Children search for information as efficiently as adults, but seek additional confirmatory evidence.
1 Motivation
2 Hierarchical Task
3 Pre-requisites
4 Question Asking
5 Intervention
6 Discussion
7 Critique
Outline for section 1

1. Motivation
2. Hierarchical Task
3. Pre-requisites
4. Question Asking
5. Intervention
6. Discussion
7. Critique
Motivation

- How do children and adults search for information to find which objects share a causal property?
- Are adults more efficient than children?
- Do they narrow down on hypothesis as fast as adults?
- Do they continue asking questions after they have sufficiently narrowed down the hypotheses space [RL15]?
Causal inference involves categorizing objects and finding out what level a property applies.

Consider a hierarchy \textit{Furniture} $\rightarrow$ \textit{Lamps} $\rightarrow$ \textit{Table Lamps} and the causal property is “produce light”.

\textit{Furniture} is too general an answer and \textit{Table Lamps} are too specific an answer.

Finding the right category involves associating a set of attributes with it.
Previous Work

Some work done before [CHM07] [Leg+13] [Mos+66] [RL15] concluded that the ability to ask questions undergoes a large developmental change from age 4 to age 11.
Previous Work

Some work done before [CHM07] [Leg+13] [Mos+66] [RL15] concluded that the ability to ask questions undergoes a large developmental change from age 4 to age 11.

But does it really?
This work aims to answer a few more questions.

- Are adults more efficient than children when it comes to asking questions?
  - Is the most efficient search path used?
  - Is a good stopping rule—which decides when to stop seeking information—used?
Fine-grained analysis

This work aims to answer a few more questions.

- Are adults more efficient than children when it comes to asking questions?
  - Is the most efficient search path used?
  - Is a good stopping rule—which decides when to stop seeking information—used?
- Are results dependent on if the task is *Question-Asking* (Study 1) or *Intervention* (Study 2) (where single objects are tested sequentially)?

Presented by: Ameet Deshpande
This work aims to answer a few more questions.

- Are adults more efficient than children when it comes to asking questions?
  - Is the most efficient search path used?
  - Is a good stopping rule—which decides when to stop seeking information—used?

- Are results dependent on if the task is *Question-Asking* (Study 1) or *Intervention* (Study 2) (where single objects are tested sequentially)?

- Is there a role of the Hierarchical structure in the questions asked?
Outline for section 2

1 Motivation

2 Hierarchical Task

3 Pre-requisites

4 Question Asking

5 Intervention

6 Discussion

7 Critique
Motivation for the Hierarchical Task

- Most studies mentioned before use 20-questions style tasks. Try out Akinator.
- Hierarchies are more complex and realistic [Rug+16].
- Will allow one to test if more abstract hypotheses are used to arrive at the answer.
## Modified 20-Questions Game

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Superordinate category</th>
<th>Basic-level category</th>
<th>Subordinate category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>Fish</td>
<td>Goldfish</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Clownfish</td>
<td></td>
</tr>
<tr>
<td>Planet Apres</td>
<td>Birds</td>
<td>Parrots</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Owls</td>
<td></td>
</tr>
<tr>
<td>Plants</td>
<td>Trees</td>
<td>Apple trees</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pine trees</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flowers</td>
<td>Tulips</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Daisies</td>
<td></td>
</tr>
<tr>
<td>Clothes</td>
<td>Shirts</td>
<td>Long sleeves</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Short sleeves</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Shoes</td>
<td>Flip-flops</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Boots</td>
<td></td>
</tr>
<tr>
<td>Machine</td>
<td>Tables</td>
<td>Dining tables</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Desk</td>
<td></td>
</tr>
<tr>
<td>Furniture</td>
<td>Chairs</td>
<td>Rocking chair</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>High chair</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cars</td>
<td>Vans</td>
<td></td>
</tr>
<tr>
<td>Vehicles</td>
<td>Planes</td>
<td>Sportcars</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Planes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Helicopters</td>
<td></td>
</tr>
<tr>
<td>Magic-box</td>
<td>Apples</td>
<td>Yellow apples</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Green apples</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Berries</td>
<td>Raspberries</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Blueberries</td>
<td></td>
</tr>
</tbody>
</table>
16 objects are given to the subjects on an iPad® screen

The subjects can either ask questions or choose objects depending on the task

The answer can be “yes”, “no” or “some”
## Modified 20-Questions Game

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Superordinate category</th>
<th>Basic-level category</th>
<th>Subordinate category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planet Apres</td>
<td>Animals</td>
<td>Fish</td>
<td>Goldfish, Clownfish</td>
</tr>
<tr>
<td></td>
<td>Birds</td>
<td>Parrots, Owls</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Trees</td>
<td>Apple trees, Pine trees</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Flowers</td>
<td>Tulips, Daisies</td>
<td></td>
</tr>
</tbody>
</table>

- One example of a causal property is, “which of these survive on an alien planet?”
- The feedback for the question “Do owls live on this planet?” could be yes
- The feedback for the question “Do animals live on this planet?” could be some
- The correct answer is “birds”
Modified 20-Questions Game

- The hierarchical structure is accessible to the subjects
- The answer to a question is a category
- And as was seen, the questions are not easy to answer because domain knowledge does not help too much
The participants were explicitly told to find the correct answer in the least number of questions.
There are 14 different hypothesis that are to be evaluated. The paper makes an assumption that the Bayesian Framework [TG01] is being used by subjects to reason. Initially all the hypotheses are equally likely.

\[
p(h|X) = \frac{p(X|h)p(h)}{\sum_{h'} p(X|h')p(h')}
\]

Simplifying assumptions:

\[
p(x|h) \in \{0, 1\}
\]

\[
p(X|h) = \prod_{i=1}^{m} p(x_i|h)
\]
How informative are the questions that were asked?

\[ I = H_{\text{prior}} - H_{\text{posterior}} \]

\( H_{\text{posterior}} \) is the predicted posterior entropy and \( H_{\text{prior}} = - \sum_h p(h) \log_2 p(h) \)

\[ H_{\text{posterior}} = p(x_1|X)H(x_1) + \cdots + p(x_n|X)H(x_n) \]

\[ H(x_i) = - \sum_h p(h|X, x_i) \log_2 p(h|X, x_i) \]

Humans are expected to ask questions which have high information gain.
Motivation

Hierarchical Task

Pre-requisites

Question Asking

Intervention

Discussion

Critique
Terminology

- $M$: Mean
- $SD$: Standard Deviation
- $SE$: Standard Error
- $p \approx$ Significance
- $F(a - 1, N - a)$: Statistical significance of multiple variables
Hypothesis Testing

It is important to check if there are any significant differences between the performance of different subjects. Especially, we want to check if “Age” is playing a role.
Analysis of Variance (ANOVA) tries to attribute the variance of an observation to some attributes. For example, say we want to explain the variability of weights in dogs.

\[
E(Y) = E(E(Y|X_i)) = E(Y|X_i)
\]

\[
V(Y) = V(E(Y|X_i))
\]

\[
V(Y|X_i)
\]

Figure 2: ANOVA: No fit
ANOVA provides a mathematical framework which supports this intuition.

Figure 1: ANOVA: Fair fit
Sometimes, we wish to find the variability of a feature while excluding individual differences. Consider the following example.

We don’t want between-subject variability to affect the within-subject variability.
Say there are multiple hypotheses to test. Say each test is not significant, so \( p < 0.05 \). If there are 20 tests, the following is the probability of observing at least one significant event.

\[
(1 - (1 - 0.05)^{20}) \approx 0.64
\]

Even though each test is not significant, the probability of at least one significant event occurring by chance is much higher than \( p \).
Bonferroni corrected multiple comparisons

It sets the significance level to $\frac{\alpha}{n}$ instead. In our previous example, the probability of occurrence of a significant event is now the following.

$\left(1 - (1 - 0.0025)^{20}\right) \approx 0.0488$

As expected, it is a conservative method.
Motivation

Hierarchical Task

Pre-requisites

Question Asking

Intervention

Discussion

Critique
Experimental Setup

- The subjects were categorized into three buckets, 7-year-olds, 10-year-olds and adults.
- All of them received different scenarios and solutions at different levels of hierarchy.
- There was no bias regarding what set of objects each group sees.
- Results were analyzed by running rANOVAs with age group (3 levels: 7-year-olds, 10-year-olds, adults) as a *between-subjects* variable.
- Trial number (3 levels: 1, 2, 3), solution condition (3 levels: subordinate-level, basic-level, superordinate-level) or scenario (3 levels: Magic box, Machine, Planet) were *within-subjects* variables.
Effect on number of questions asked

Does age have an effect?

Effect on number of questions asked

Does age have an effect?

Effect on number of questions asked

Does age have an effect?

Effect on number of questions asked

Does age have an effect?
Effect on number of questions asked

Does age have an effect?

\[ F(2, 67) = 5.29 \text{ and } p = 0.007. \]
A Bonferroni corrected multiple comparisons confirmed that 7-year olds \((M_7 = 4.92, SE = 0.34)\) asked more questions than adults \((M_{adults} = 3.36, SE = 0.35, p = 0.006)\). There was however no difference between 7 and 10 year olds and 10 year olds and adults.
Effect on number of questions asked

Does age have an effect?

$F(2, 67) = 5.29$ and $p = 0.007$. A Bonferroni corrected multiple comparisons confirmed that 7-year olds ($M_7 = 4.92, SE = 0.34$) asked more questions than adults ($M_{adults} = 3.36, SE = 0.35, p = 0.006$). There was however no difference between 7 and 10 year olds and 10 year olds and adults.

What about the level of the answer

As expected, it took lesser questions for superordinate level than the basic level, and basic level than subordinate level.
Effect on number of questions asked

Does age have an effect?

$F(2, 67) = 5.29$ and $p = 0.007$. A Bonferroni corrected multiple comparisons confirmed that 7-year olds ($M_7 = 4.92, SE = 0.34$) asked more questions than adults ($M_{adults} = 3.36, SE = 0.35, p = 0.006$). There was however no difference between 7 and 10 year olds and 10 year olds and adults.

What about the level of the answer

As expected, it took lesser questions for superordinate level than the basic level, and basic level than subordinate level.
Does age have an effect?

$F(2, 67) = 5.27$ and $p = 0.007$. A Bonferroni corrected multiple comparisons confirmed that 7-year olds asked less informative questions on average than adults. There was, however, no difference between 7 and 10-year-olds and 10-year-olds and adults.
Does age have an effect?

Adults asked more questions at the superordinate level than older children, who in turn asked more questions at that level than younger children.

Does age have an effect?

Adults asked less questions at the subordinate level than older children, who in turn asked less questions at that level than younger children. The number of basic questions did not vary significantly.
To disentangle information search efficiency and stopping rule, the number of questions asked before a single hypothesis remains is checked. **There was no effect of age on how many questions it requires.**
Analysis of shortest path

It can be concluded that developmental differences between adults and children are mainly because of an efficient stopping rule used by the former.

Presented by: Ameet Deshpande
Comparison with optimal model and random model

- Optimal model chooses the question with highest information gain at each step
- The Random model chooses a random question

The participants did better than random and worse than optimal model.
- Optimal model chooses the question with highest information gain at each step
- The Random model chooses a random question
- The participants did better than random and worse than optimal model
Presented by: Ameet Deshpande

Children search for information as efficiently as adults, but seek additional confirmatory evidence.
Intervention task

- In this task, the subjects can only choose and ask questions about objects.
- Naturally, this seems like a harder task.
- Same analysis as in Question Asking is performed.
How many objects were chosen?

Does age have an effect?
Younger children picked more objects than older children who in turn picked more objects than adults. Note the slight difference between this and Question Asking.
Information Gain

**Does age have an effect?**

A Bonferroni corrected multiple comparisons confirmed that 7-year olds asked less informative questions on average than adults. There was, however, no difference between 7 and 10-year-olds and 10-year-olds and adults. There was **no** significant difference between the average information gains when only objects before narrowing down the hypothesis are considered.

![Bar chart showing average information gain for 7-year-olds, 10-year-olds, and adults.](chart.png)

Presented by: Ameet Deshpande

Children search for information as efficiently as adults, but seek additional confirmatory evidence.
The number of objects selected before narrowing down to one hypothesis was **more** for younger children as compared to older children and adults, and there was no significant difference between older children and adults.

However, the difference was much weaker than the number of objects picked before giving a solution.

And as mentioned before, there was no significant difference when average information gain was considered.

<table>
<thead>
<tr>
<th>Solution</th>
<th>Young Children</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full path</td>
<td>7.79</td>
<td>4.11</td>
</tr>
<tr>
<td>Shortest path</td>
<td>5.74</td>
<td>4.02</td>
</tr>
</tbody>
</table>
The number of objects selected before narrowing down to one hypothesis was more for younger children as compared to older children and adults, and there was no significant difference between older children and adults.

However, the difference was much weaker than the number of objects picked before giving a solution.

And as mentioned before, there was no significant difference when average information gain was considered (Does this make sense?)

<table>
<thead>
<tr>
<th>Solution</th>
<th>Young Children</th>
<th>Adults</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full path</td>
<td>7.79</td>
<td>4.11</td>
</tr>
<tr>
<td>Shortest path</td>
<td>5.74</td>
<td>4.02</td>
</tr>
</tbody>
</table>
Children search for information as efficiently as adults, but seek additional confirmatory evidence.
Conclusion

- Information search by children is better than random
- All age-groups leveraged hierarchies in the tasks. Further evidence for this is the ease with which higher level answers are retrieved. *Intervention* does not have these advantages though.
- Children do not have an efficient stopping rule, but otherwise, search like adults
  - Children may seek additional confirmatory evidence before telling the answer
  - It is possible that children considered disjunctive hypotheses
1 Motivation

2 Hierarchical Task

3 Pre-requisites

4 Question Asking

5 Intervention

6 Discussion

7 Critique
### Can we find a naive yet useful strategy?

<table>
<thead>
<tr>
<th>Scenario</th>
<th>Superordinate category</th>
<th>Basic-level category</th>
<th>Subordinate category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Animals</td>
<td>Fish</td>
<td>Goldfish, Clownfish</td>
<td></td>
</tr>
<tr>
<td>Planet Aparis</td>
<td>Birds</td>
<td>Owls, Parrots</td>
<td></td>
</tr>
<tr>
<td>Plants</td>
<td>Trees</td>
<td>Apple trees, Pine trees</td>
<td></td>
</tr>
<tr>
<td>Clothes</td>
<td>Flowers</td>
<td>Tulips, Daisies</td>
<td></td>
</tr>
<tr>
<td>Machine</td>
<td>Shirts</td>
<td>Long sleeves, Short sleeves</td>
<td></td>
</tr>
<tr>
<td>Furniture</td>
<td>Shoes</td>
<td>Flip-flops, Boots</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Tables</td>
<td>Dining tables, Desk</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Chairs</td>
<td>Rocking chair, High chair</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Cars</td>
<td>Vans</td>
<td></td>
</tr>
<tr>
<td>Vehicles</td>
<td>Planes</td>
<td>Sportcars, Planes</td>
<td></td>
</tr>
<tr>
<td>Magic-box</td>
<td>Apples</td>
<td>Helicopters</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Berries</td>
<td>Yellow apples</td>
<td></td>
</tr>
</tbody>
</table>

Children search for information as efficiently as adults, but seek additional confirmatory evidence.
Why were the participants told not to consider disjunctive hypotheses?

 Were the participants the same for both the tests?

 The hierarchical 20-questions task could have been explained better


