ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

CS-411 : Digital Education & Learning Analytics

Chapter 10:

Learning Modeling

Pierre Dillenbourg and Patrick Jerman

Learner Modelling

$$5^2 = ?$$

Learner Modelling

From the learner’s behaviour, infer his/her knowledge state

$5^2 = 25$ ➔ knows X^2

$5^2 \neq 25$ ➔ doesn’t know X^2

Cognitive Diagnosis

$5^2 = ??$	Knowledge States										
Behavior (Answer)	$5^2 = 25$	$5^n = ...$	$n^2 = n . N$	$x^n = x.x....$	$x^n = x . x$ but bad mult.	$x^n = x.n$	$x^n = x + n$	$x^n = ???$	Sum	Entropy	Normalized entropy
25	**0.10**	**0.20**	**0.30**	**0.40**	0.00	0.00	0.00	0.00	1	1.89	0.63
35	0.00	0.00	0.00	0.00	**0.40**	0.10	0.00	**0.50**	1	1.41	0.47
10	0.00	0.00	0.00	0.00	0.00	**0.79**	0.00	0.20	1	0.79	0.26
27	0.00	0.00	0.00	0.00	0.40	0.00	0.00	**0.59**	1	1.03	0.34
7	0.00	0.00	0.00	0.00	0.00	0.00	**0.59**	0.40	1	1.03	0.34
											0.41

Diagnosis Power

Learner Modelling

$5^2 =$

- [ ] 52
- [x] 25
- [ ] 15
- [ ] 10
- [ ] 7

p (knows X^2)= 1 ?

Learner Modelling

From the learner's behaviour, infer his/her knowledge state

p (state = knows | correct-answer) = 1 – Guess

p (state = knows | incorrect-answer) = 0 + Slip

Factors that depend upon the response modality

Bayesian Knowledge Tracing, Corbezt & Anderson

Learner Modelling

$5^2 = ??$	Knowledge States										
Behavior (Answer)	$5^2 = 25$	$5^n = ...$	$n^2 = n . N$	$x^n = x.x....$	$x^n = x . x$ but bad mult.	$x^n = x.n$	$x^n = x + n$	$x^n = ???$	Sum	Entropy	Normalized entropy
25	**0.10**	**0.20**	**0.30**	**0.40**	0.00	0.00	0.00	0.00	1	1.89	0.63
35	0.00	0.00	0.00	0.00	**0.40**	0.10	0.00	**0.50**	1	1.41	0.47
10	0.00	0.00	0.00	0.00	0.00	**0.79**	0.00	0.20	1	0.79	0.26
27	0.00	0.00	0.00	0.00	0.40	0.00	0.00	**0.59**	1	1.03	0.34
7	0.00	0.00	0.00	0.00	0.00	0.00	**0.59**	0.40	1	1.03	0.34
											0.41

$5^2 = ??$	Knowledge States (with SLIP/GUESS factors)										
Behavior (Answer)	$5^2 = 25$	$5^n = ...$	$n^2 = n.n$	$x^n = x.x....$	$x^n = x . x$ but bad mult.	$x^n = x.n$	$x^n = x + n$	$x^n = ???$	Sum	Entropy	Normalized entropy
25	**0.05**	**0.15**	**0.25**	**0.35**	0.05	0.05	0.05	0.05	1	2.53	0.84
35	0.05	0.05	0.05	0.05	**0.34**	0.05	0.00	0.41	1	2.16	0.72
10	0.05	0.05	0.05	0.05	0.00	**0.64**	0.00	0.15	1	1.72	0.57
27	0.05	0.05	0.05	0.05	0.25	0.00	0.00	**0.54**	1	1.87	0.62
7	0.05	0.05	0.05	0.05	0.00	0.00	**0.50**	0.30	1	1.92	0.64
											0.68

From the learner's behaviours, infer his/her knowledge state

$X(S)$={lost, active, fine, brilliant}

b(s) = watch video with many pauses

x(s)=[.15 .40 .30 .15]

H_0= 0.94

b(s) = select correct definition of SD in a quiz with 5 possible definitions

x(s)= [.05 .15 .25. 55]

H_0 = .80

b(s) = post a message "There is a mistake on the slide" (and there is one indeed)

x(s)= [.01 .02 .02 .95]

H_0 = .18

Normalized entropy of the diagnosis vector

Diagnosis Power

State "fine": the learner is performing well
State "active": the learner is working but does not seem to succeed well
State "lost": the learner does not understand at all or did not complete the activities
State "drop": the learned has dropped out (e.g. no login since N days)

x(s)=[.15 .40 .30 .15]

$$H(X) = -\sum_i P(x_i) \log_b P(x_i)$$

$\boxed{H_0 = 0.94}$

The uncertainty of the diagnosis can be estimated by Shannon's entropy applied to the vector fo probabilities for the different states.

Since this value depends upon the number of states, we normalize it on a 0->1 scale by dividing it by the maximal entropy which $\log_2$ of the number of states

The diagnosis power of a question can be measured by the entropy of the diagnosis vector

Which question has the highest diagnosis power ?

Question 1

The standard deviation of a distribution if the ....of the sum of …............. from the mean

Question 2

Remove two numbers from this distribution to minimize it’s standard deviation : [1 3 3 5 9 9 9 10 11 18 19 25 29]

Basic approach to reduce uncertainty

Decrease uncertainty by collecting multiple answers

$5^2 = ??$	Knowledge States (with SLIP/GUESS factors)										
Behavior (Answer)	$5^2 = 25$	$5^n = ...$	$n^2 = n.n$	$x^n = x.x....$	$x^n = x . x$ but bad mult.	$x^n = x.n$	$x^n = x + n$	$x^n = ???$	Sum	Entropy	Normalized entropy
25	**0.05**	**0.15**	**0.25**	**0.35**	0.05	0.05	0.05	0.05	1	2.53	0.84
35	0.05	0.05	0.05	0.05	**0.34**	0.05	0.00	0.41	1	2.16	0.72
10	0.05	0.05	0.05	0.05	0.00	**0.64**	0.00	0.15	1	1.72	0.57
27	0.05	0.05	0.05	0.05	0.25	0.00	0.00	**0.54**	1	1.87	0.62
7	0.05	0.05	0.05	0.05	0.00	0.00	**0.50**	0.30	1	1.92	0.64
											0.68

$7^2 = ??$	Knowledge States (second question)										
Behavior (Answer)	$7^2 = 25$	$7^n = ...$	$n^2 = n.n$	$x^n = x.x....$	$x^n = x . x$ but bad mult.	$x^n = x.n$	$x^n = x + n$	$x^n = ???$	Sum	Entropy	Normalized entropy
49	**0.02**	**0.10**	**0.35**	**0.45**	0.02	0.02	0.02	0.02	1.00	1.95	0.65
56	0.02	0.02	0.02	0.02	**0.61**	0.00	0.00	0.31	1.00	1.43	0.48
14	0.02	0.02	0.02	0.02	0.00	**0.82**	0.00	0.10	1.00	1.04	0.35
72	0.02	0.02	0.02	0.02	0.32	0.00	0.00	**0.60**	1.00	1.44	0.48
9	0.02	0.02	0.02	0.02	0.00	0.00	**0.72**	0.20	1.00	1.27	0.42
											0.48

$3^3 = ??$	Knowledge States (second question)										
Behavior (Answer)	$3^3 = 25$	$3^n = ...$	$n^3 = n.n.n$	$x^n = x.x....$	$x^n = x . x$ but bad mult.	$x^n = x.n$	$x^n = x + n$	$x^n = ???$	Sum	Entropy	Normalized entropy
49	**0.01**	**0.10**	**0.20**	**0.65**	0.01	0.01	0.01	0.01	1.00	1.53	0.51
26	0.01	0.01	0.01	0.01	**0.81**	0.00	0.00	0.15	1.00	0.94	0.31
9	0.01	0.01	0.01	0.01	0.00	**0.82**	0.00	0.14	1.00	0.92	0.31
33	0.01	0.01	0.01	0.01	0.12	0.00	0.00	**0.84**	1.00	0.86	0.29
6	0.01	0.01	0.01	0.01	0.00	0.00	**0.91**	0.05	1.00	0.63	0.21
											0.33

How does the teacher/system chooses the next question ?

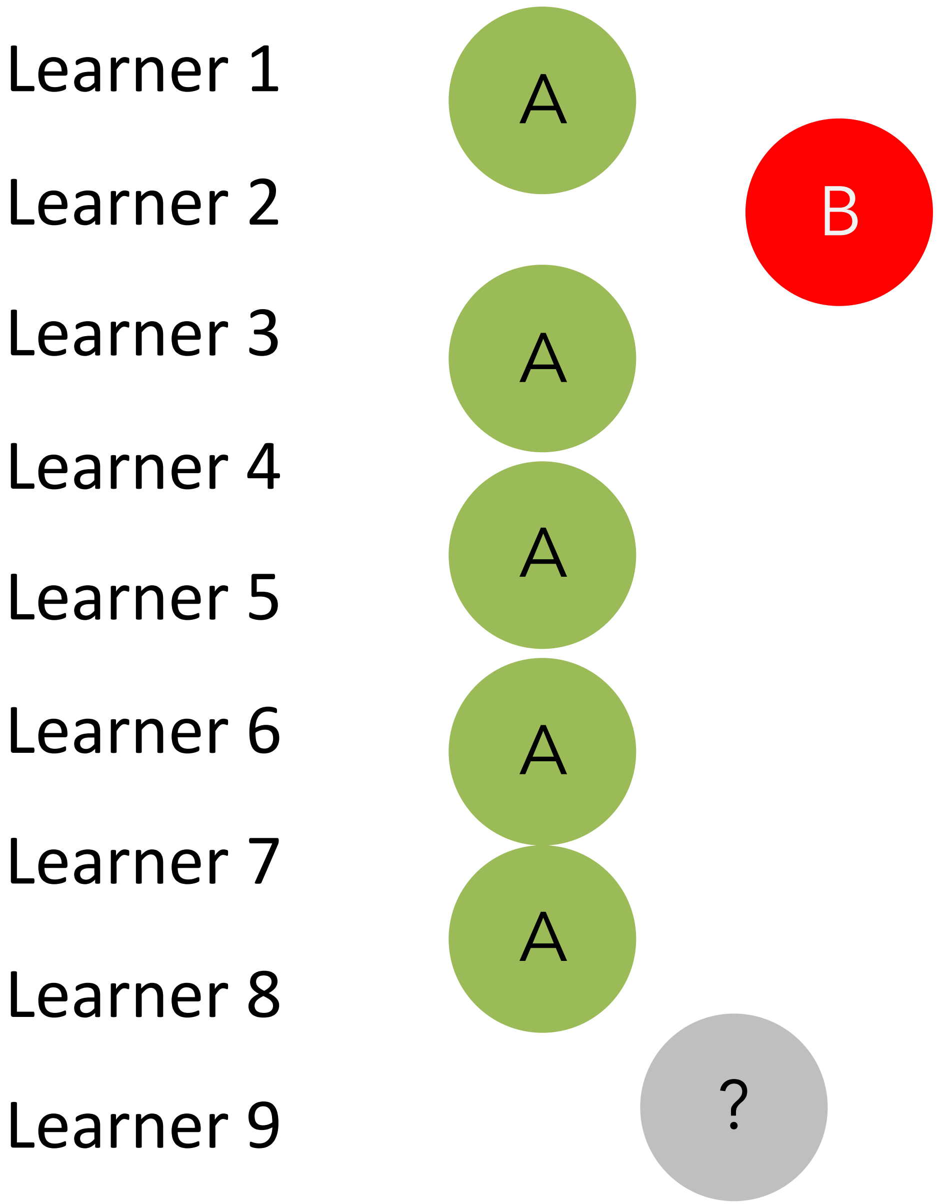
- Because it will maximize the learning gain of the learner ?

Exploitation

- Because it will maximize the system knowledge about the learner ?

Exploration

Exploration Exploitation Tradeoff



Learner Modelling

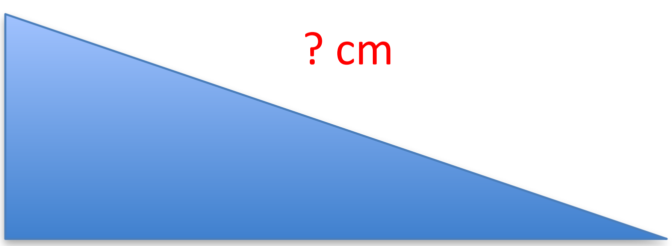
From the learner's behaviour, **infer** his/her knowledge state

Inference mechanisms

△ *(learner-knowledge, correct-knowledge) = ƒ (△ (learner-answer, correct-answer))*

? cm

L1: 3 cm



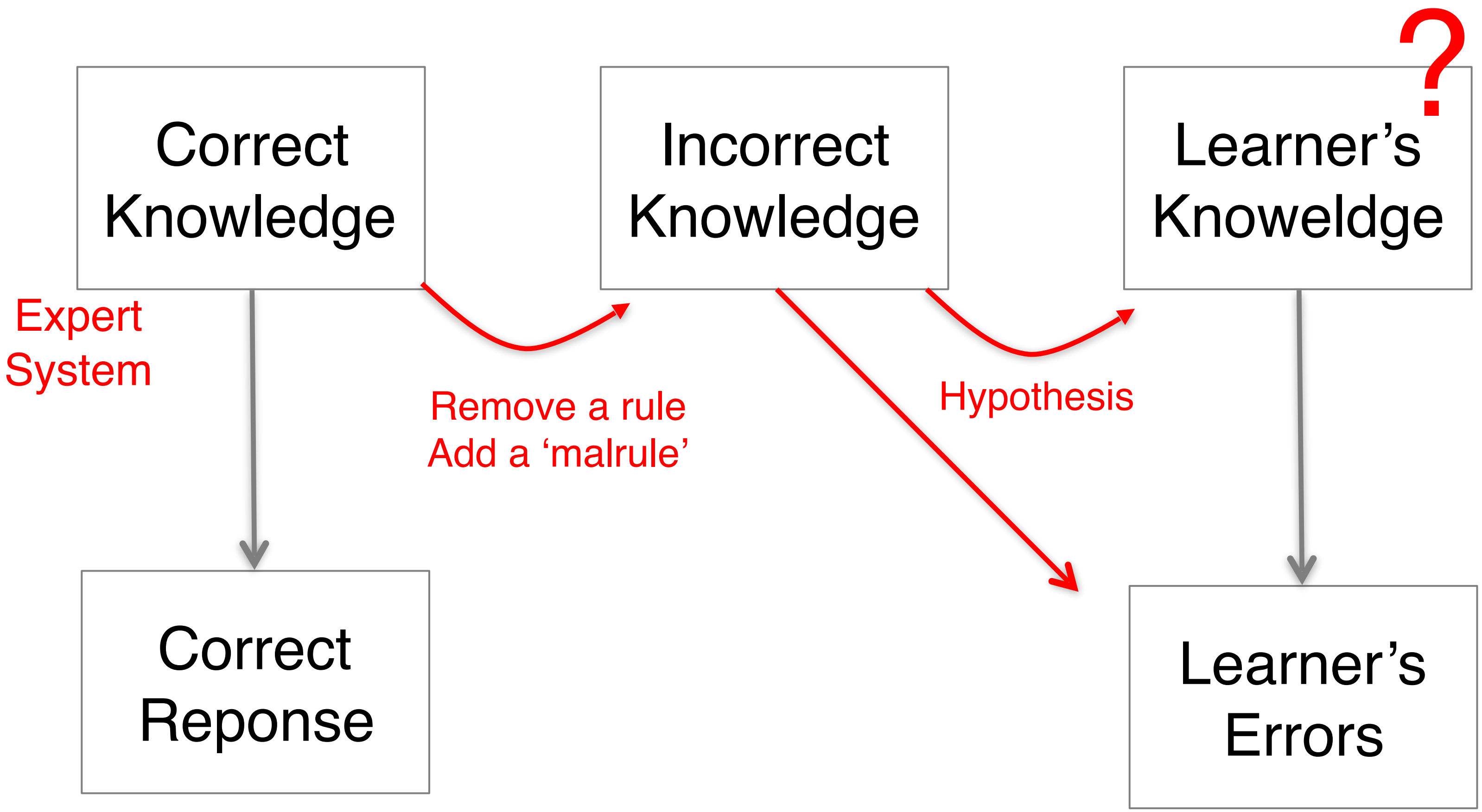
L2 : 8 cm

Correct answer = 8.54

Learner Answer = 8.18 = SQRT ($8^2 + 3^1$)

AI approach to inference mechanisms:

If, when bringing perturbation X to an expert system, it produces the same mistake as the learner, X is a good hypothesis of what the learner did not understand



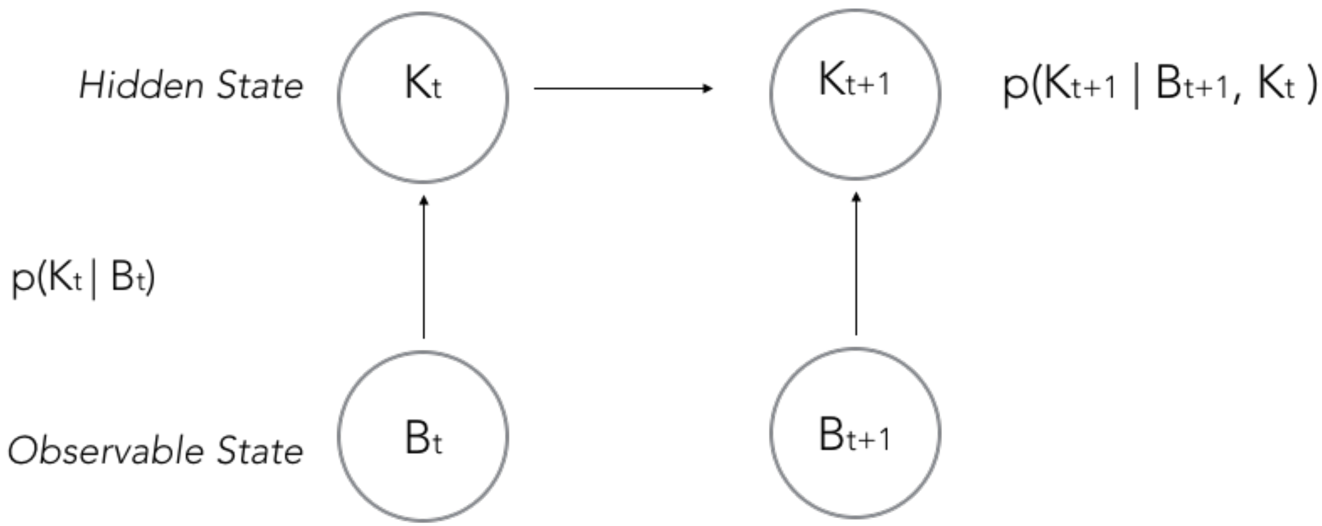
Learner Modelling

From the learner's behaviour, **infer** his/her knowledge state

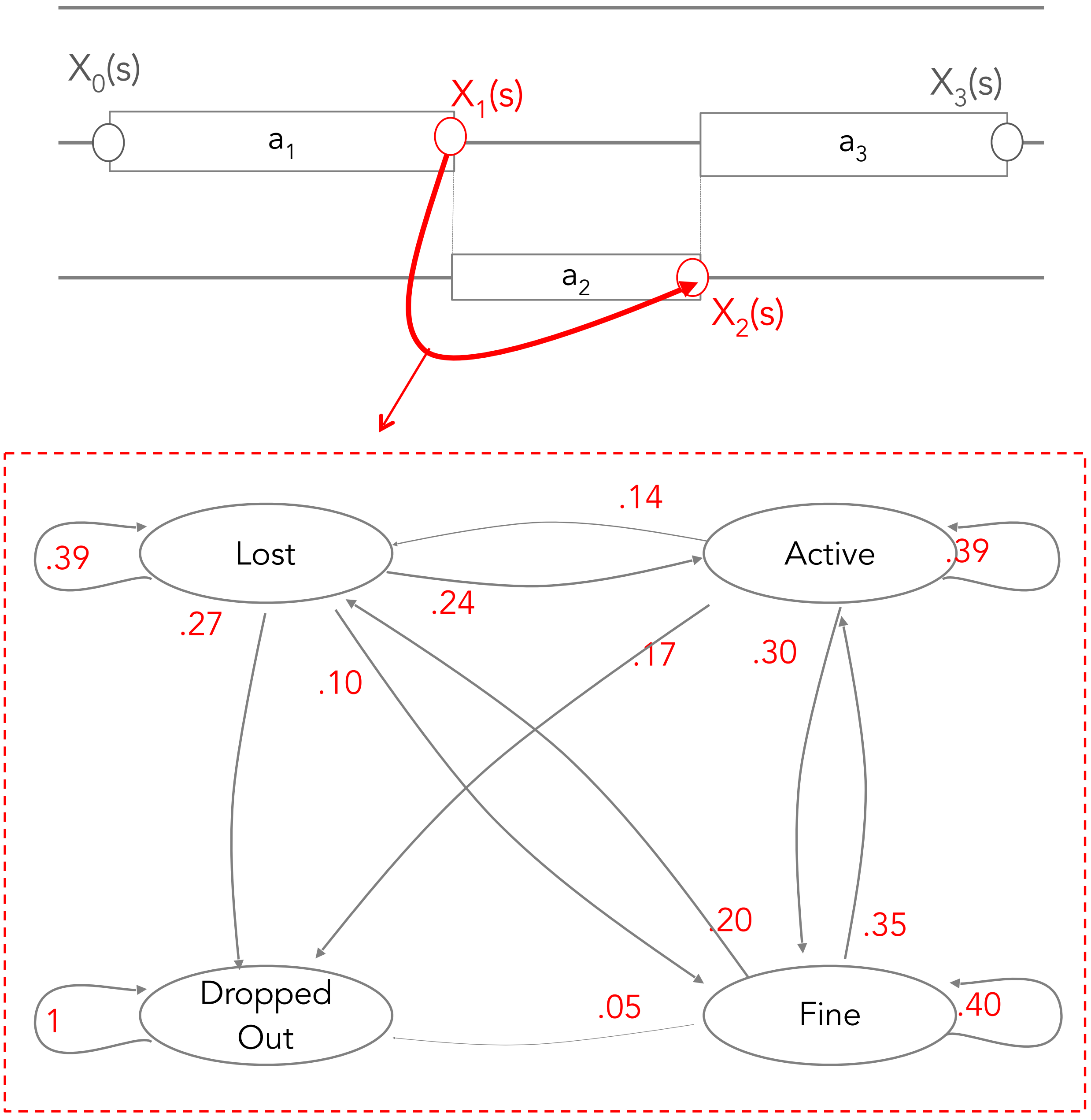
+

From the learner's previous state, **predict** his/her knowledge state

From the learner's behaviour and **his previous state,** infer his/her knowledge state



Bayesian Knowledge Tracing



Predicting from transition matrix

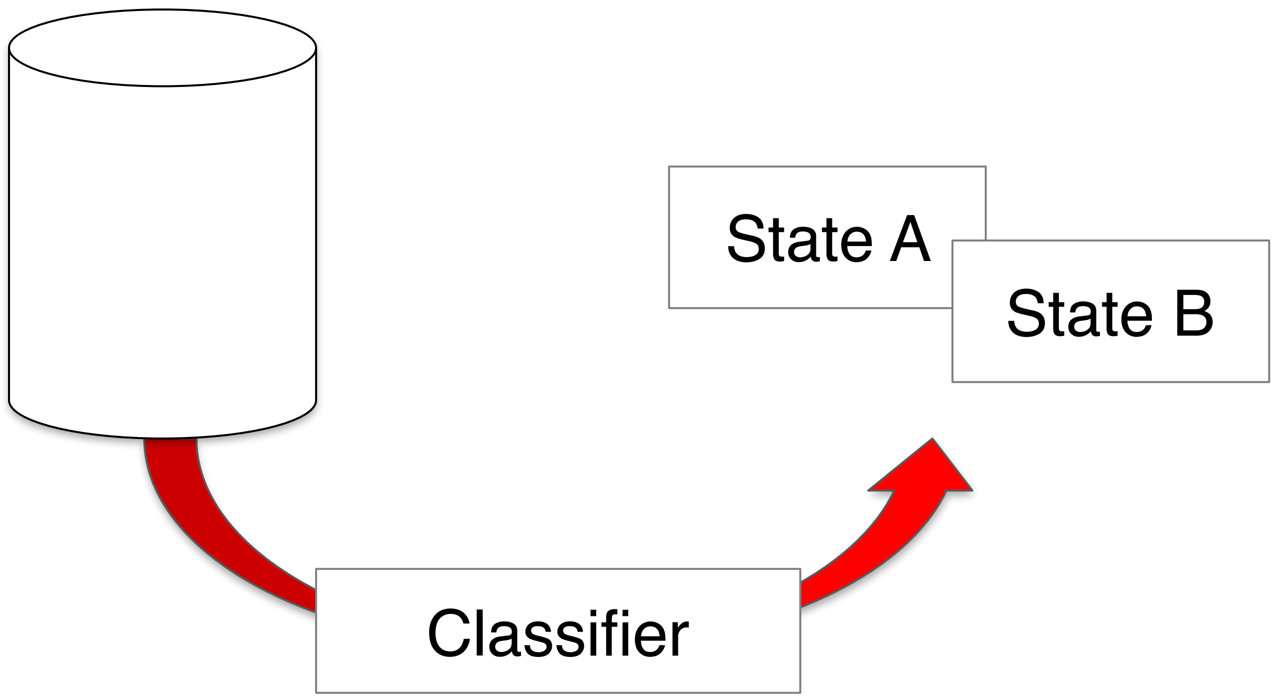
Learner Modelling

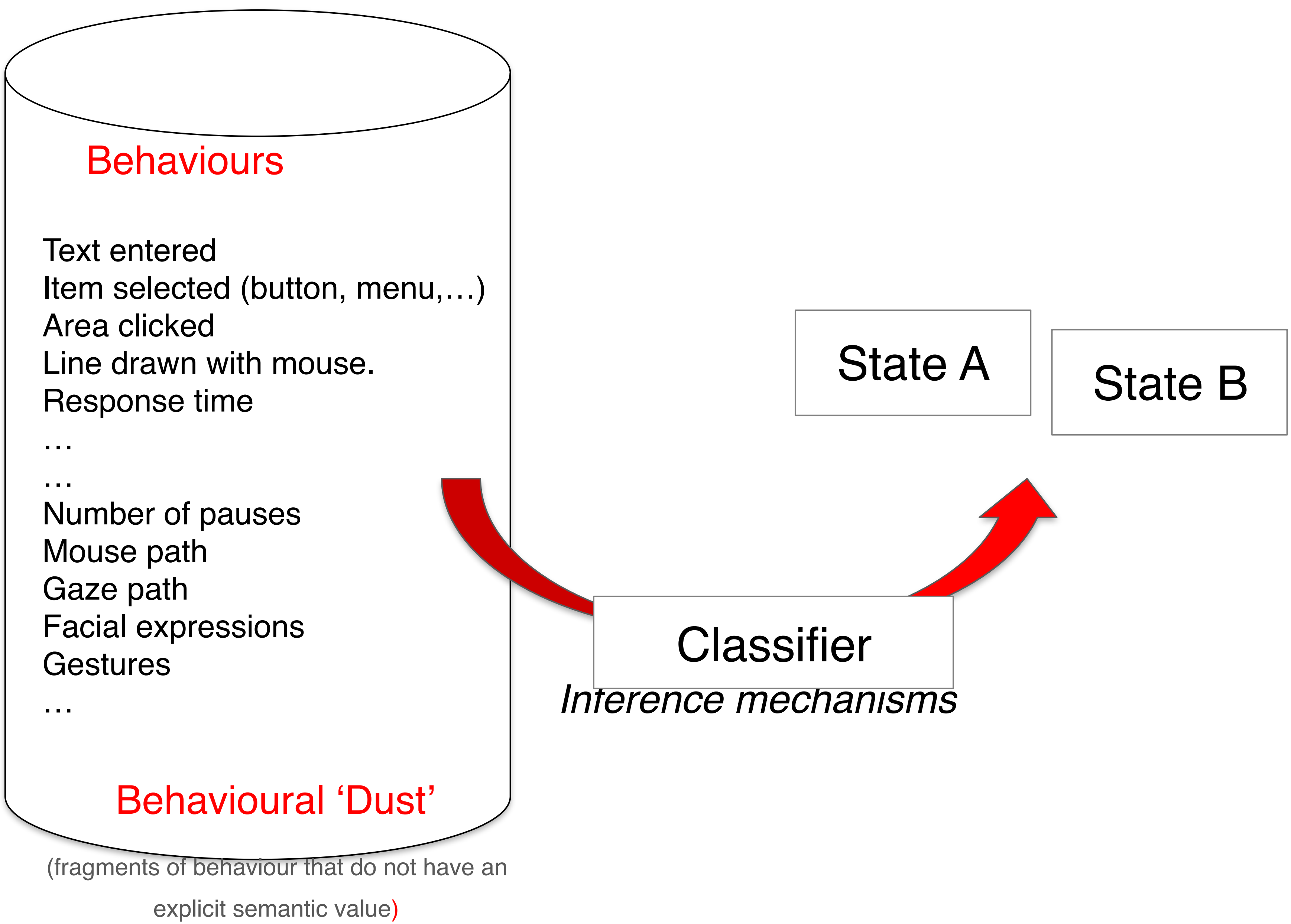
From the learner's behaviour, infer his/her knowledge state

?

From the learner's behaviour, **infer** his/her learner's knowledge state

From the learner's behaviour, infer his/her knowledge state



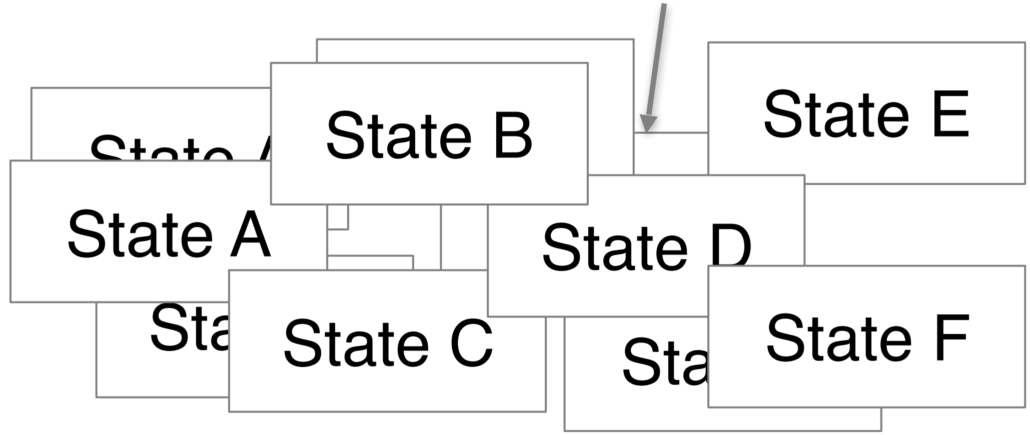

Behaviours
Text entered
Item selected (button, menu,…)
Area clicked
Line drawn with mouse.
Response time
…
…
Number of pauses
Mouse path
Gaze path
Facial expressions
Gestures
…
Behavioural 'Dust'
(fragments of behaviour that do not have an explicit semantic value)
Classifier
Inference mechanisms
State A
State B

From the learner's behaviour, **infer** his/her knowledge state

	Behaviours	Behavioural 'Dust'
3 Class Plane	The # messages in the forum	Head Co-Rotation
2 Team Plane	The concept map produced by a pair	Gaze Recurrence
1 Individual Plane	The learner answer to a quizz	Video 'Withmeness'

« Don't diagnose what you can't treat »

John A. Self



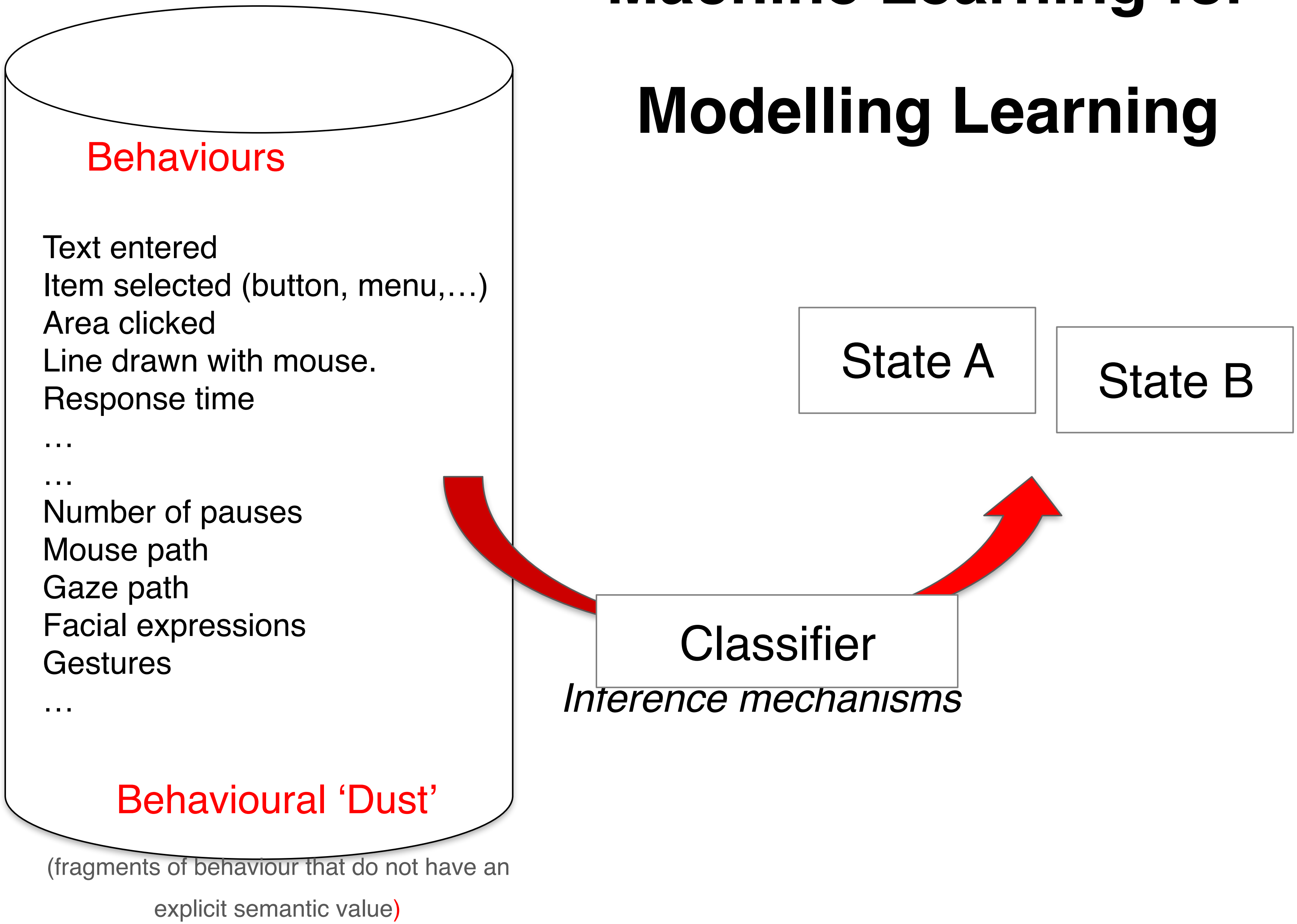
How many states ?

(Beyond the sake of cognitive research),

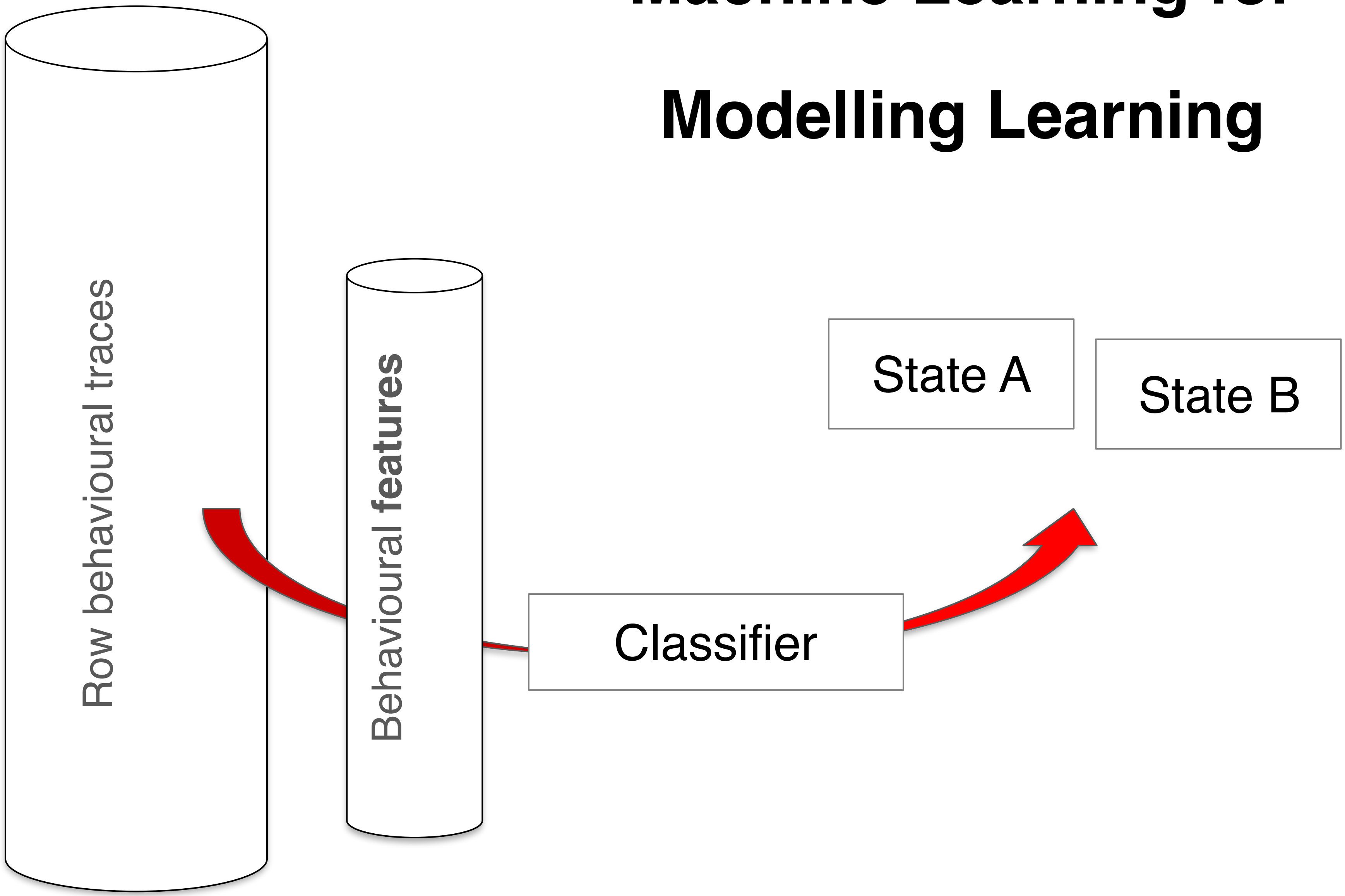
it is only interesting to discriminate state X and Y,

if the next decision will be different for X and Y

Machine Learning for Modelling Learning



Machine Learning for Modelling Learning



Relevant Behavioral Abstractions (Features)

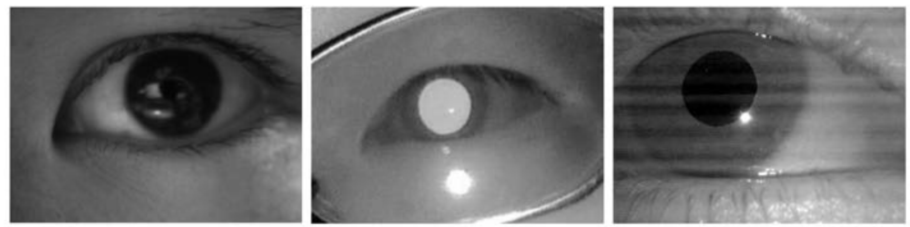
Computational Models

Education Research

Relevant Behavioral Abstractions (Features)

$$\mathrm{gaze}(a) = f(\mathrm{gaze}(b))$$

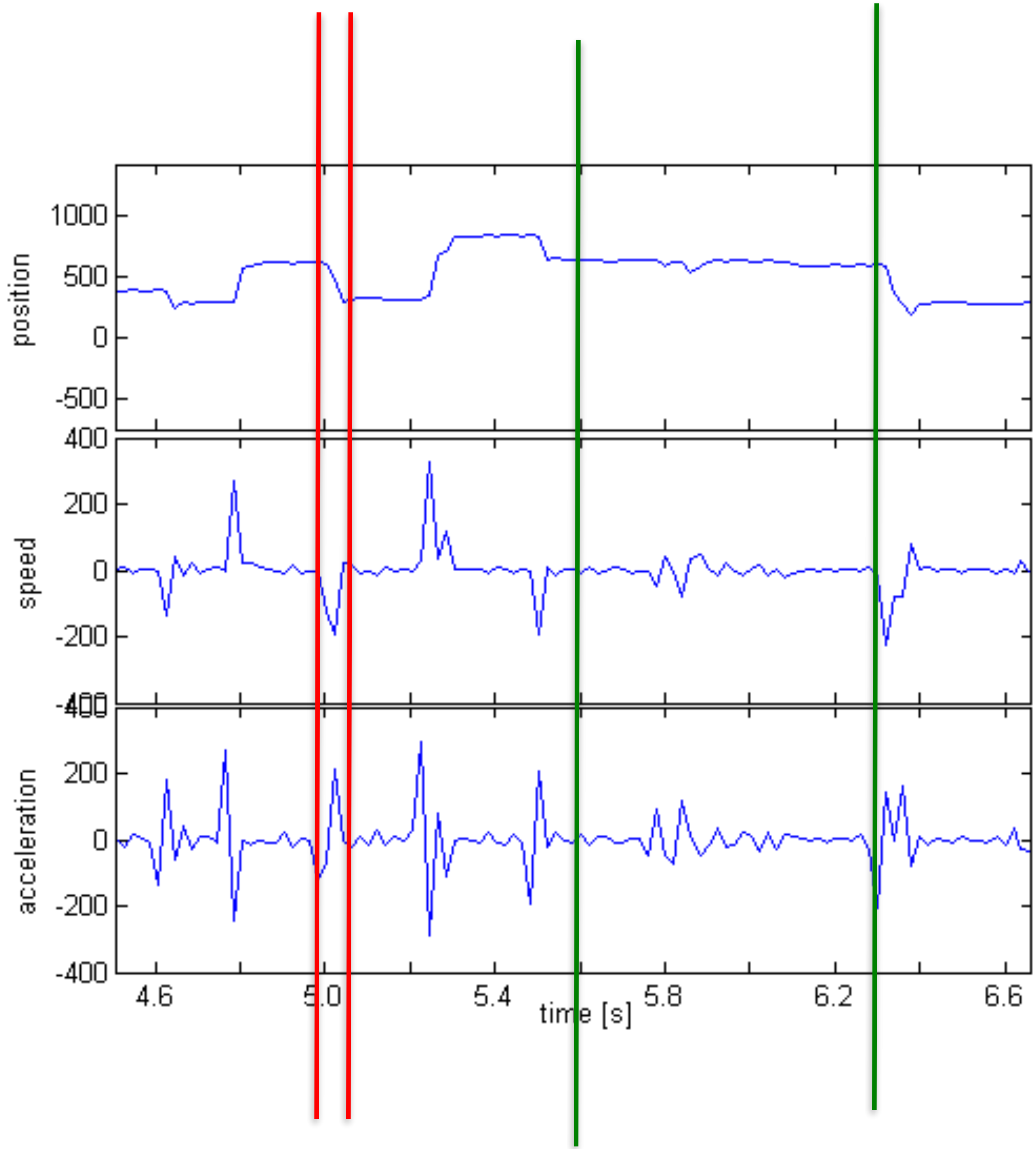
Eye Tracking



(a) (b) (c)

Tobii 1750

Youri Marko, Design of an eye tracker system for analyzing reading behavior, EPFL Master Thesis, 2011



Fixations

120 - 1000 ms

en général 200 -600

± 3 fois par sec.

Saccades

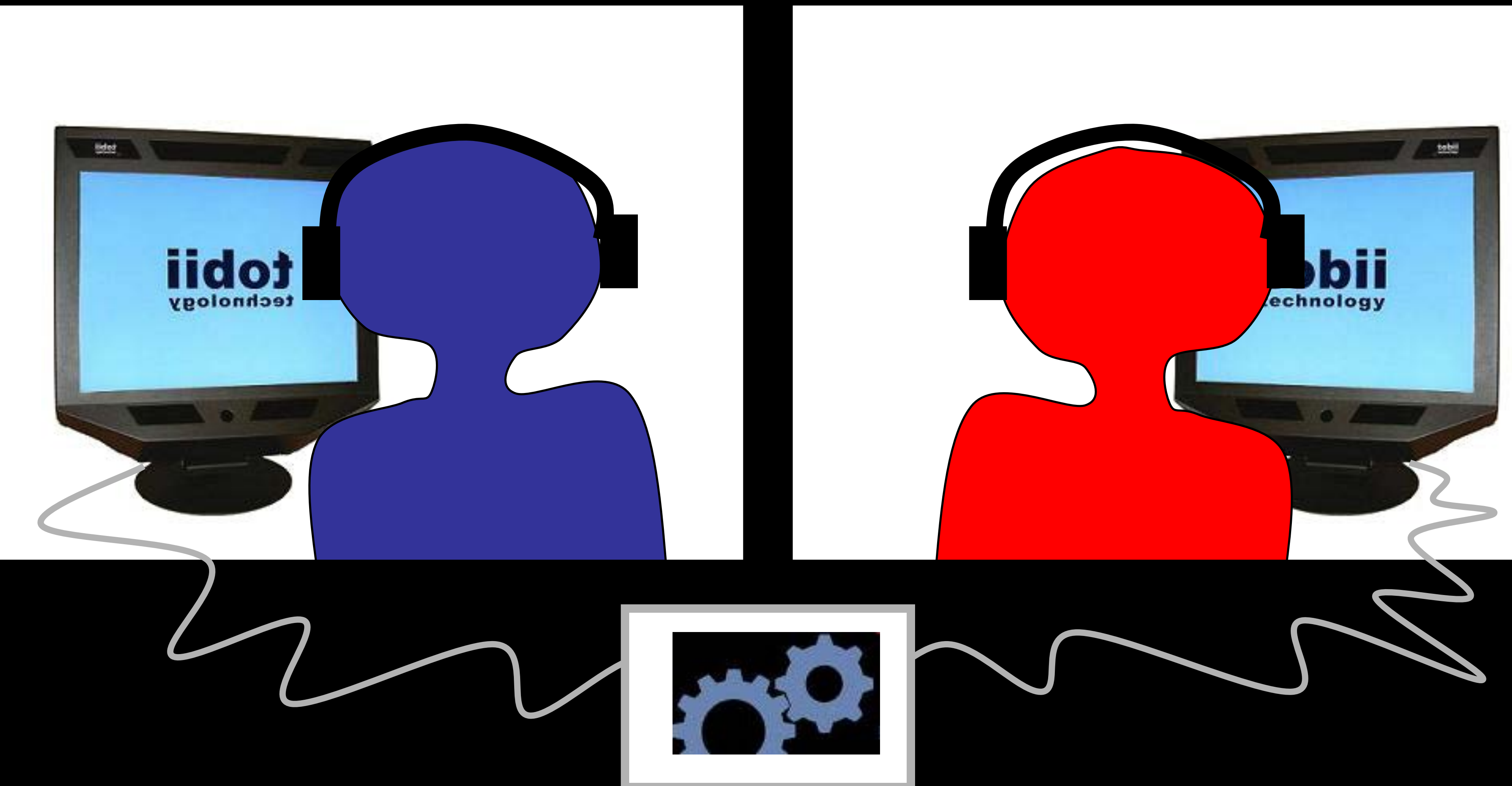
sauts rapides de l'œil

entre 40 et 120 ms

Nous sommes aveugles pendant la saccade

The (raw) eyetracking data

Timestamp [ms]	Category	Pupil size R [mm]	Pupil size L [mm]	Point of regard X	Point of regard Y	...
87542.5	Blink	3	2.9	936.3	691.7	
87575.7	Blink	3	2.8	908.6	639.5	
87609.2	Visual Intake	3	2.9	873.7	613.7	
87642.5	Visual Intake	3	2.9	851.3	608.9	
87675.8	Visual Intake	3	3	828.5	603.1	
87709.2	Visual Intake	3	3	809.1	613.9	
87742.3	Visual Intake	3.1	3	794.1	618.1	
87775.6	Visual Intake	3.1	3.1	783.7	627.1	
87808.8	Visual Intake	3.2	3.1	771.4	633.7	
87842.1	Saccade	3.1	3.2	769.3	651.5	
87875.3	Saccade	3.2	3.2	767.7	671.3	
87908.6	Saccade	3.2	3.2	764	679.8	
87941.8	Visual Intake	3.2	3.2	759	686.1	
87975.3	Visual Intake	3.2	3.2	758.9	690.9	
...	...	...	...	...	...	...



DUET - Dual Eye-Tracking
Pair programming experiment

Low gaze recurrence

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

P. Jermann, M.-A. Nüssli & P. Dillenbourg

Supported by the Swiss National Science Foundation
(grants #K-12K1-117909 and #PZ00P_126611)

DUET - Dual Eye-Tracking
Pair programming experiment

High gaze recurrence

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

P. Jermann, M.-A. Nüssli & P. Dillenbourg

Supported by the Swiss National Science Foundation
(grants #K-12K1-117909 and #PZ00P_126611)

Relevant Behavioral Abstractions

$$\mathrm{gaze}(\mathrm{listener}) = f(\mathrm{gaze}(\mathrm{speaker}))$$

Feature: *Gaze recurrence*

Context: *Collaborative learning*

Eye tracking experiment on MOOC Video

Following teacher's references

Gaze of students' watching Scala course by Prof. Martin Odersky (EPFL, Switzerland)

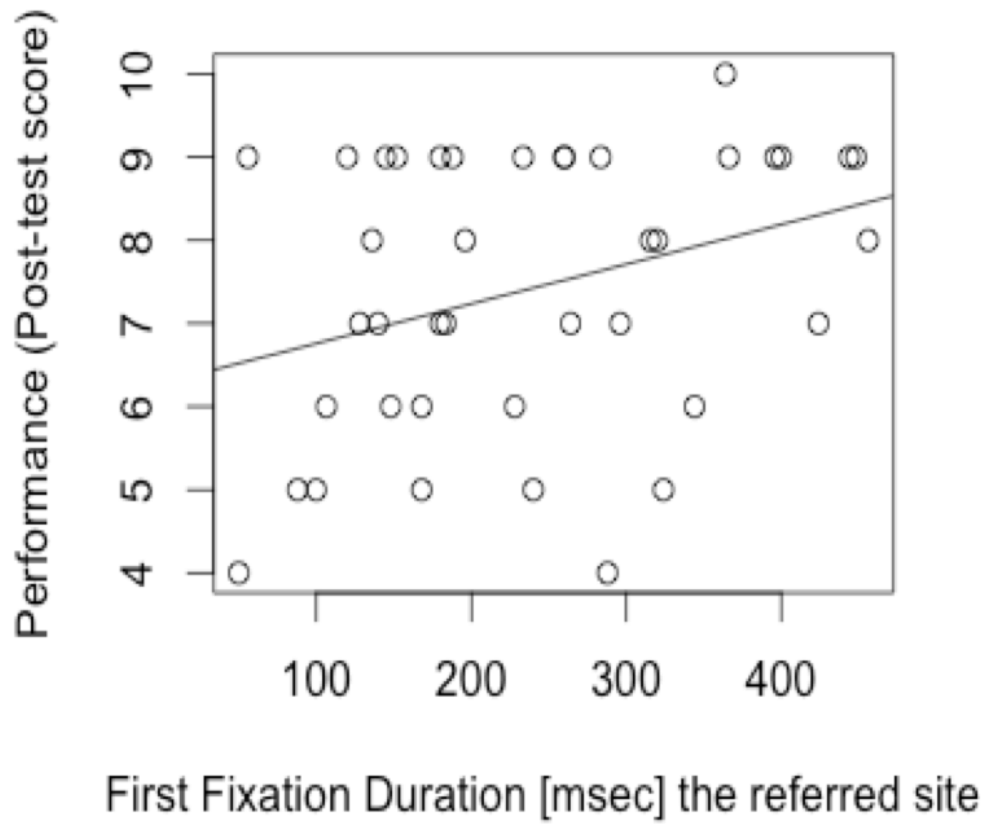
ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

K. Sharma, P. Jermann, P. Dillenbourg
@ CHILI – http://chili.epfl.ch

Supported by the Swiss National Science Foundation
(Grants CR1211_132996 and PZ00P2_126611)

Relevant Behavioral Abstractions

gaze (learner) = *f* (reference (teacher))



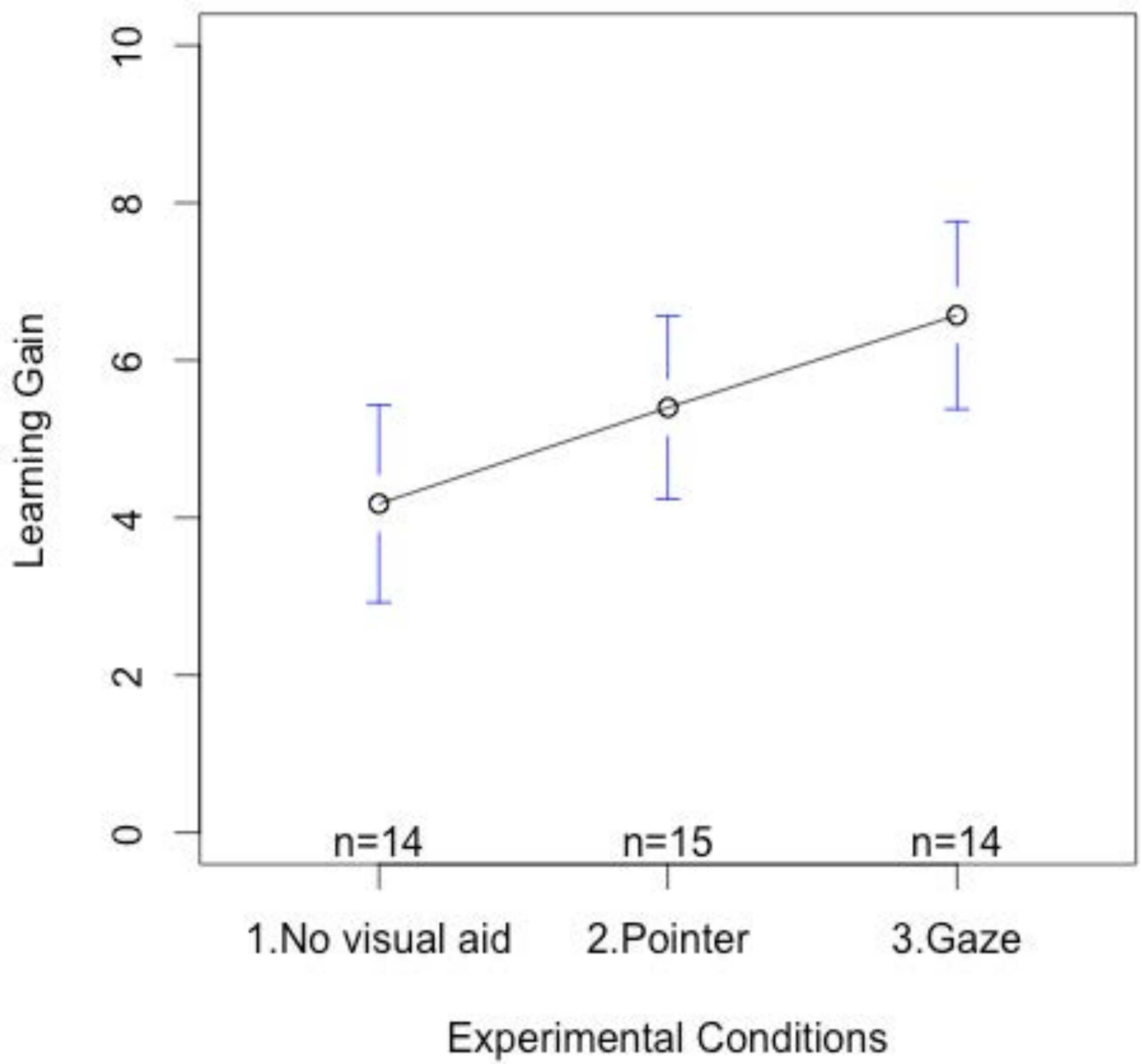
Feature: *Withmeness*

Context: *Lecturing*

"...they look like a bunch of little grains arranged together...typically a group of very small elements"

Sarah d'Angelo, Kshitij Sharma, Darren Gergle, Pierre Dillenbourg (2016)

Do finger-based or gaze-based deictics enhance learning ?



Sarah d'Angelo, Kshitij Sharma, Darren Gergle, Pierre Dillenbourg (2016)

Relevant Behavioral Abstractions

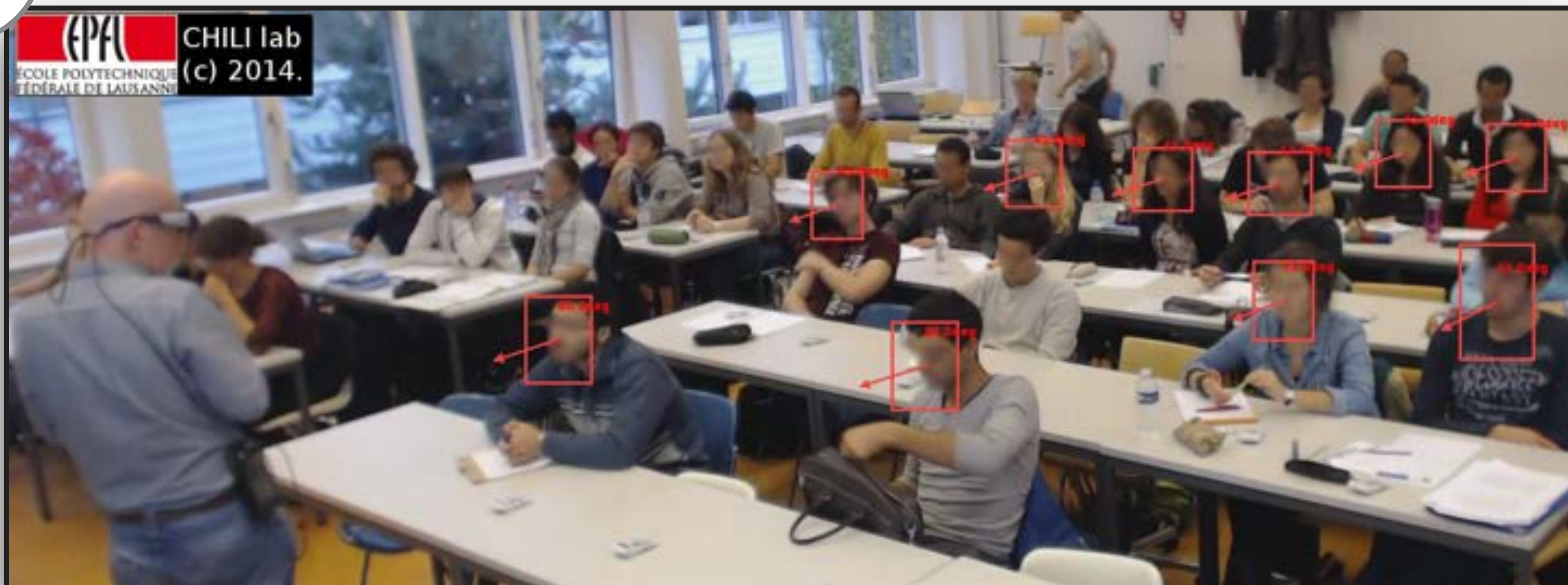
$$\text{gaze (learner)} = f \text{ (gaze (teacher))}$$

<u>Feature</u>: *'Withmeness'*

<u>Context</u>: *Lecturing*

Modeling in the wild ?

K_t



Kernel	Features	Score	Cohen's kappa
RBF(c=1.31, g=0.0211)	Distance, Head travel norm., Num. still periods	61.86%	0.30
BF(c=1.21, g=0.11)	Period, Row, Head travel norm., Mean duration still	61.72%	0.32
BF(c=1.11, g=0.061)	Head travel norm., Mean duration still	60.42%	0.28
BF(c=1.4, g=0.04)	Period, Distance, Row, Mean duration still	59.23%	0.30

B_t

Raca, Tormey & Dillenbourg

Relevant Behavioral Abstractions

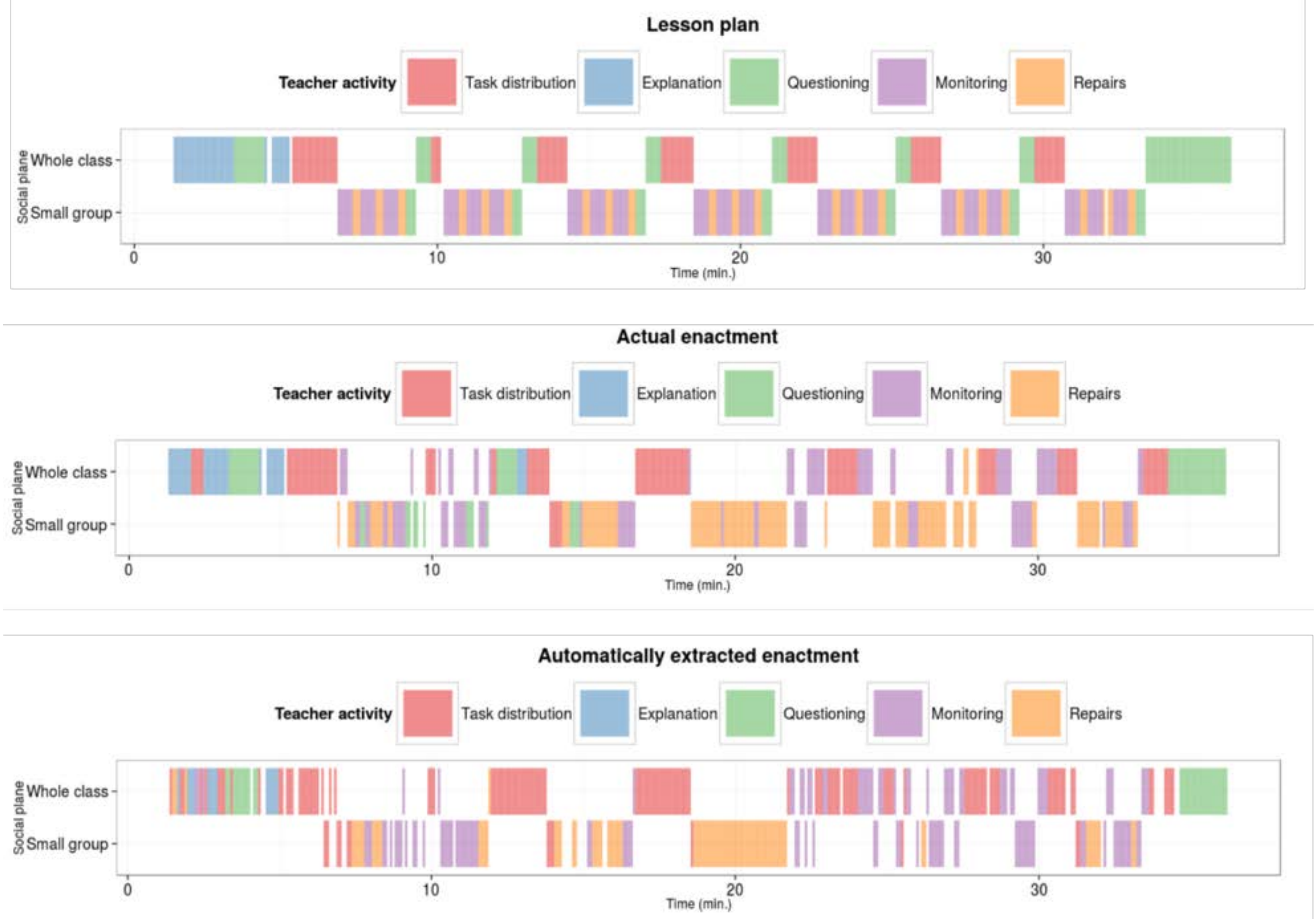
$$\text{gaze (learner)} = f\ \text{(location (teacher))}$$

<u>Feature</u>: *Head rotations*

<u>Context</u>: *Lecturing*

activity (teacher) = f (gaze (teacher))

L. Prieto, K. Sharma, L. Kidzinsky, P. Dillenbourg



Computational Modelling of Education

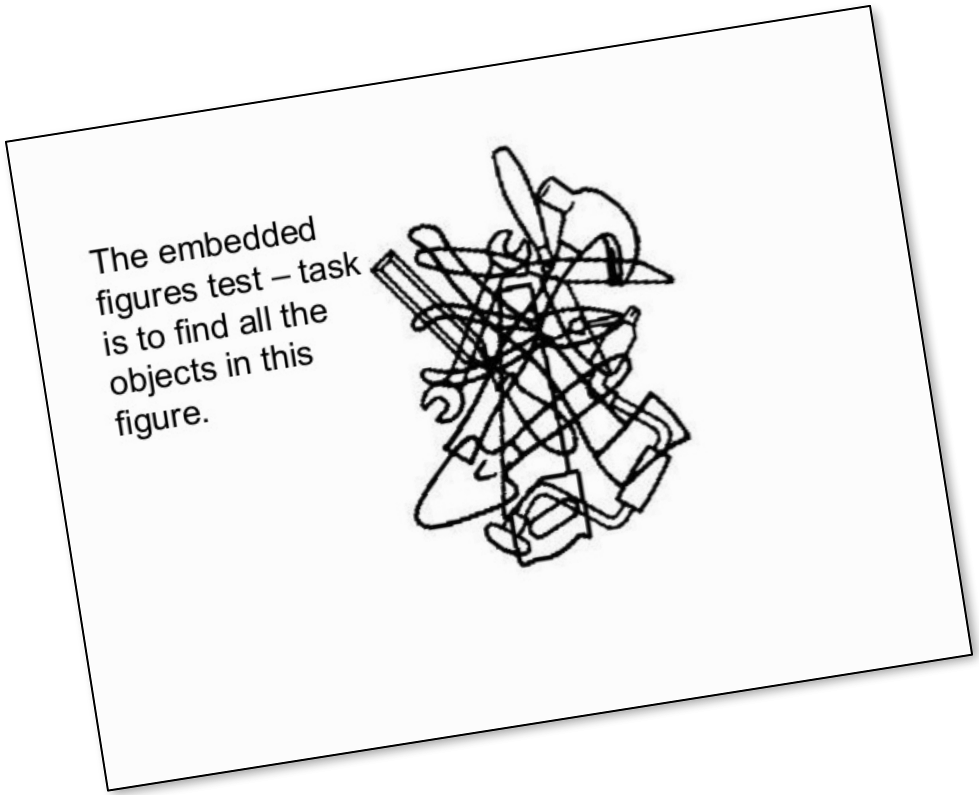
Relevant Behavioral Abstractions

$$\text{activity(teacher)} = f\ (\text{location (teacher)})$$

Feature: *pupil diameter, #faces in field of view*

Context: *Lecturing*

Limitations



Measured at time t

Stable in time

State ≠ Trait

State	Trait
• Knows	• Anxious / Self-confident
• Understands	• Risk-aversive / Risk-seeking
• Performs	• Aural / visual / kinesthetic
• Engaged	• Deep / Surface
• Bored	• Field-dependent / independent

Learning styles

Severe criticisms:

- Contextual rather than personal
- No clear effects of adaptation
- Should education mimic style or counterbalance them ?
- Labels produce self-fulfilling prophecies
-

BEWARE OF
the medicalisation of Education !!!

- Learning disabilities, LD
- Attention-deficit disorder, ADD
- Attention-deficit hypeactivty disorder, ADHD
- Dysgraphia, dyscalculia, dyslexia, ….
- High-potential chidren, HP
- ….

Labels help Sales

Ethics of Algorithms !

- Social Determinism
- The ’halo effect’
- Biased data sets (over-representation of a subpopulation)
- Over-fitting
- …

Education raises challenges to data science

- *Explainaibility : which features make sense*
- *Exploration/Exploitation trade-off*
- *Cold Start: simulations, expert's knowledge*

Education is a computational science

EPFL Center for Learning Sciences