, given the prior experience. In order to compare kids performance with a random choice, we define:

$$\mathcal{L}_{\mathcal{K}} = -\log \left( \frac{p_K^+}{p_M^+} \right) \quad (12)$$

for the kids and

$$\mathcal{L}_{\mathcal{R}} = -\log \left( \frac{p_R^+}{p_M^+} \right) \quad (13)$$

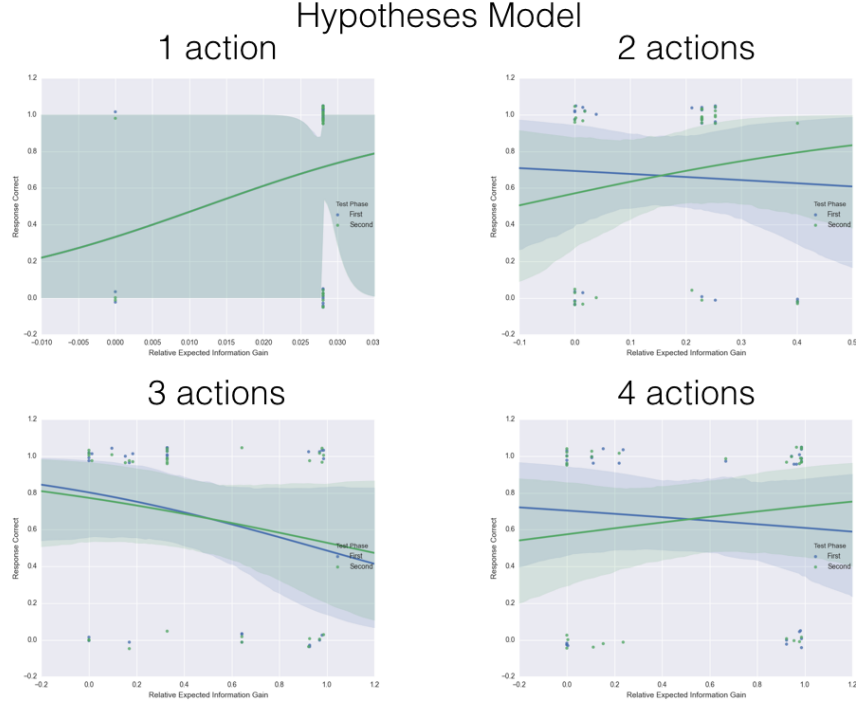


Figure 10: Correctness of response in phases 2a (blue) and 2b (green) as a function of EIG relative to hypotheses model in the active learning phase, for the four first actions in this phase.

for the random choice, where  $p^+$  are the probabilities of success for the action chosen for the kids ( $K$ ), model ( $M$ ) and random ( $R$ ).

In figure 12 we show the distribution of these variables for the first four actions ( $\epsilon = 0.001$ ). As before, children are plotted in red ( $\mathcal{L}_K$ ) and the random model in blue ( $\mathcal{L}_R$ ). A value close to 0 indicates play consistent with activation probability maximization.

In mean value, children are consistently worse than random for this measure too. We can try to get rid of the effect of high values by considering the fraction of kids that are better than random for each action number, and test this against chance with a binomial of proportion equal to 1/2. The results are as follows:

action	$n_{better}$	$n_{total}$	p
1	3	31	$< 10^{-5}$
2	19	30	0.200
3	15	30	1
4	19	29	0.136

where  $n_{better}$  is the number of kids that are better than random and  $n_{total}$  is the total number of kids for that action number (which decreases because some kids performed a smaller amount of actions). We see that for the second and fourth actions a non-significant majority of kids are better than random, and it is tied for the third. In the first actions most kids perform worse than random, but this is likely an artifact of the model's priors.

Finally, we can try to interpret the apparent bimodality of the plots in figure 12 as a division in a “player” population which tries to maximize the probability of activation, and a “learner” population which tries to maximize the information gain. The problem with this is that these populations do not

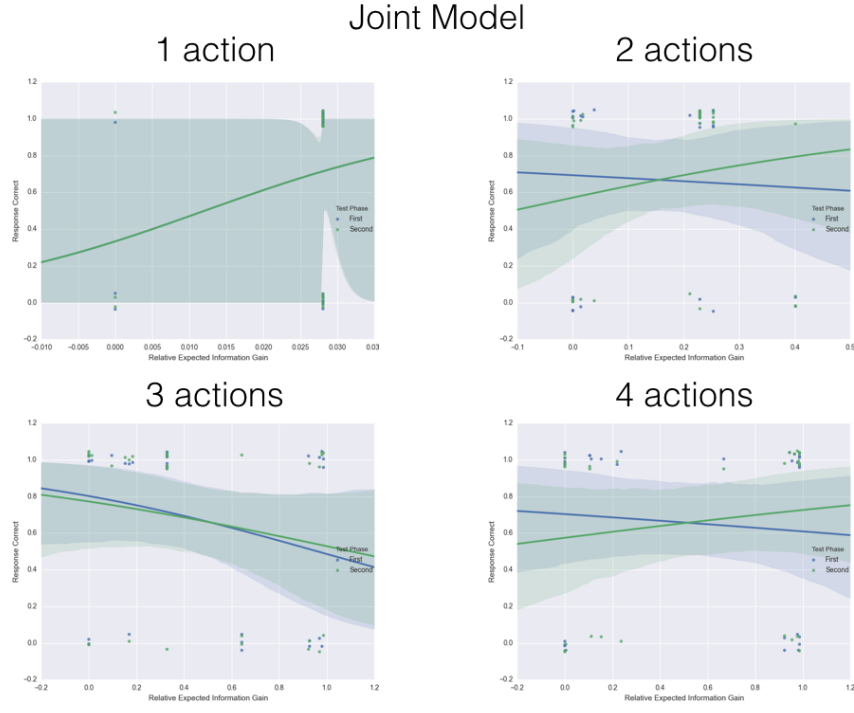


Figure 11: Correctness of response in phases 2a (blue) and 2b (green) as a function of EIG relative to joint model in the active learning phase, for the four first actions in this phase. y values (0: incorrect, 1: correct) are jittered for clarity.

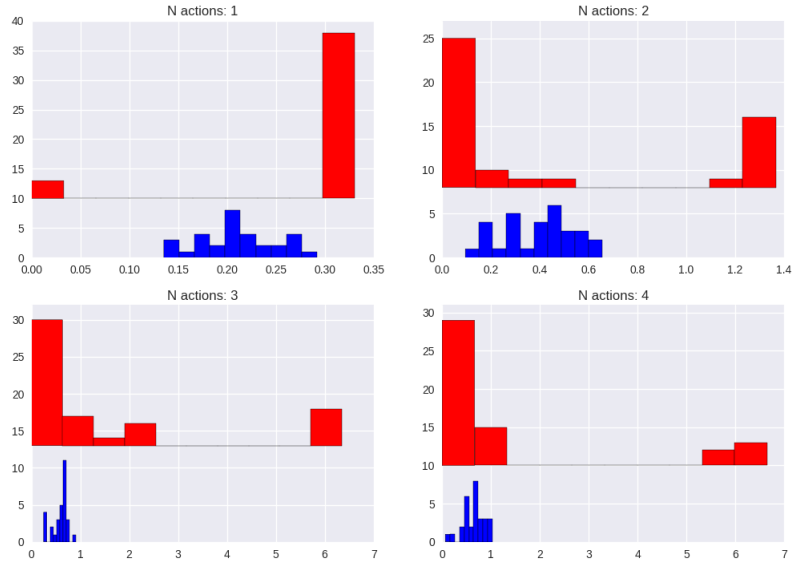


Figure 12: Distributions of  $\mathcal{L}_K$  (red) and  $\mathcal{L}_R$  (blue).

seem to be consistent for the different actions. The set of children that perform in a range of 10% of the worst  $\mathcal{L}$  score in the second action (8 kids) has no overlap with that in the third (5 kids). In the fourth action (3 kids) 2 of the kids in the second action reappear, but none of those in the third.