A logical model for Deductive Mastermind

Nina Gierasimczuk

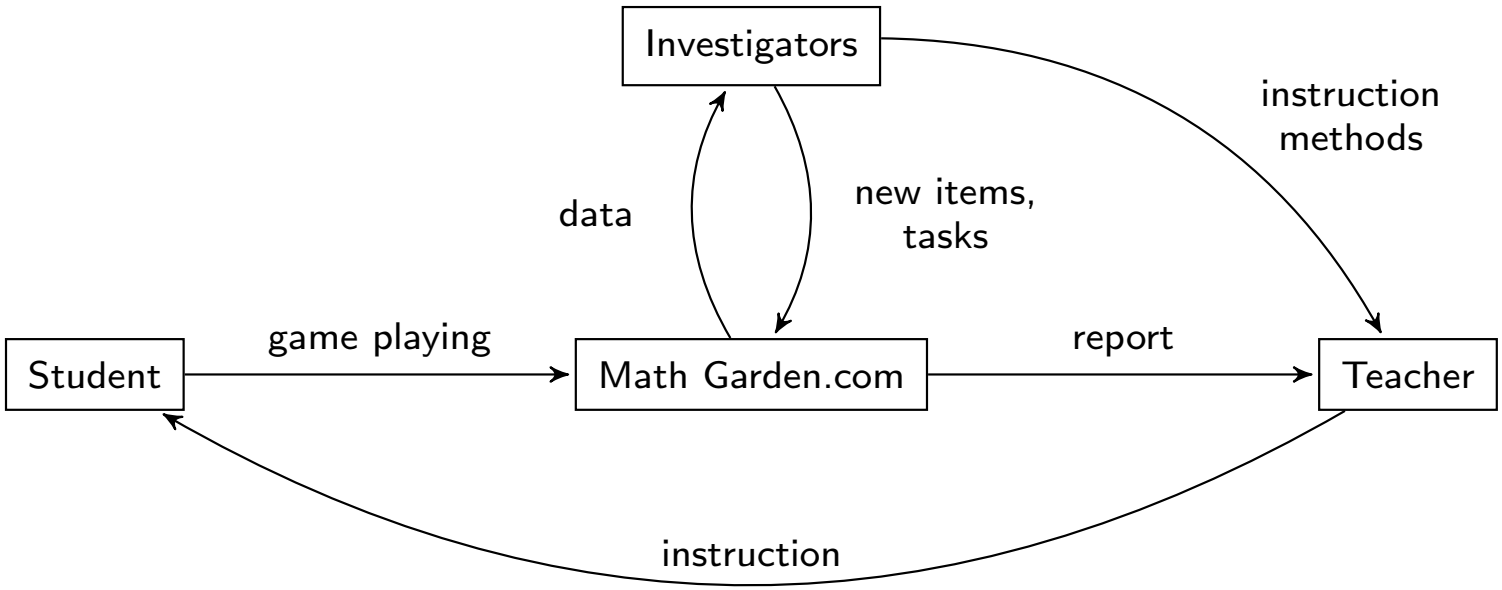
based on a joint work with Han van der Maas and Maartje Raijmakers

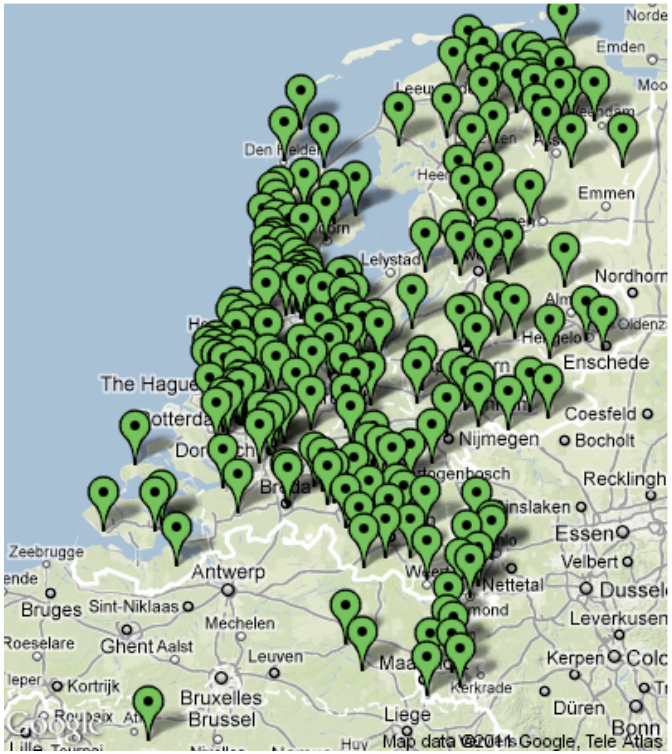
Institute for Logic, Language and Computation
University of Amsterdam



False-Belief task Mini-workshop
Amsterdam
Dec 5, 2013

- adaptive training environment with educational games
- abstract thinking development
- 15 arithmetic games and 2 complex reasoning games





- In three years the number of schools has grown from 8 to over 700.
- Currently, over 90,000 students have generated over 200 billion.

Difficulty Levels

- students play game-items suited for their level (75% correctly)
- the tasks' difficulty and the students' level are continuously estimated
- via the Elo rating system (used for ranking chess players, Elo 1978)
- i.e.: students are ranked by playing, and items are rated by getting played
- ratings depend on accuracy and speed of item solving (Klinkenberg 2011)

By-products:

1) rating of all items (item difficulty parameters)
2) rating of children (reflecting the reasoning ability)

introduce a dedicated logical reasoning training in primary schools

understand the empirically established item difficulty parameters

by means of a logical analysis of the items

computational paradigms for cognition

Mastermind Game

- Meirowitz 1970, but similar to the earlier Bulls and Cows
- an inductive inquiry game, trials of experimentation and evaluation

Mastermind Game: A Code-Breaking Game

- the set consists of:
  - a decoding board
  - code pegs of k colors
  - and feedback pegs of black and white
- players:
  - the code-maker: chooses a secret pattern of ℓ code pegs
  - the code-breaker: guesses the pattern, in a given n rounds
- rounds:
  - code-breaker makes a guess by placing a row of ℓ code pegs
  - code-maker provides the feedback:
    one black for each code peg of correct color and position, and
    one white for each peg of correct color but wrong position
  - repeat until either the code-breaker guesses correctly, or n incorrect guesses have been made
- winning:
  - for the code-breaker: if obtains the solution within n rounds
  - the code-maker wins otherwise

- question of the underlying logical reasoning and its difficulty
- mathematical results only on existence of efficient strategies (Knuth 1976, Irving 1978, Koyama 1993, Kooi 2005)

Static Mastermind (Chvatal 1983)

Static Mastermind is a version of the Mastermind game

- the goal is to find out the minimum number of guesses
- the code-breaker can make all at once
- at the beginning of the game
- without waiting for the individual feedbacks
- and upon receiving them all at once
- completely determine the code in the next guess

.

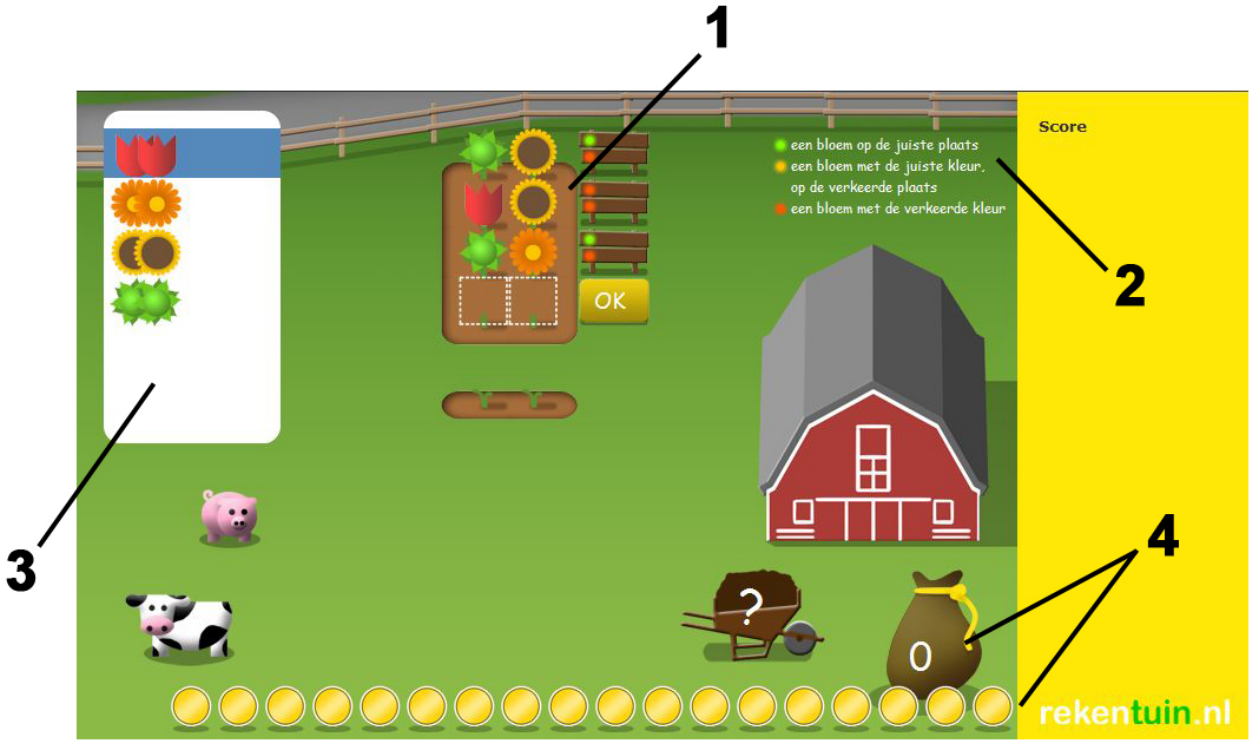
Definition (Mastermind Satisfiability Decision Problem)

Input A set of guesses G and their corresponding scores.

Question Is there at least one valid solution?

Theorem

Mastermind Satisfiability Decision Problem is NP-complete wrt ℓ (positions).



1) decoding board
2) short feedback instruction
3) domain of flowers to choose from
4) timer in the form of disappearing coins

Some Facts about Deductive Mastermind

- unlike the classic, DM does not have the trial conjectures
- DM gives the clues upfront
- unlike SM, ensuring exactly one correct solution

reduced complexity: from inductive inference to logical-reasoning game

DM is a combo of the classic (the goal is the same) and SM (no trial-and-error)

emphasis on the atomic logical steps of non-linguistic logical reasoning

DM is easily adaptable as a single-player game (Math Garden)

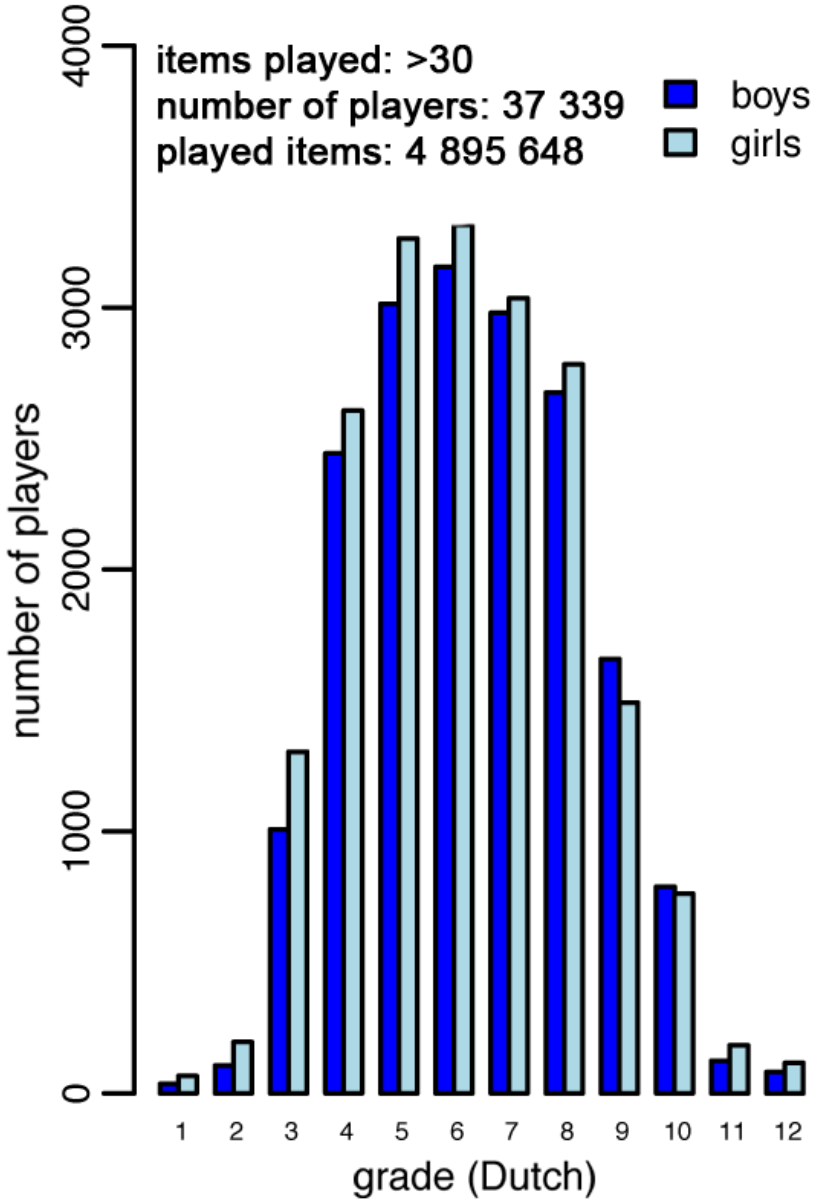
Some Facts about Flowercode

- running since November 2010
- 321 game-items, 1-5 flowers, 2-5 colors
- by December 2012, 4,895,648 items had been played
- by 37,339 primary school students (grades 1-6, age: 6-12 years)
- in over 700 locations (schools and family homes)

We can access:

- the individual progress of individual players on a single game
- the most frequent mistakes with respect to a game-item
- the relative difficulty of game-items
- correlations with different, mathematical games
- etc.

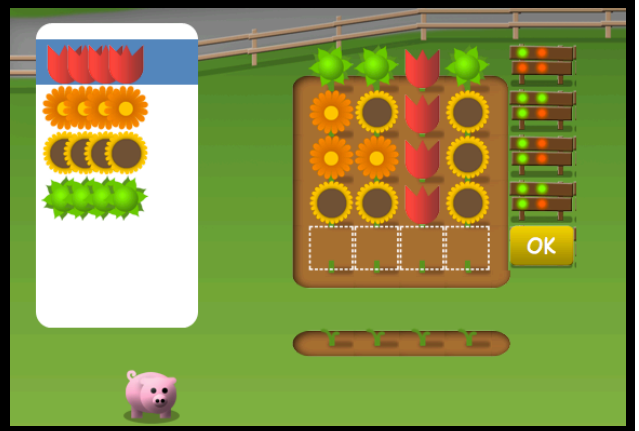
Some Facts about Flowercode

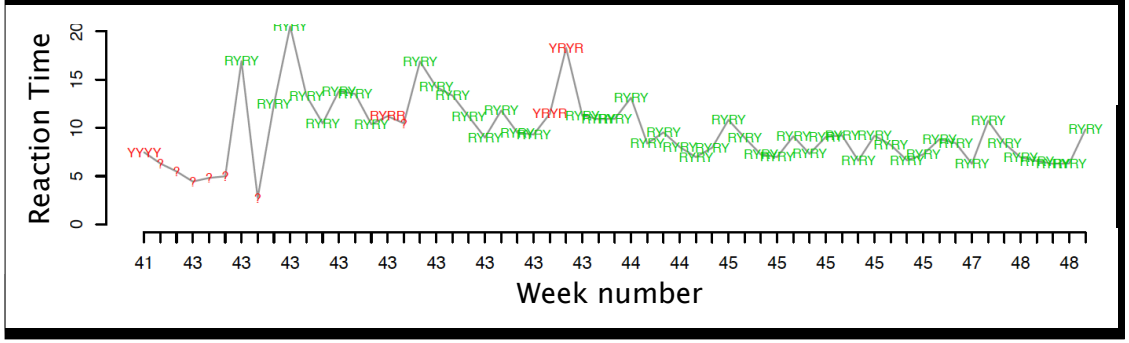


Some Facts about Flowercode

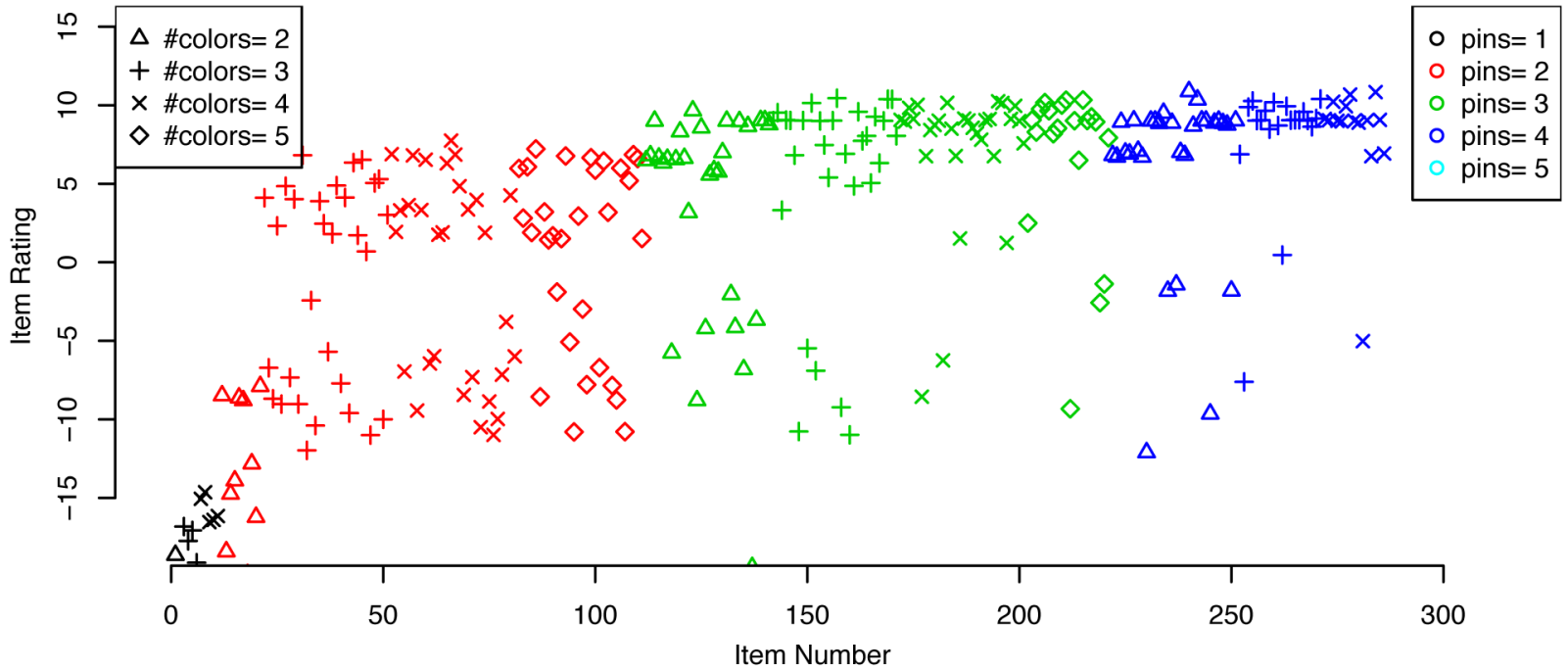
user 163545
played 3601 items

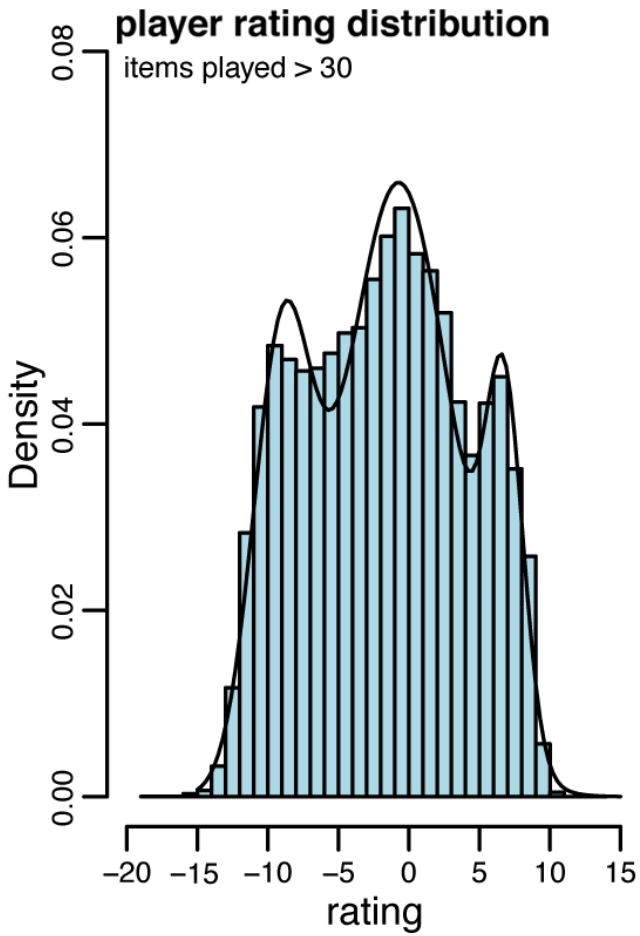
59 times this item

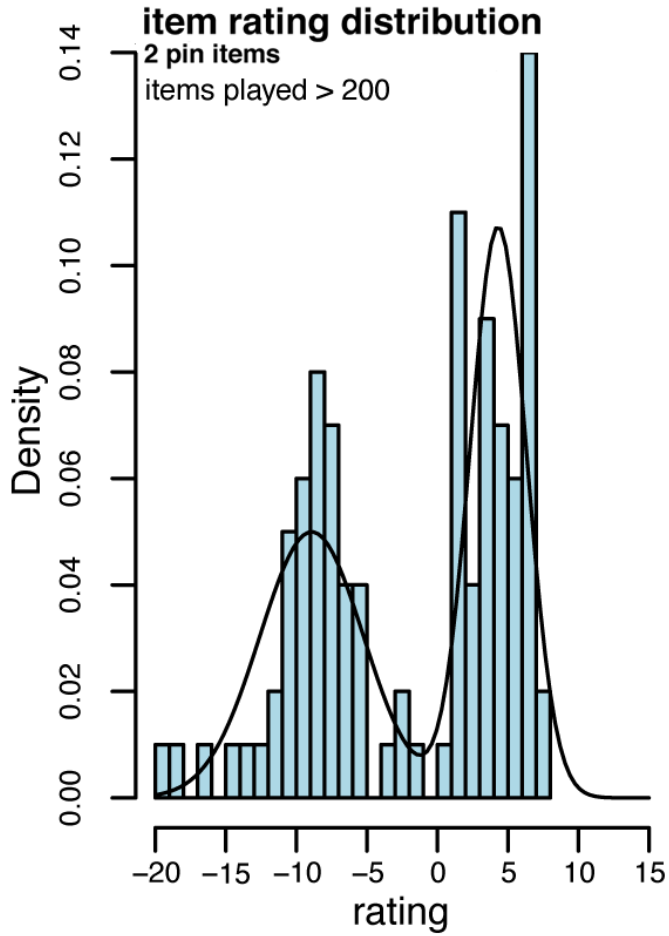




initial difficulty estimation in terms of non-logical aspects
(number of flowers, colors, lines, the rate of the hypotheses elimination, etc.)







how to fix this to facilitate the training effect?

Each Deductive Mastermind game-item consists of a sequence of conjectures.

Definition

A conjecture of length l over k colors is any sequence given by a total assignment, $h : \{1, \ldots, \ell\} \to \{c_1, \ldots, c_k\}$. The *goal sequence* is a distinguished conjecture, $goal : \{1, \ldots, \ell\} \to \{c_1, \ldots, c_k\}$.

- every non-goal conjecture is accompanied by a feedback
- that indicates how similar h is to the given goal assignment
- feedback colors g, o, r

- set $\mathbb{G} := \{G | G \subseteq \{1, \ldots, \ell\} \wedge card(G) = a\}$, and,
- if $G \subseteq \{1, \ldots, \ell\}$, then $\mathbb{O}^G = \{O | O \subseteq \{1, \ldots, \ell\} \setminus G \wedge card(O) = b\}$

Definition

Finally, we can set $Bt(h, f)$, the Boolean translation of (h, f) to be given by:

$$Bt(h, f) := \bigvee_{G \in \mathbb{G}} (\varphi^g_G \wedge \bigvee_{O \in \mathbb{O}^G} (\varphi^o_{G,O} \wedge \varphi^r_{G,O})).$$

Example

Let us take $\ell = 2$ and (h, f) such that: $h(1){:=}c_1$, $h(2){:=}c_2$; $f{:=}or$. Then $\mathbb{G}{=}\{\emptyset\}$, $\mathbb{O}^{\{\emptyset\}}{=}\{\{1\},\{2\}\}$. The corresponding formula, $Bt(h, f)$, is:

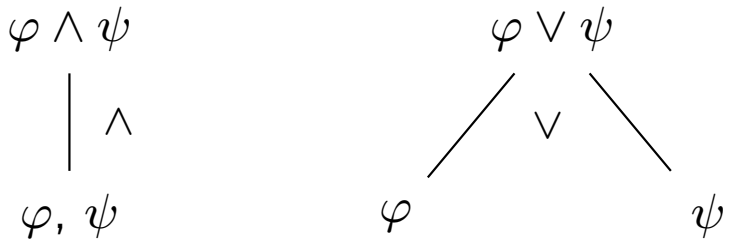
$(goal(1){\neq}c_1 \wedge goal(2){\neq}c_2) \wedge ((goal(1){=}c_2 \wedge goal(2){\neq}c_1) \vee (goal(2){=}c_1 \wedge goal(1){\neq}c_2))$

Definition

A Deductive Mastermind game-item over ℓ positions, k colors and n lines, $DM(l, k, n)$, is a set $\{(h_1, f_1), \ldots, (h_n, f_n)\}$ of pairs, each consisting of a single conjecture together with its corresponding feedback. Respectively, $Bt(DM(l, k, n)) = Bt(\{(h_1, f_1), \ldots, (h_n, f_n)\}) = \{Bt(h_1, f_1), \ldots, Bt(h_n, f_n)\}$.

- hence, each DM game-item is a set of Boolean formulae
- moreover, by the construction this set is satisfiable
- and, even more, there is a unique valuation

- analytic tableau is a decision procedure for propositional logic
- it solves satisfiability of finite sets of formulas of propositional logic
- by giving an adequate valuation
- building a formula-labeled tree rooted at the set
- unfolding breaks them into smaller formulae
- until contradiction is found or no further reduction is possible

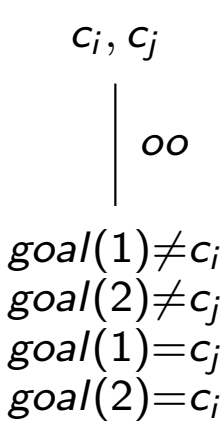


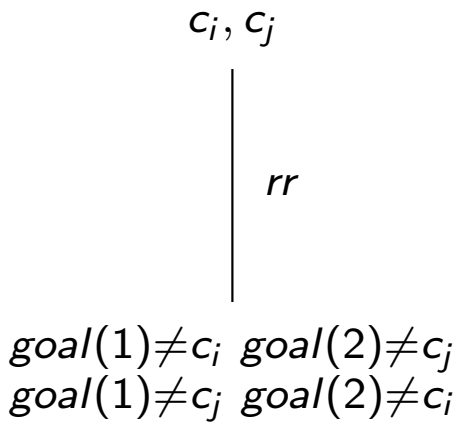
By construction of DM

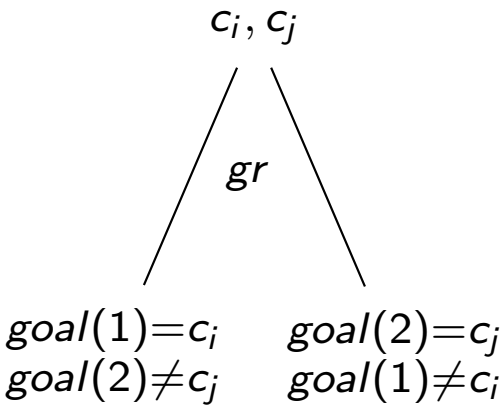
Applying the analytic tableaux method to the Boolean translation of a Deductive Mastermind game-item will give the unique missing assignment *goal*.

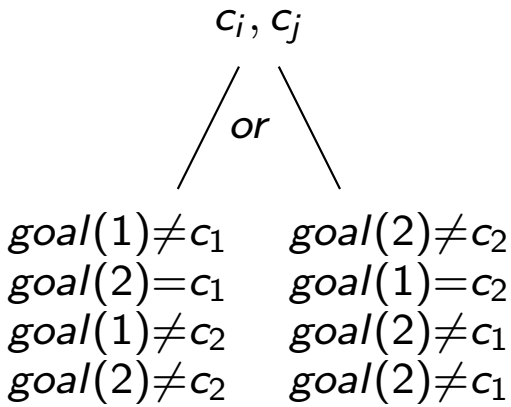
2-placed Deductive Mastermind Game-items

gg, go, oo, rr, gr, or



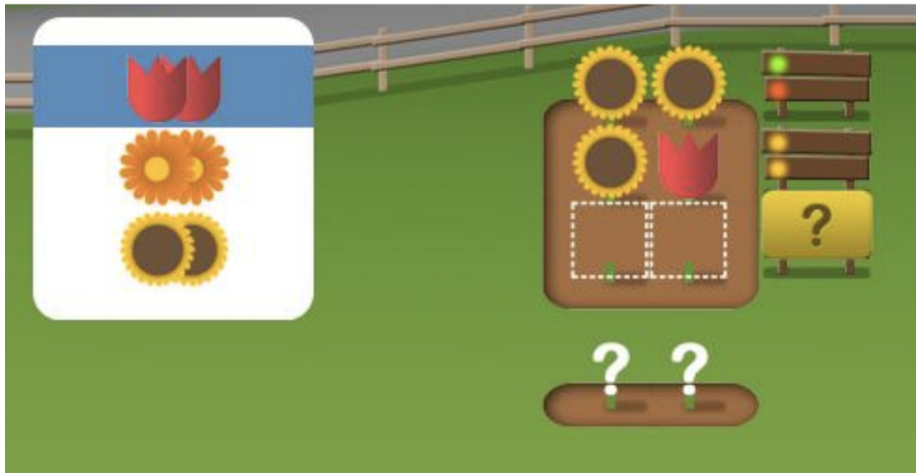


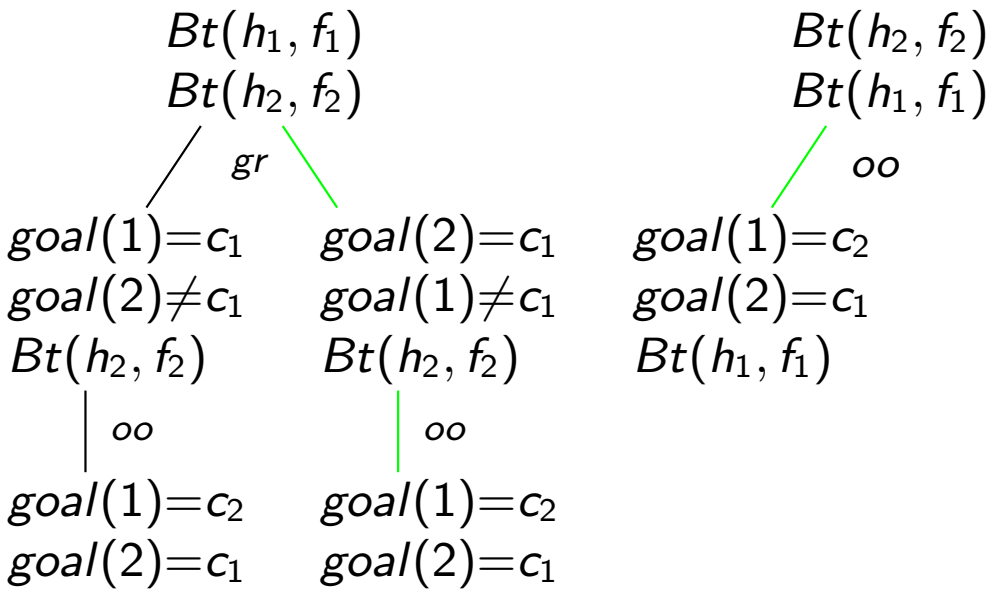




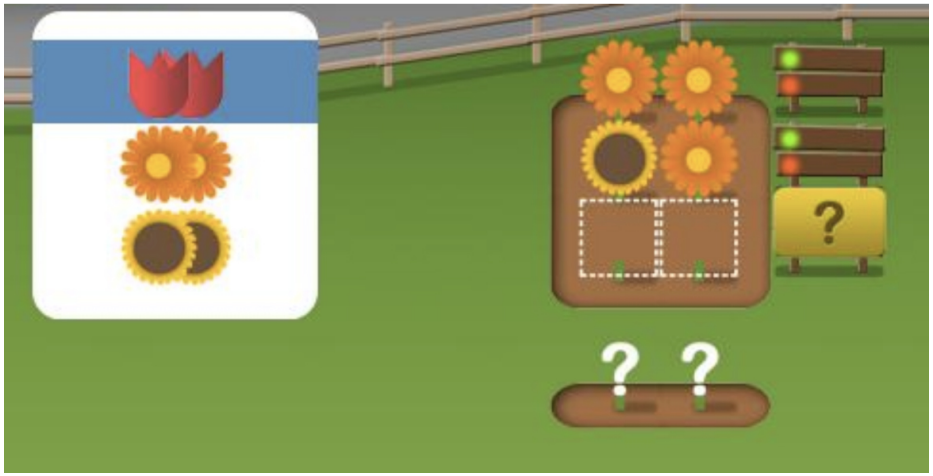
$oo < rr < gr < or$

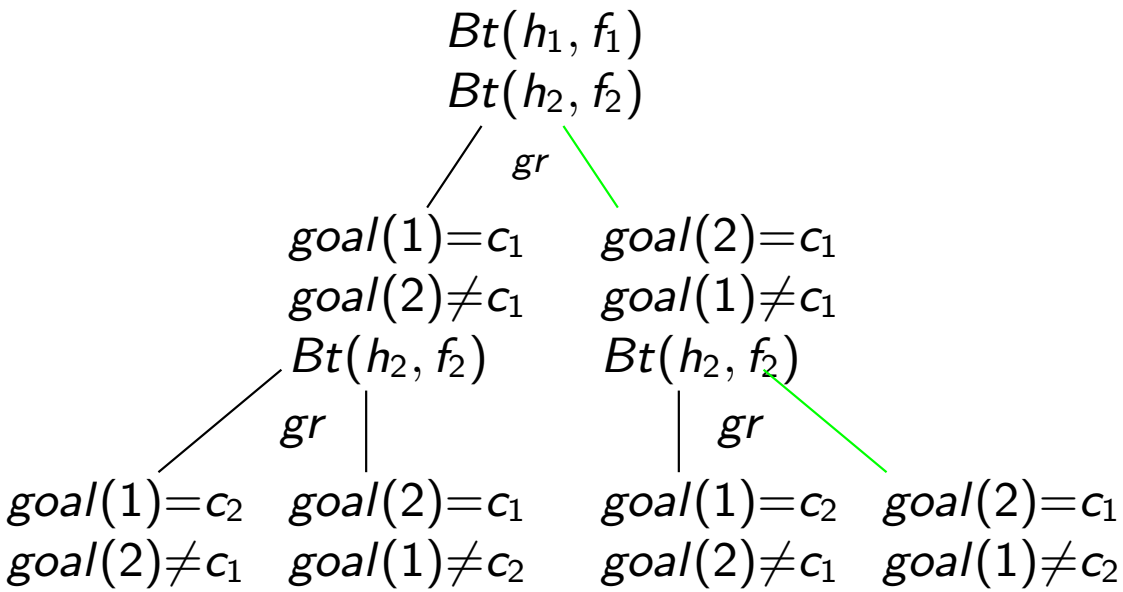
An Example


?
? ?


$Bt(h_1, f_1)$
$Bt(h_2, f_2)$
gr
$goal(1)=c_1$
$goal(2)\neq c_1$
$Bt(h_2, f_2)$
oo
$goal(1)=c_2$
$goal(2)=c_1$
$goal(2)=c_1$
$goal(1)\neq c_1$
$Bt(h_2, f_2)$
oo
$goal(1)=c_2$
$goal(2)=c_1$
$Bt(h_2, f_2)$
$Bt(h_1, f_1)$
oo
$goal(1)=c_2$
$goal(2)=c_1$
$Bt(h_1, f_1)$

Another Example





- tableau give 'ideal' reasoning scheme
- abstract complexity measure (tree size)
- shape and size of the tree depends on what goes first (minimal size)
- reasoning optimization

items' initial difficulty corresponds to the size of top-botom trees

items' logical difficulty corresponds to the size of the minimal trees

the reasoning is optimized according to feedback complexity

- participants: 28,247 students from grades 1-6, of age: 6-12 years.
- played: 2,187,354 items between November 2010 and January 2012
- items: 321 DM items among them 100 two-places items.

all factors but one (*gr*) were significant in predicting item difficulties

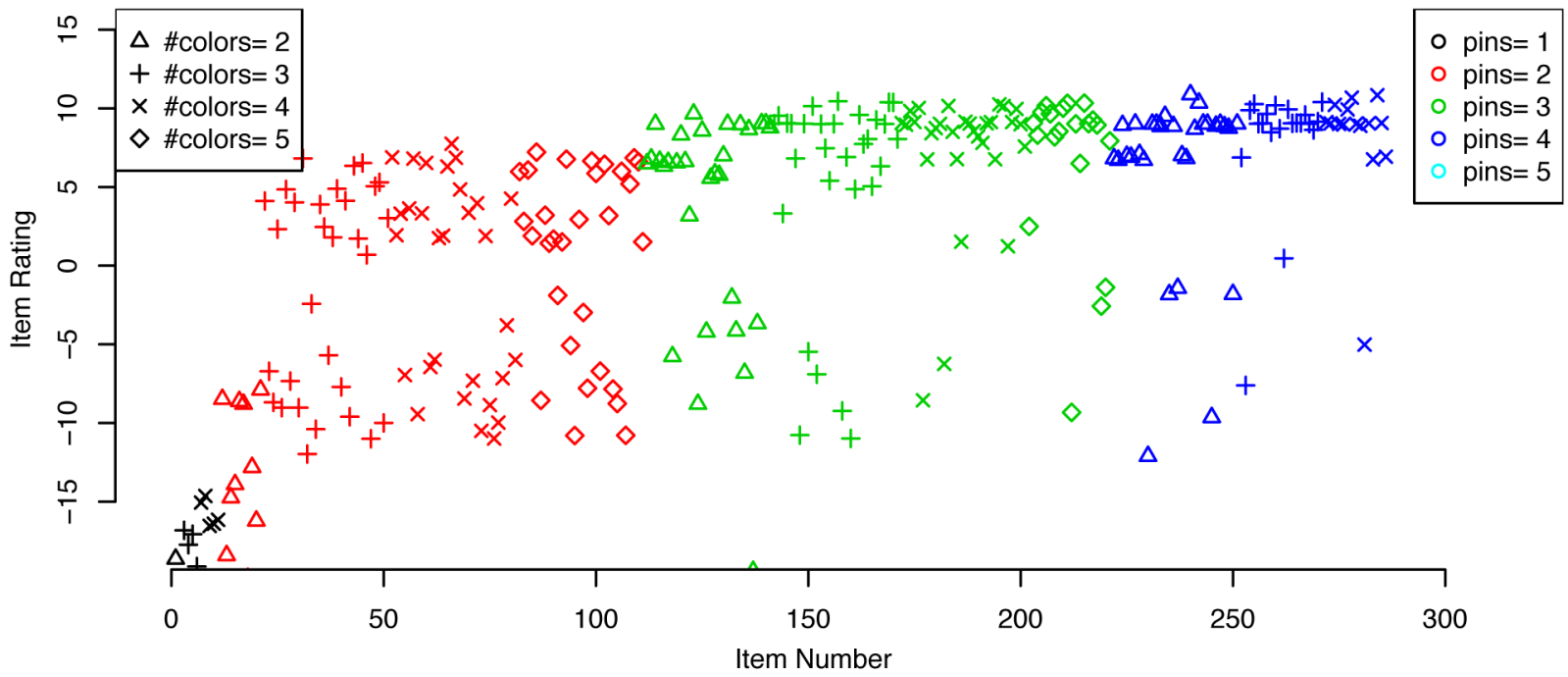
two difficulty clusters

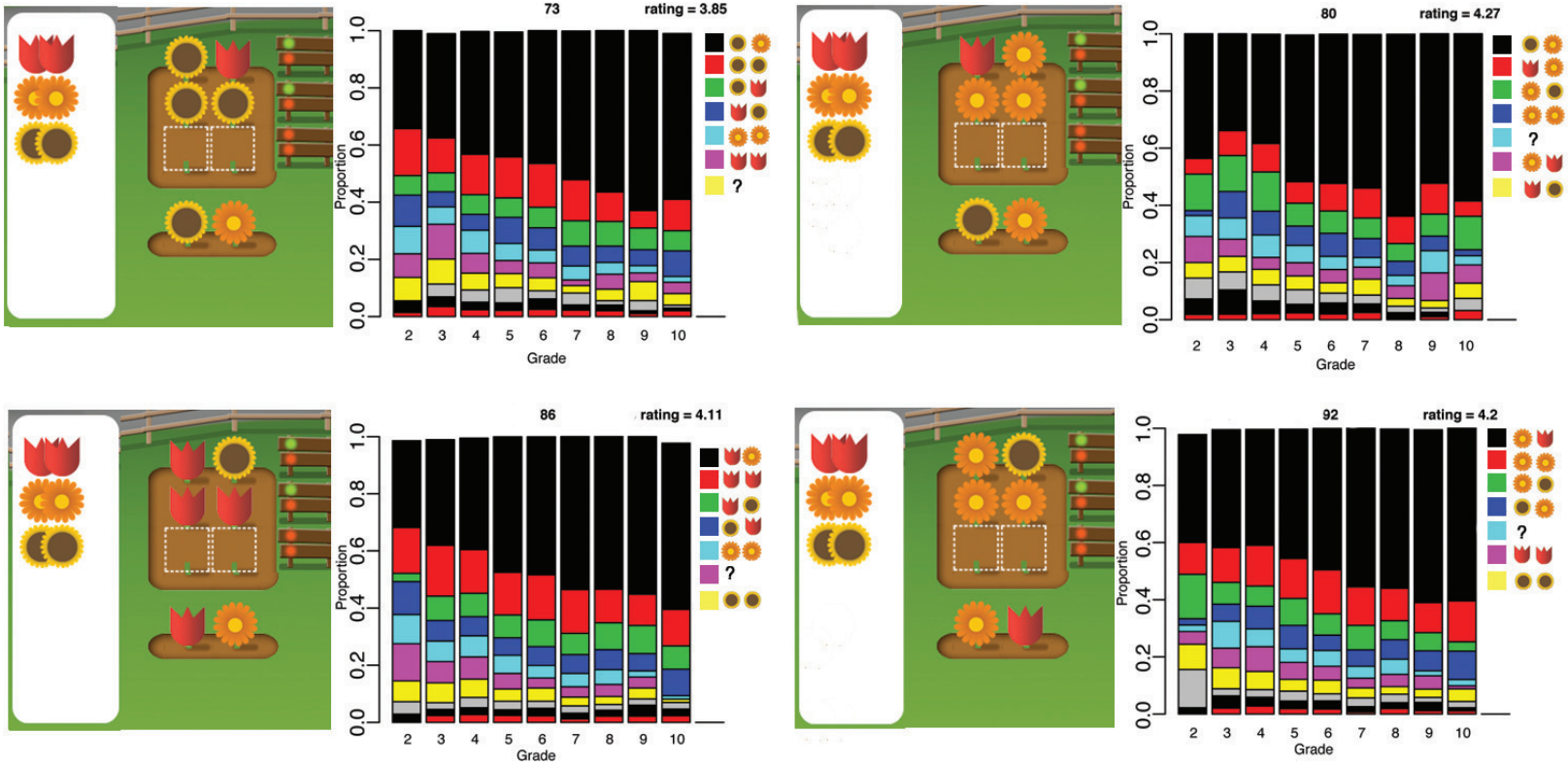
1) easy

   no *or* feedback and no *gr* feedback

   no *or* feedback, at least one *gr* feedback, and all colors are included

2) difficult: otherwise




73
rating = 3.85
80
rating = 4.27
86
rating = 4.11
92
rating = 4.2
Proportion
Grade
0.0
0.2
0.4
0.6
0.8
1.0
2 3 4 5 6 7 8 9 10
?

- operationalization of false-belief reasoning as online games
- testing ACT-R models predictions
- based on logical reasoning
- give a preliminary classification of lines along which logic can help

THANK YOU!


Nina Gierasimczuk, Han van der Maas, and Maartje Raijmakers, An analytic tableaux model for Deductive Mastermind empirically tested with a massively used online learning system, Journal of Logic, Language and Information 2013.


Nina Gierasimczuk, Han van der Maas, and Maartje Raijmakers, Logical and Psychological Analysis of Deductive Mastermind, Proceedings of the Logic & Cognition Workshop at ESSLLI 2012, Opole, Poland, 13-17 August, 2012, CEUR Workshop Proceedings.