CS-411 : Digital Education & Learning Analytics

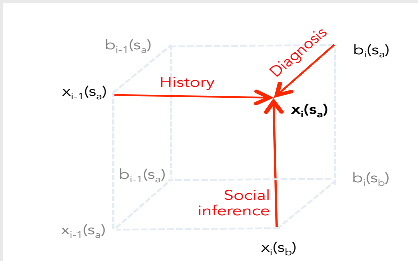
Chapter 9:

Learner modeling

Pierre Dillenbourg and Patrick Jerman

You are here

Learning Design
Orchestration
Graphs
Activities



Learning Analytics
Statistical models
Probabilistic inference

CS-411 : Digital Education & Learning Analytics

$$p\,(\text{Trump} \mid \text{Snow}) = ?$$

If the first snow comes on a Tuesday…

something will happen somewhere someday

Learner Modelling

$$5^2 = ?$$

Learner Modelling

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

$5^2 = 25$ ➔ knows X^2

$5^2 \neq 25$ ➔ doesn't know X^2

Learner Modelling

$5^2 = ??$	Knowledge States								
Behavior (Answer)	$5^2 = 25$	$5^n = \dots$	$n^2 = n\,.\,N$	$x^n = x\,.\,x.\,\dots$	$x^n = x\,.\,x$ but bad mult.	$x^n = x.n$	$x^n = x + n$	$x^n = ???$	
25	0.25	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1
35	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1
10	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1
27	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1
7	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1

Learner Modelling

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

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

Potential values

Actual value

$$x_i\,(s) \in X_i(S) = \{\text{fine, active, lost, drop}\}$$

At activity a_i

Actual student

All students

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)

From the learner's behaviours, infer his/her learner's 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) = post a message "There is a mistake on the slide"
(and there is one indeed)

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

H_0 = .80

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

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

H_0 = .18

Normalized entropy of the diagnosis vector

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

x(s)=[.15 .40 .30 .15] $\longrightarrow$ $\boxed{H_0 = 0.94}$

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

Write a question that

- determines if the learner understood the concept of standard deviation;
- has a high diagnosis power
- can be automatically graded

Learner Modelling

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

1. The basic approach
2. The good old AI approach
3. The data crunching approach
4. The Bayesian approach
5. The Markov approach

(1) Basic approach to learner Modelling

Decrease uncertainty by collecting multiple answers

5^2 = ??	7^2 = ??	Knowledge States								
Behavior (Answer)		$5^2 = 25$	$5^n = \ldots$	$n^2 = n . N$	$x^n = x . x. \ldots$	$x^n = x . x$ but bad mult.	$x^n = x.n$	$x^n = x + n$	$x^n = ???$	
25	49	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1
25	21	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1
35	49	0.125	0.125	0.125	0.125	0.125	0.125	0.125	0.125	1

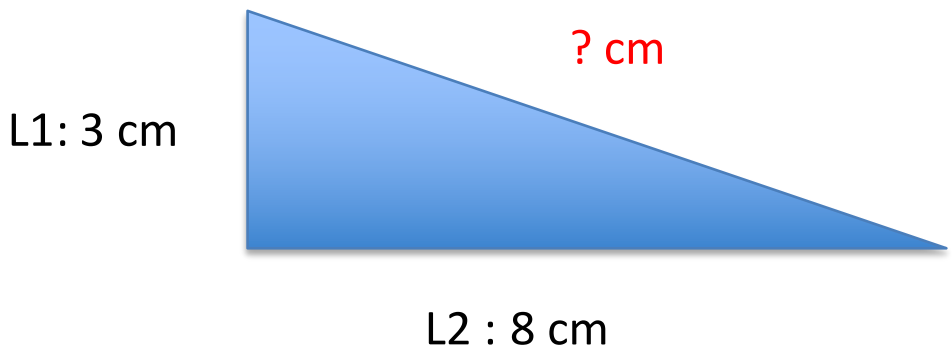
If the learner makes more than n% errors in a_i,

then (s)he is in state « low understanding »

(2) Learner Modelling in symbolic AI

To compute the length of the hypothenuse

1. Measure the length, L1 and L2
2. Compute $L1^2$ and $L2^2$
3. Sum them
4. Extract the square root

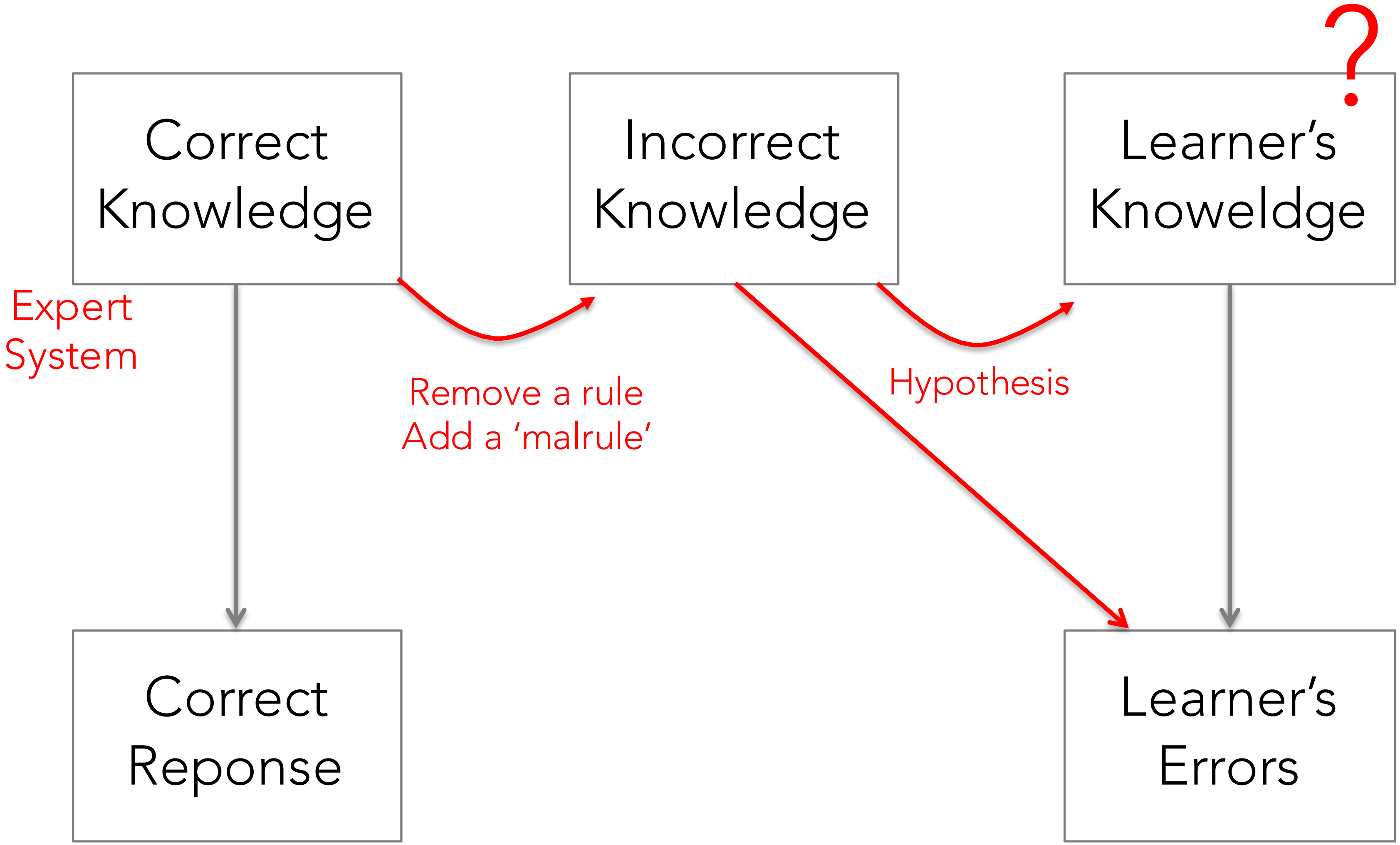


Correct answer = 8.54

Learner Answer = 8.18

What did he do wrong ?

(2) Learner Modelling in symbolic AI



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

(3) Learner Modelling with data sciences

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

Behaviours

- Text entered
- Item selected (button, menu,…)
- Area clicked
- Line drawn with mouse.
- Response time
- …

…

Behavioural 'Dust'

(fragments of behaviour that do not have an explicit semantic value)

- Number of pauses
- Mouse path
- Gaze path
- Facial expressions
- Gestures
- …

« Social Signal Processing »

Example 1: From the learner's gaze, infer the « withmeness »

Eye tracking experiment on MOOC Video

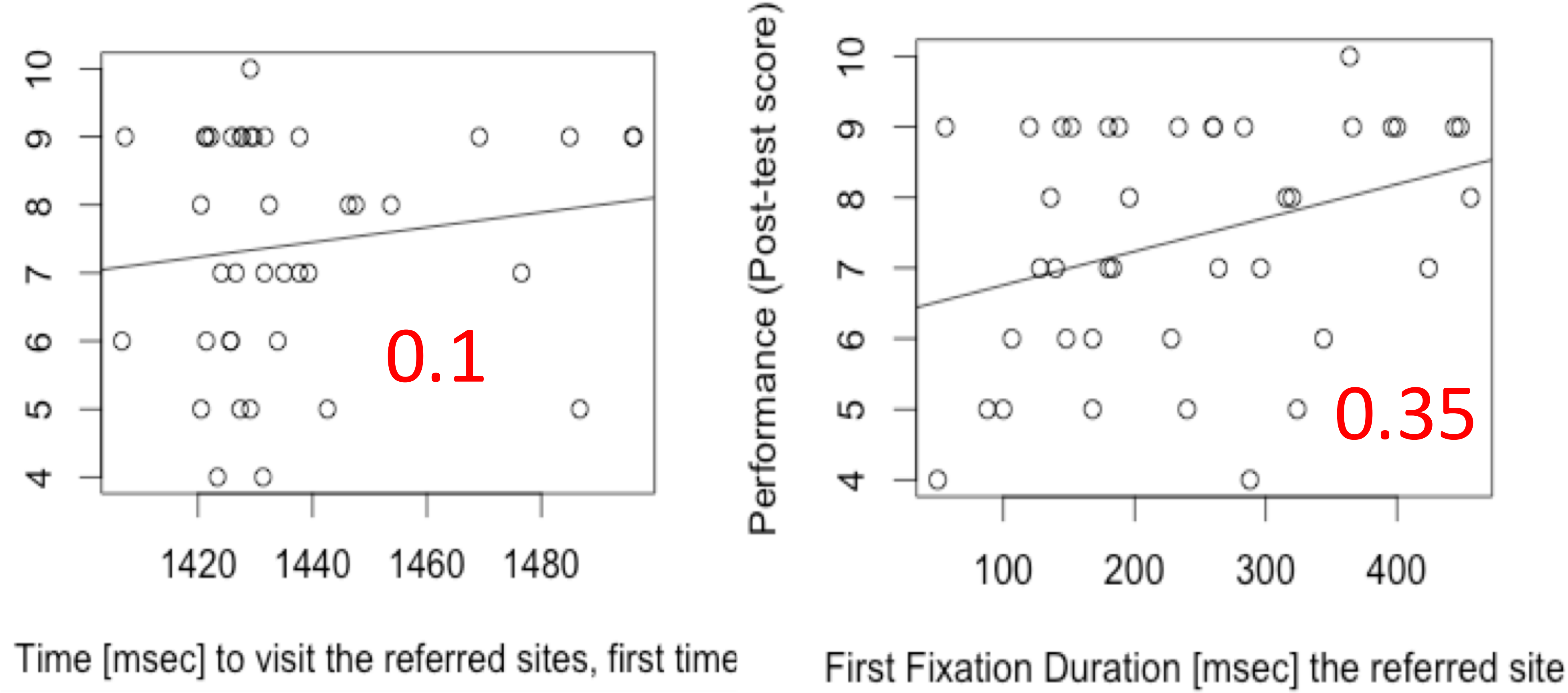
Following teacher's references

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

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

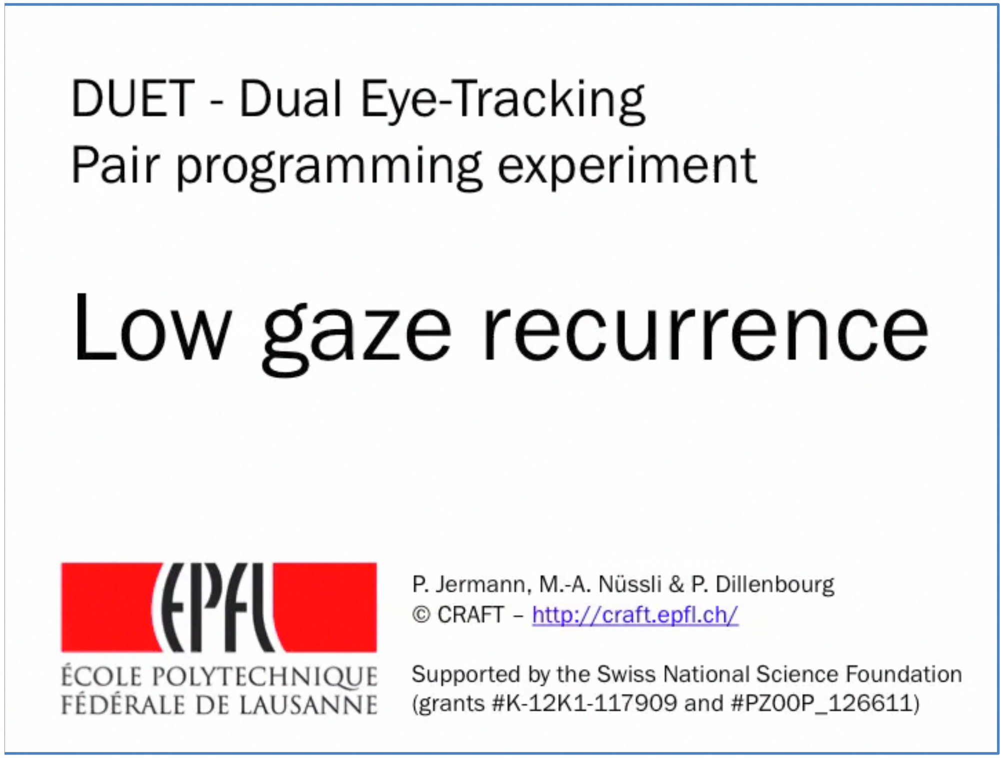
K. Sharma, P. Jermann, P. Dillenbourg
@ CHILI – http://chili.epfl.ch
Supported by the Swiss National Science Foundation
(Grants CR1211_132996 and PZ00P2_126611)

Example 1: From the learner's gaze, infer the « withmeness » because it predicts learning gains

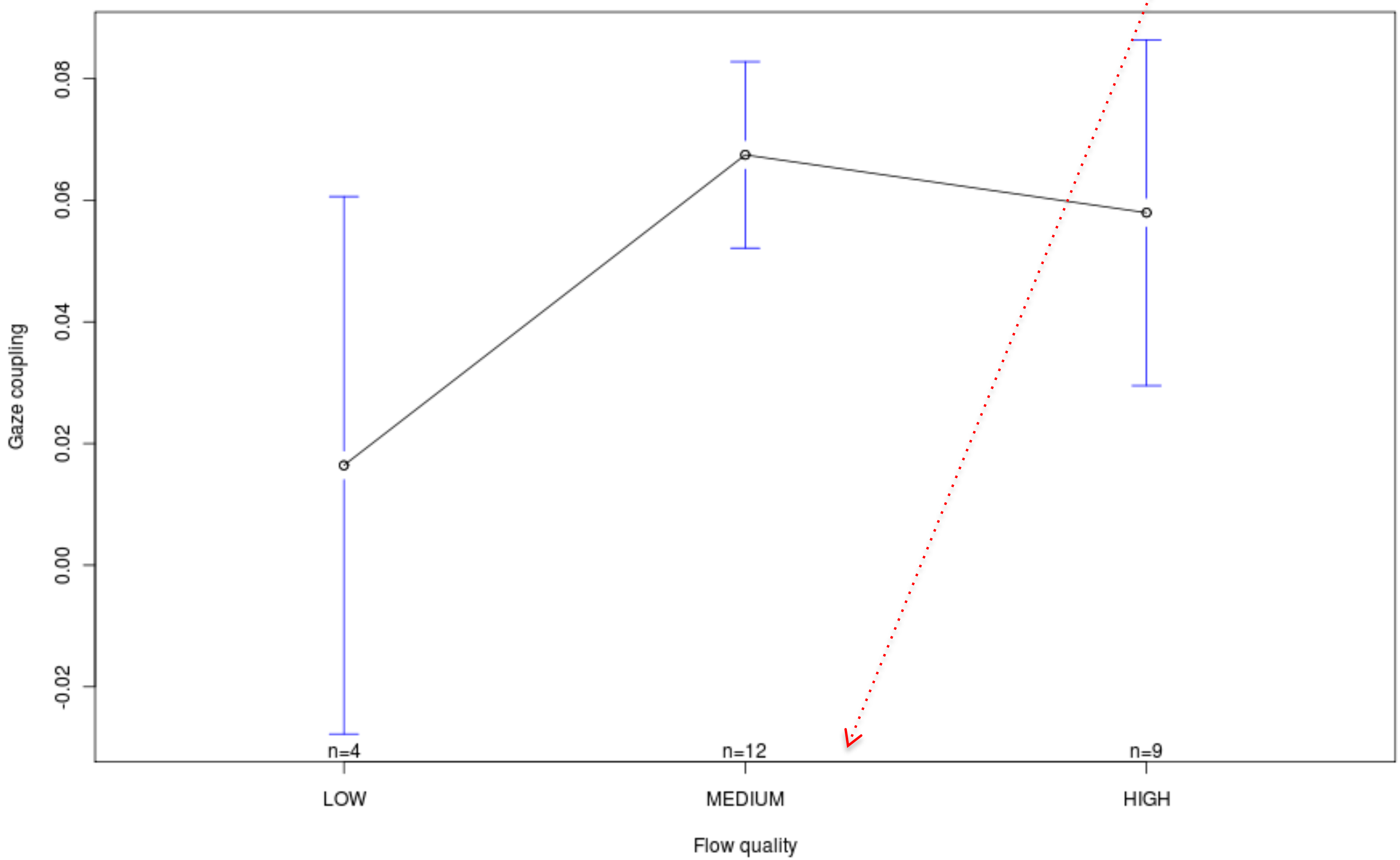


Kshitij Sharma, Patrick Jermann, Pierre Dillenbourg
EPFL Center for Digital Education

Example 2: From 2 learners gazes, infer the quality of collaboration



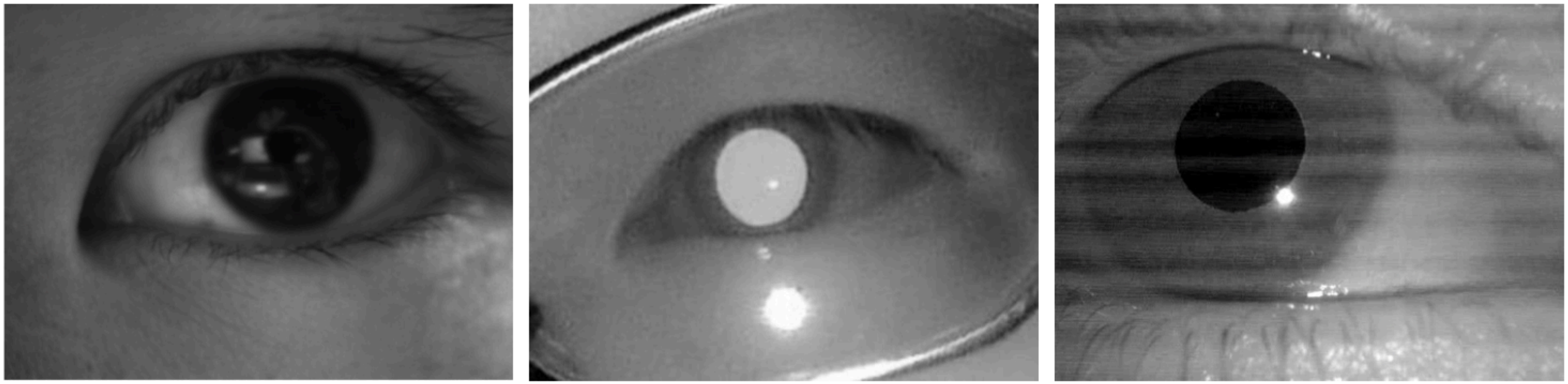
Example 2: From 2 learners gazes, infer the quality of collaboration



The pairs that collaborate well tend to look ± at the same time at ± the same object

Next Week

08:15 - 10:00	Eye tracking methods Kshitij Sharma
10:15 – 12:00	Learning analytics Try an eye tracker



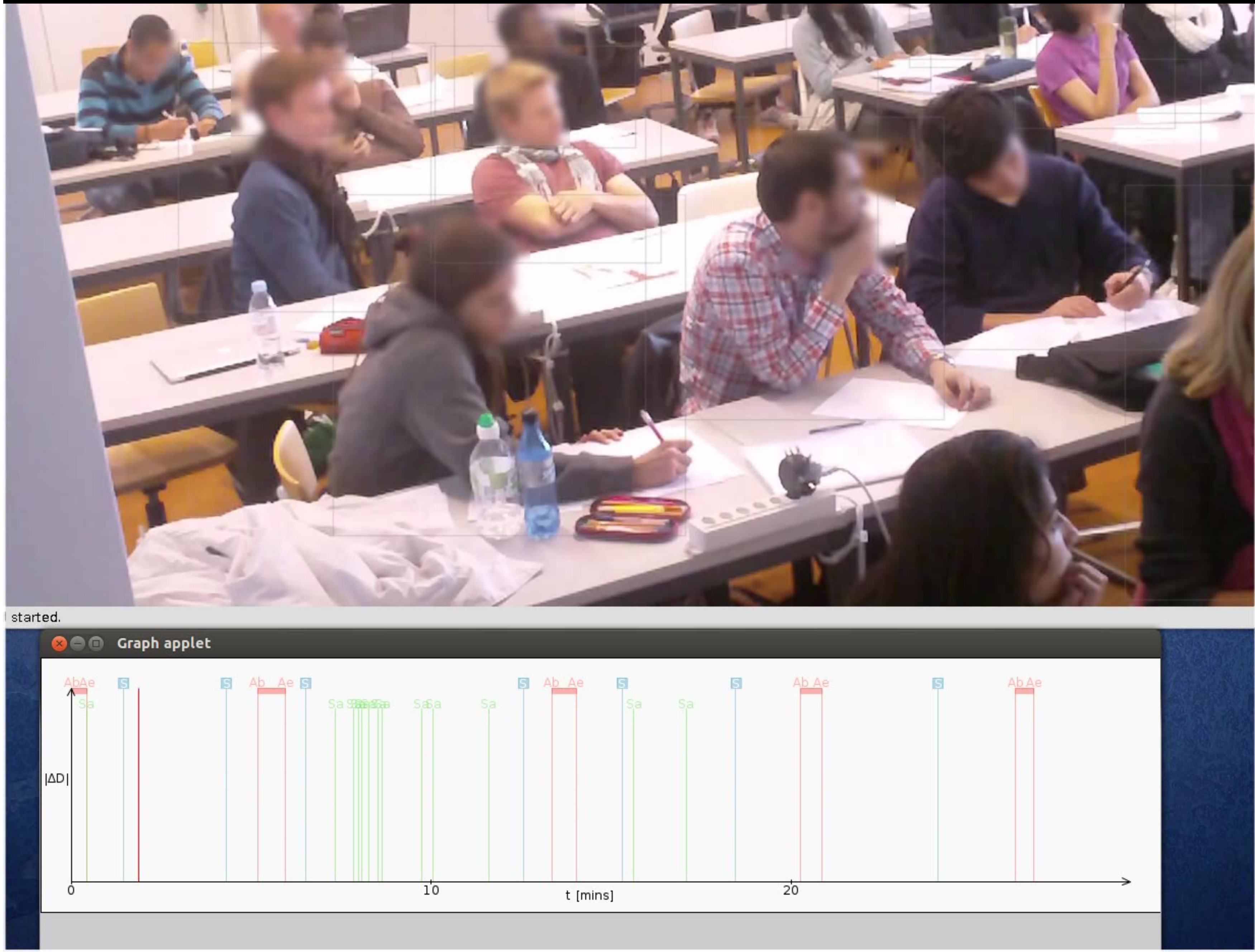
(a) (b) (c)

Learner Modelling

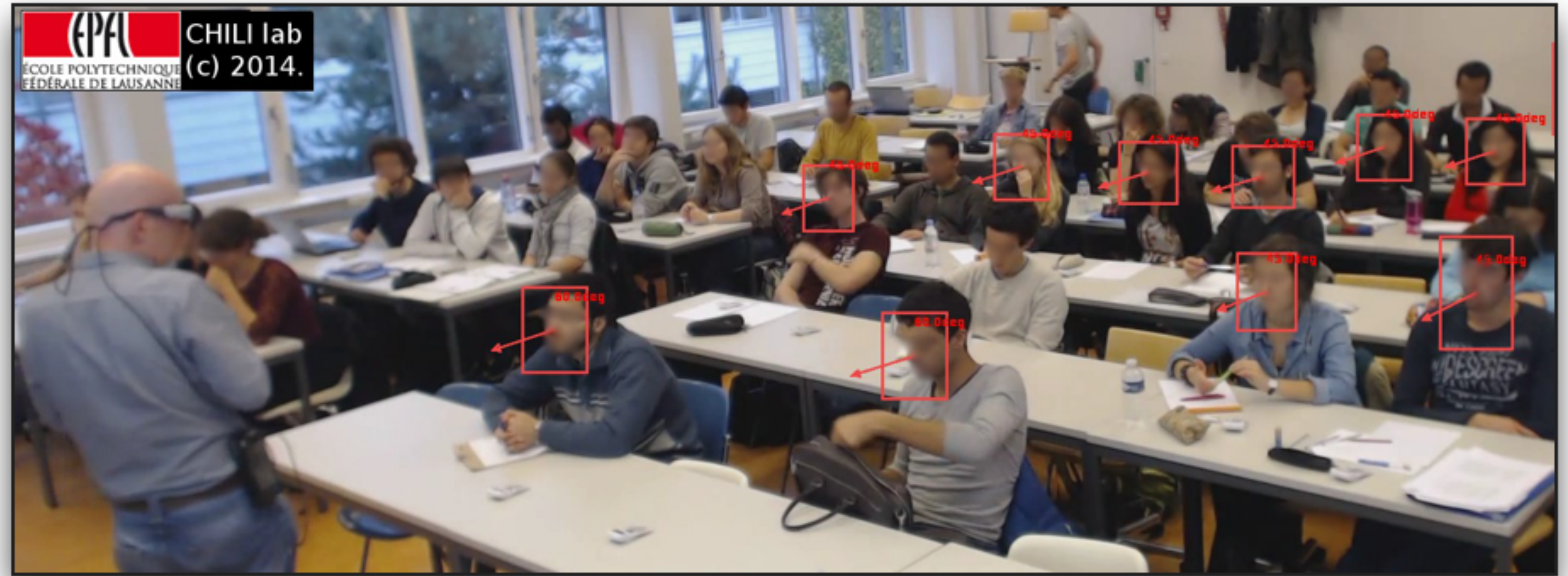
From the learner's behaviour, infer his/her learner's knowledge state

?

Example 3: From the learner's (co)movements, infer the class level of attention



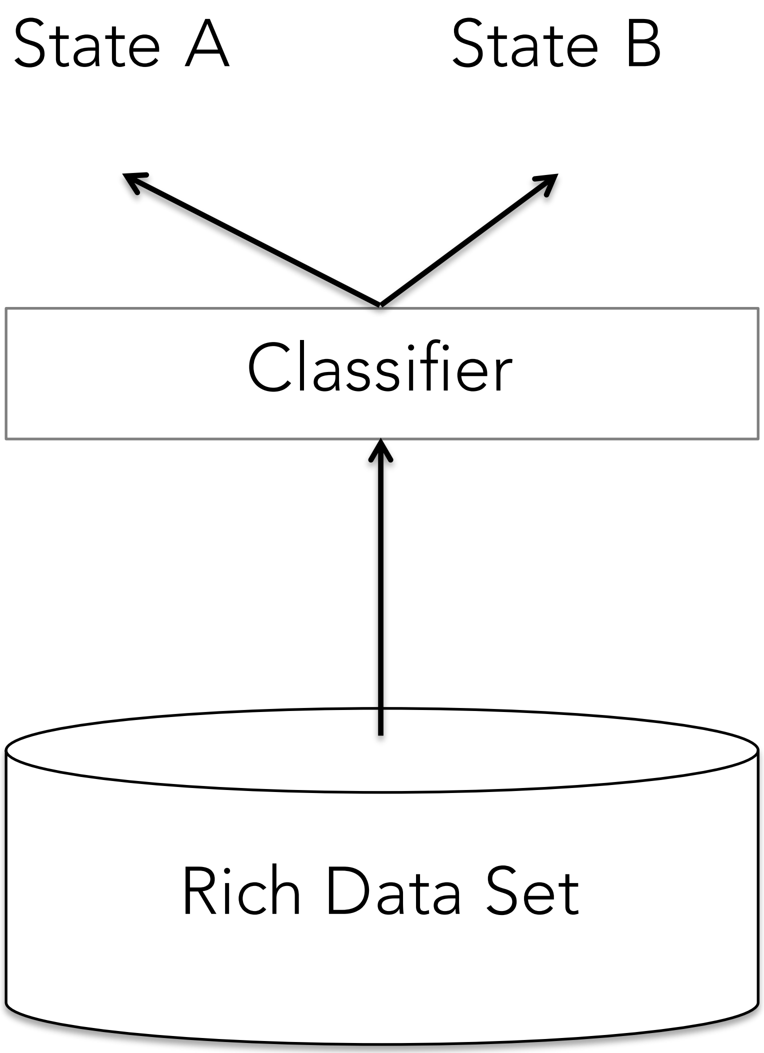
Mirko Racca, CHILI Lab, EPFL



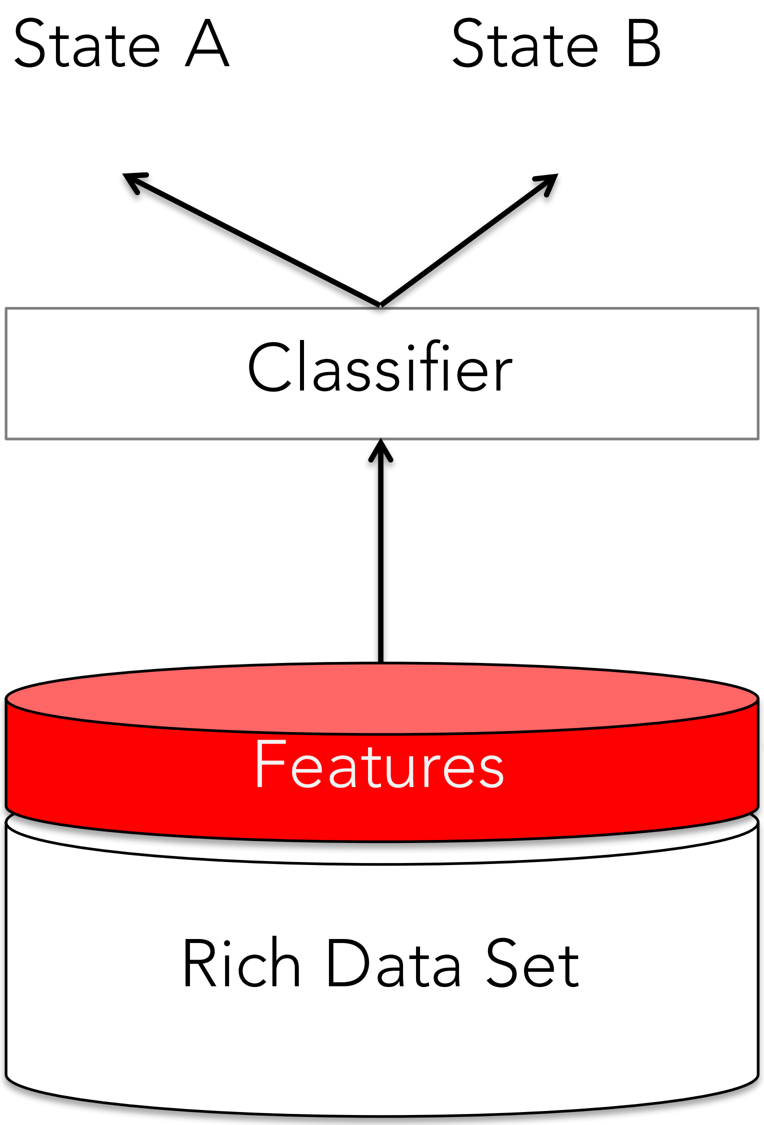
Kernel	Features	Score	Cohen's kappa
RBF(c=1.31, g=0.0211)	Distance, Head travel norm., Num. still periods	61.86%	0.30
RBF(c=1.21, g=0.11)	Period, Row, Head travel norm., Mean duration still	61.72%	0.32
RBF(c=1.11, g=0.061)	Head travel norm., Mean duration still	60.42%	0.28
RBF(c=1.4, g=0.04)	Period,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

	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'

Learner Modelling @ DataScience Times



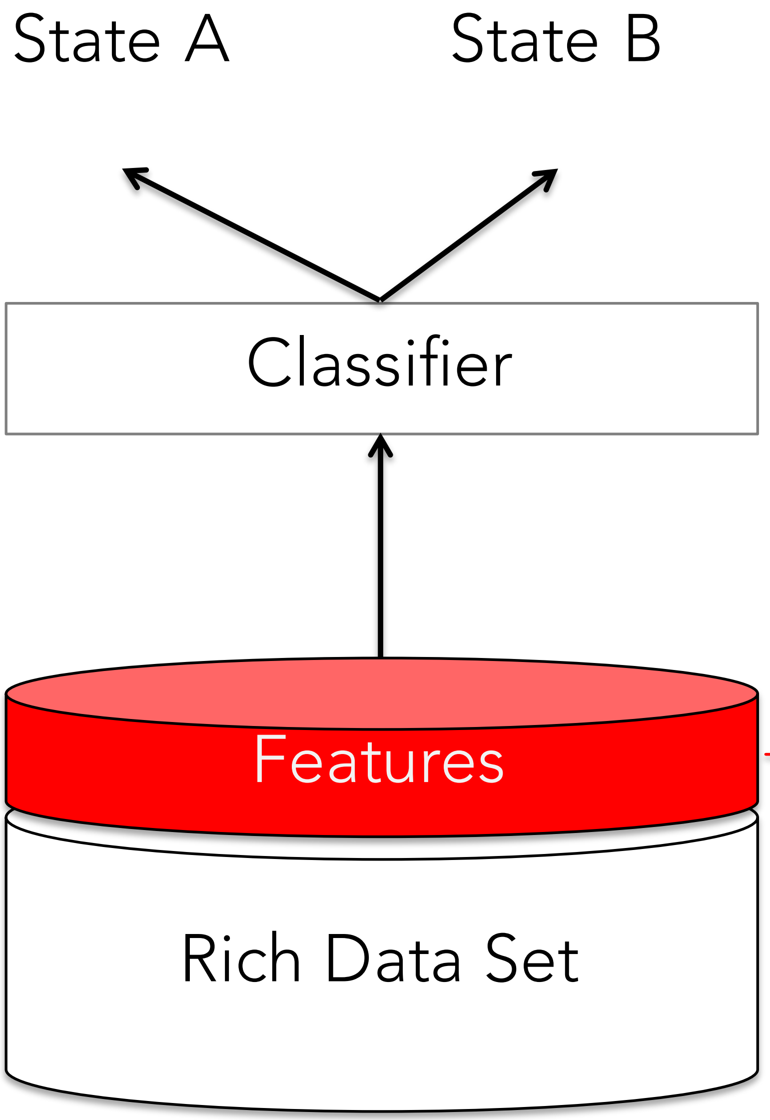
Learner Modelling @ DataScience Times



Π_1

- Rate of back/cancel actions in a navigation task.
- Redundancy: Did the learner ask a question for which he already had an answer?
- "With-me-ness": Did the learner look at the object mentioned by the lecturer in the video?
- Attention map: Which areas does the learner look at most often?
- …..

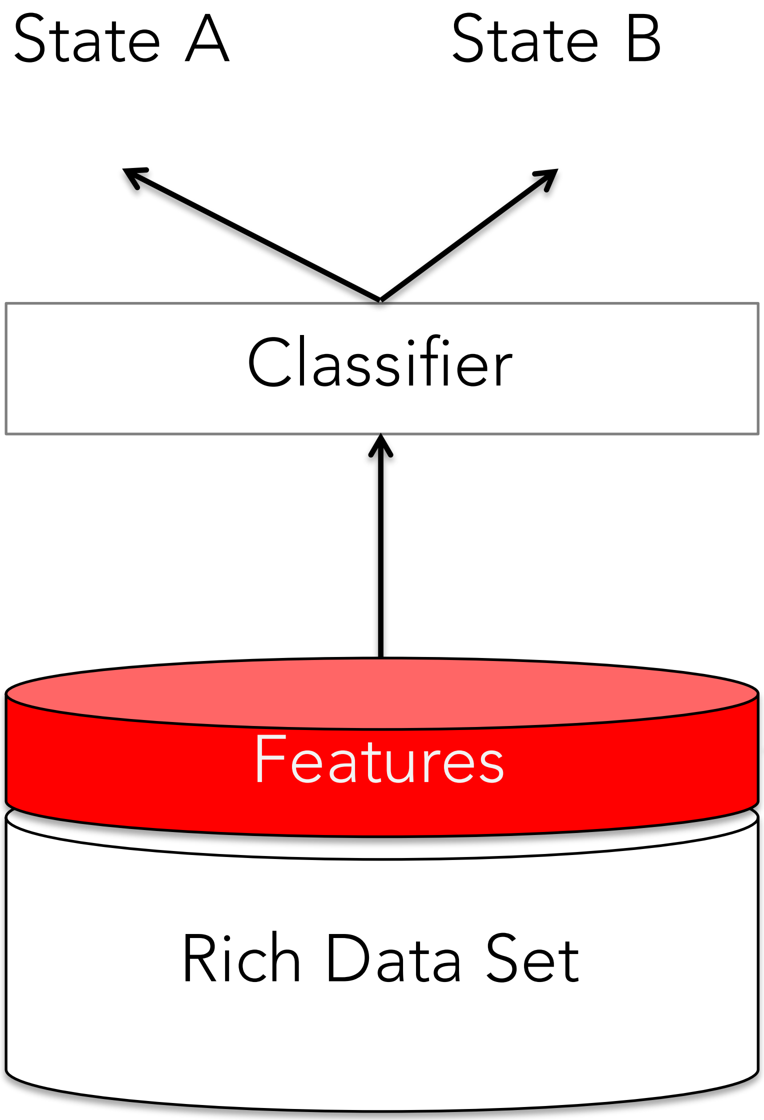
Learner Modelling @ DataScience Times



Π_2

- Balance of participation: Did all team members do a fair share of the workload?
- Task-distribution: Do team members perform specific subsets of the tasks?
- Rate of acknowledgement: What percentage of utterances from a learner received acknowledgement—from a simple nod to an acknowledging action
- Transactivity: Did team members build utterances upon the utterances produced by their peers?
- Cross-recurrence: Did team members look at the same object at (more or less) the same time?
- Rate of redundancy: Did the learner ask a question for which another team member already had the answer?

Learner Modelling @ DataScience Times



Π_3

- Conversation depth: The average depth of conversation threads in forums.
- Connectivity: What is the minimal number of students that need to be removed from the social network to disconnect the other nodes from each other (Diestel, 2005)?
- Homophily: Do students form ties with similar versus dissimilar students ? Ties can be forums postings; similarity is measured through students' profiles.
- Reciprocity: If student A often replies to another student B in the forum, is the opposite true?
- Propinquity: The tendency for actors to have more ties with those who are geographically close (Kadushin, 2012).
- Density: The proportion of direct interactions between two students relative to the total number of possible interactions between all students (Xu et al., 2010).
- …..

(4) Learner Modelling @ BayesianTimes

Activity a_5 . In order to reduce the variance of the set [1 2 2 3 3 3 4 5 8], 3 numbers can be removed. Which ones?

a) Remove all occurrences of number 3
b) Remove the numbers that appear several times
c) Remove 1, 5, and 8
d) Remove 4, 5, and 8

$X_5(S)$ = {**m**isunderstanding, **g**ood understanding)

if $b_5(s)=c$, then $x_5(s)=g$
if $b_5(s)=a$, then $x_5(s)=m$

$P\ (x_5(s)=g \mid b_5(s)=c) = 1$
$P\ (x_5(s)=g \mid b_5(s)=a) = 0$

(4) Learner Modelling @ BayesianTimes

Activity a_5 . In order to reduce the variance of the set [1 2 2 3 3 3 4 5 8], 3 numbers can be removed. Which ones?

a) Remove all occurrences of number 3
b) Remove the numbers that appear several times
c) Remove 1, 5, and 8
d) Remove 4, 5, and 8

$X_5(S)$ = {**m**isunderstanding, **g**ood understanding)

~~if $b_5(s)=c$, then $x_5(s)=g$~~
~~if $b_5(s)=a$, then $x_5(s)=m$~~

$P(x_5(s)=g \mid b_5(s)=c) = 75\%$ (he had 25% to succeed by chance)
$P(x_5(s)=g \mid b_5(s)=a) \approx 10\%$ (e.g. typing mistake)

(4) Learner Modelling @ BayesianTimes

$X_5(S)$ = {**m**isunderstanding, **g**ood understanding)

$P(x_5(s)=g \mid b_5(s)=c)$ = 75% (he had 25% to succeed by chance)

But if one knows a priori that this a difficult concept, e.g. that only 20% of students are usually in state « good understanding », one may apply Bayes Theorem

$$P(A|B) = \frac{P(A)\,P(B|A)}{P(B)}$$

$$P(x_5(s)=g \mid b_5(s)=c) = \frac{\overset{0.20}{P(x_5(s)=g)} \cdot \overset{\substack{\text{1-distraction} \\ 0.90}}{P(b_5(s)=c \mid x_5(s)=g)}}{\underset{0.90}{P(b_5(s)=c \mid x_5(s)=g)} \cdot \underset{0.20}{P(x_5(s)=g)} + \underset{\substack{0.25 \\ \text{randomness}}}{P(b_5(s)=c \mid x_5(s)\neq g)} \cdot \underset{0.80}{P(x_5(s)\neq g)}}$$

$P(x_5(s)=g \mid b_5(s)=c) = 0.47$

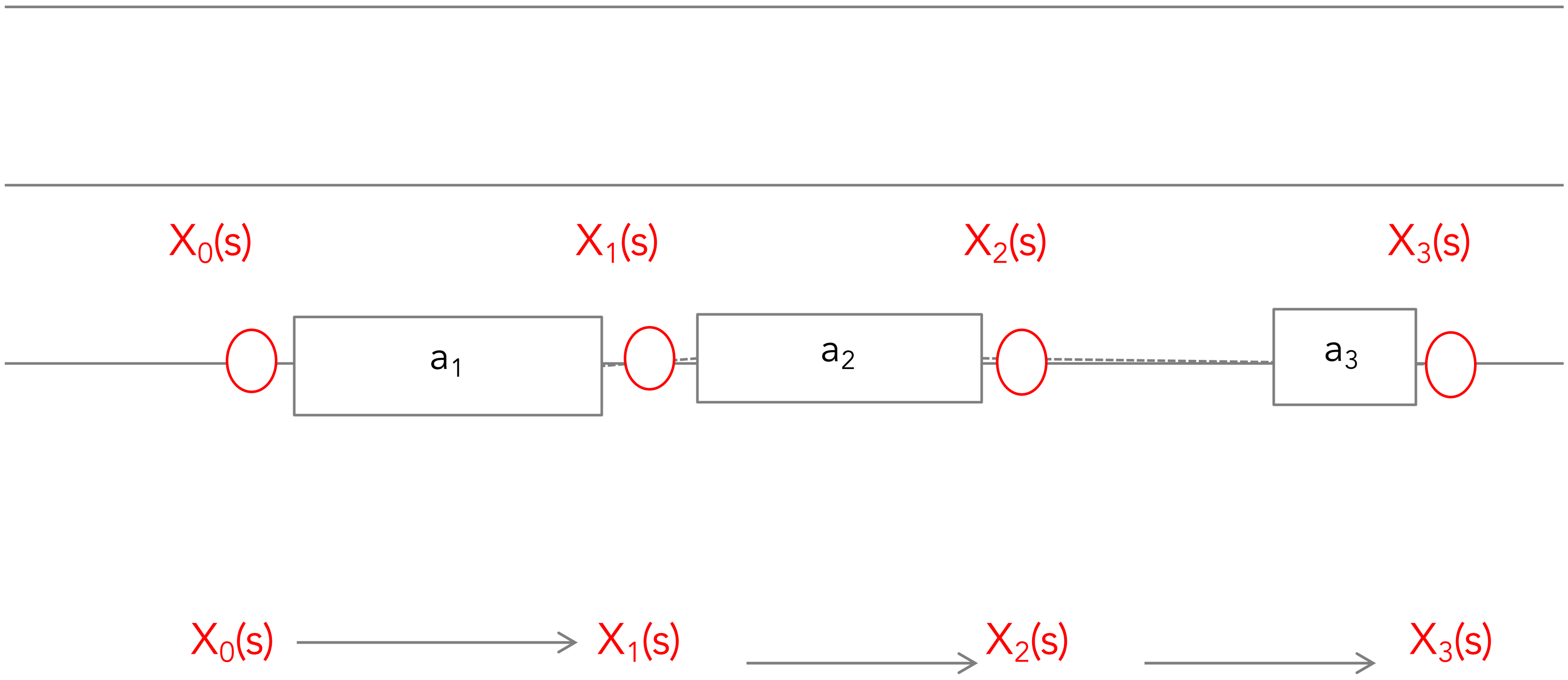
(4) Learner Modelling @ BayesianTimes

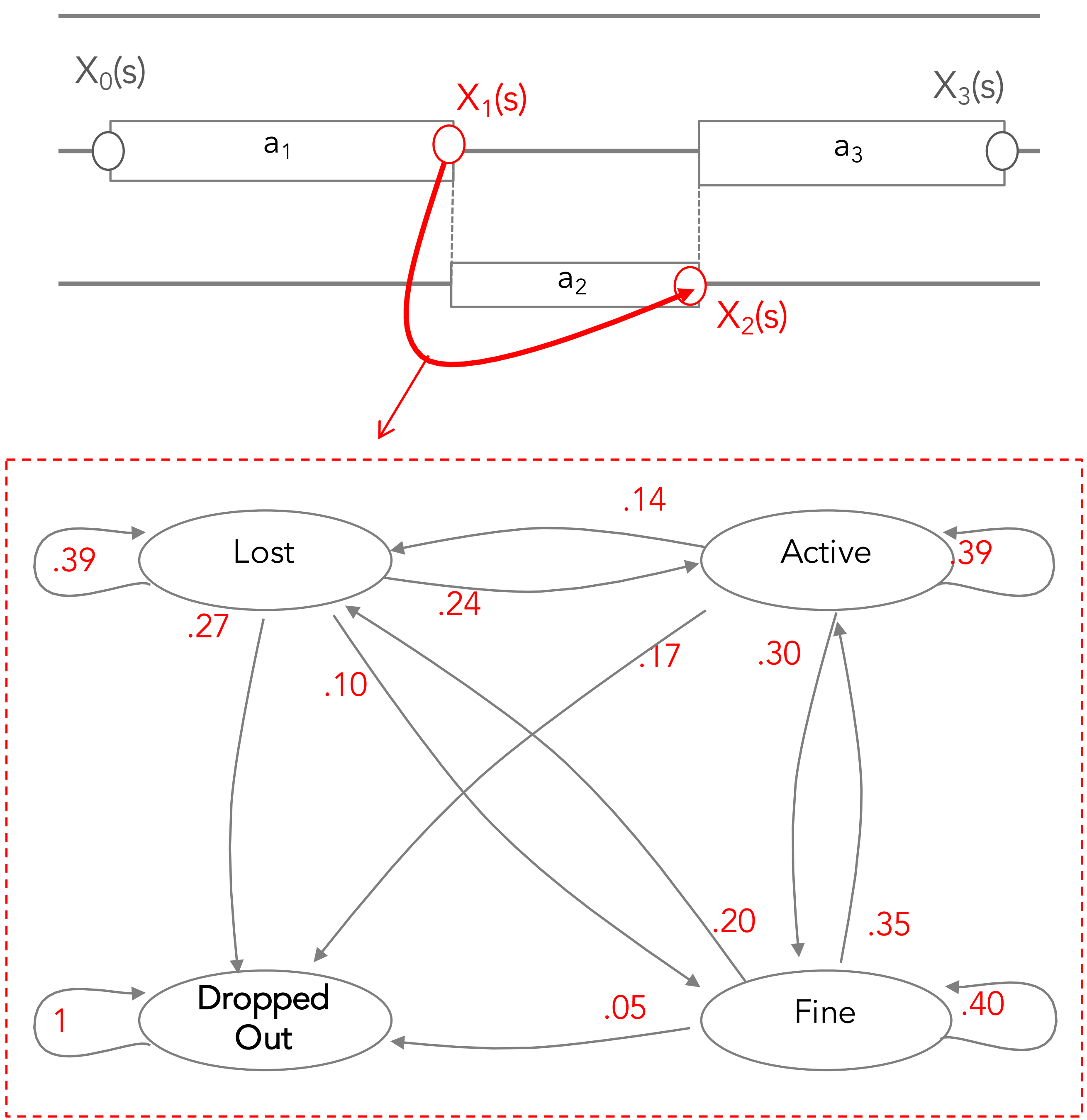
$$\left.\begin{aligned} P\,(x_5(s)=g \mid b_5(s)=c) &= 0.47 \\ P\,(x_5(s)=m \mid b_5(s)=c) &= 0.53 \end{aligned}\right\} \quad b_5(s) = c \rightarrow x_5(s) = [0.47\ 0.53]$$

The diagnosis power of this question is not great, close to 50/50. Entropy is very high !

(5) Learner Modelling @ MarkovTimes

Inferring the learner's state from his previous state





The weight of edges

State Transition Matrix

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

Activity 2

Activity 1

	Lost	Active	Fine	Great	**H**	**H0**
Lost	0.26	0.39	0.29	0.06	1.80	0.90
Active	0.19	0.34	0.26	0.21	1.96	0.98
Fine	0.11	0.28	0.45	0.16	1.81	0.90
Great	0.05	0.15	0.25	0.55	1.60	0.80
					1.H0 =	**0.10**

M6	Lost	Active	Fine	H	H0
Lost	0.01	0.24	0.75	0.87	0.55
Active	0.01	0.24	0.75	0.87	0.55
Fine	0.01	0.24	0.75	0.87	0.55
				ω(M5)	0.45

M7	Lost	Active	Fine	H	H0
Lost	0.75	0.24	0.01	0.87	0.55
Active	0.75	0.24	0.01	0.87	0.55
Fine	0.75	0.24	0.01	0.87	0.55
				ω(M6)	0.45

State Transition Matriy Utopy

M8				
0.2	0.2	0.2	0.2	0.2
0.2	0.2	0.2	0.2	0.2
0.2	0.2	0.2	0.2	0.2
0.2	0.2	0.2	0.2	0.2
0.2	0.2	0.2	0.2	0.2
			U(M)	**0**

M9				
1	0	0	0	0
0	1	0	0	0
0	0	1	0	0
0	0	0	1	0
0	0	0	0	1
			U(M)	**0**

M10				
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
0	0	0	0	1
			U(M)	**1**

M11				
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
1	0	0	0	0
			U(M)	**-1**

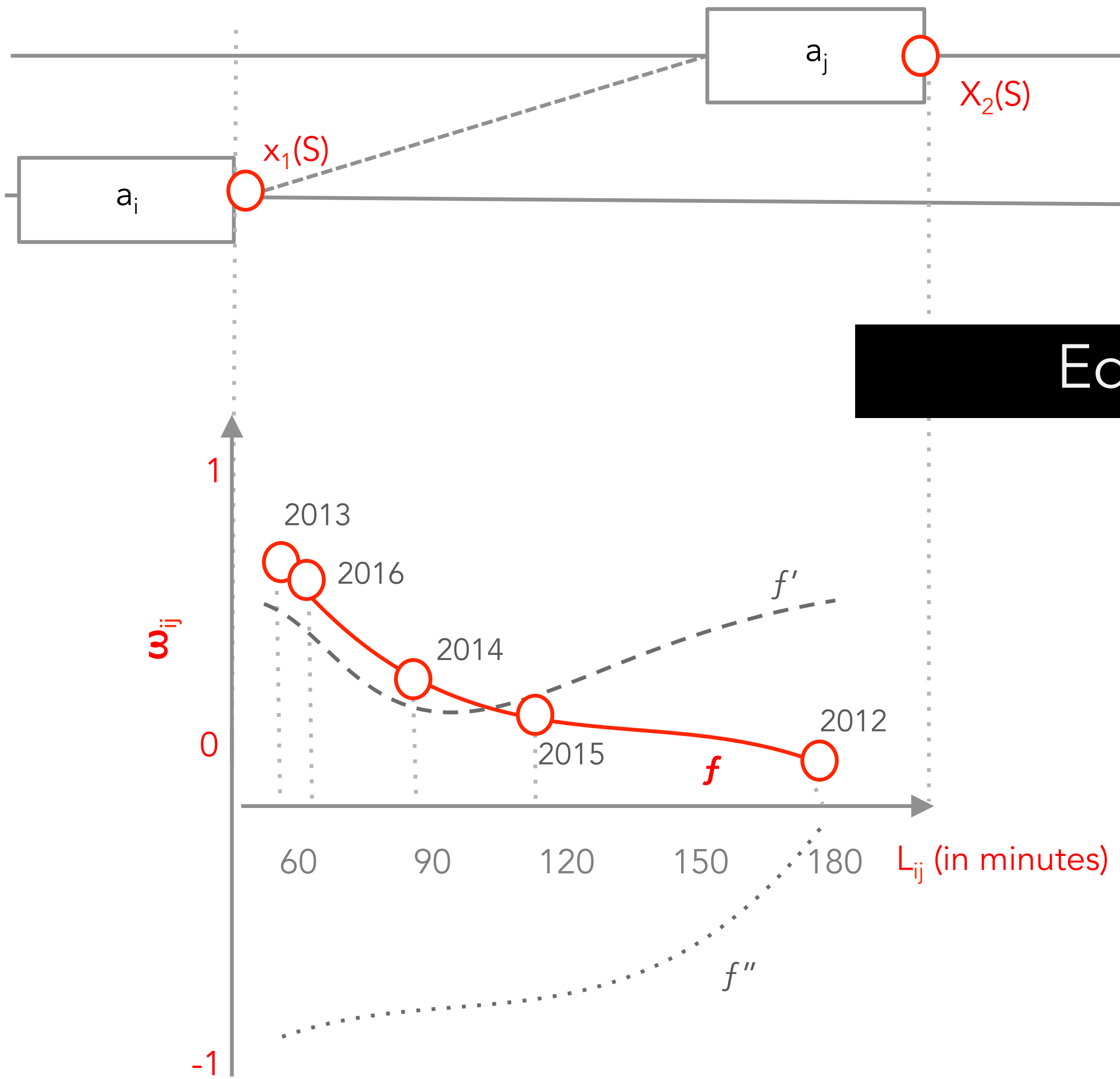
M12				
0.2	0.2	0.2	0.2	0.2
0.1	0.1	0.2	0.3	0.3
0	0	0.2	0.3	0.5
0	0.1	0.2	0.2	0.4
0	0	0	0.2	0.8
			U(M)	**0.47**

M13				
0.5	0.1	0.2	0.1	0.1
0.2	0.2	0.2	0.2	0.2
0.7	0.2	0.1	0	0
0.2	0.2	0.2	0.2	0.2
0.8	0.2	0	0	0
			U(M)	**-0.42**

$$\gamma(M) = \frac{2}{m(m-1)} \sum_{k=1,l>k}^{m-1} (l-k)m_{kl} - \sum_{k=2,l<k}^{m} (k-l)m_{kl}$$
$$= \frac{2}{m(m-1)} \sum_{k=1}^{m}\sum_{l=1}^{m} (l-k)m_{kl}$$

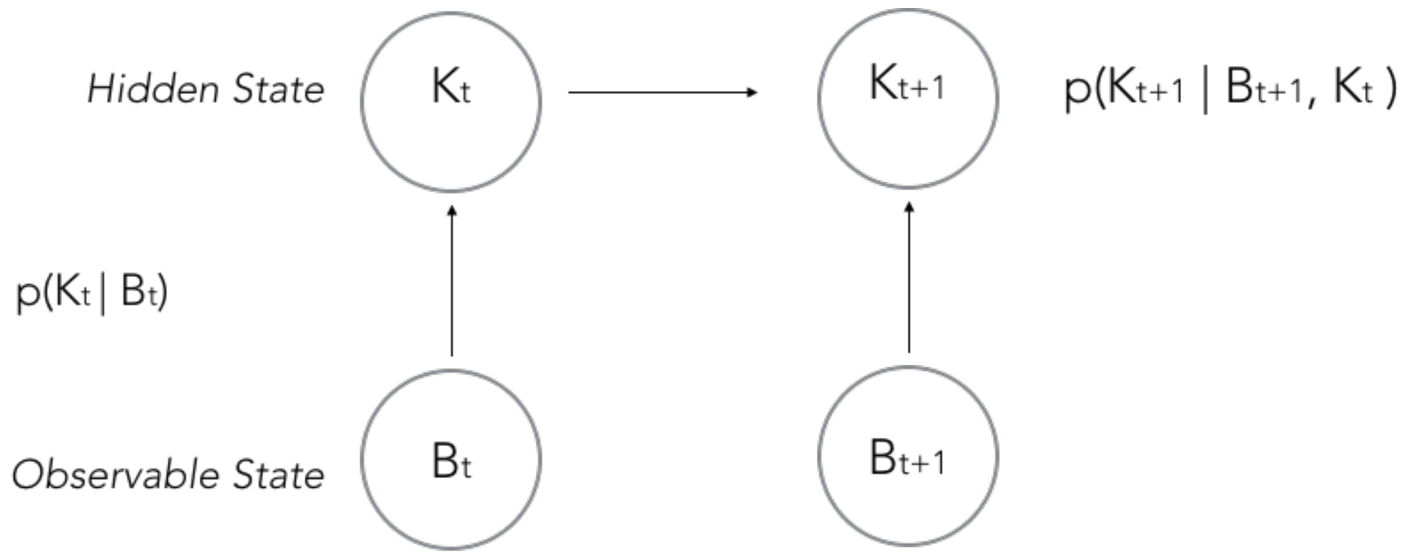
For the exam, I don’t ask you to learn home-made formulas but to understand the principles. Formulas could be replaced by visalisations

Edge Eslasticity



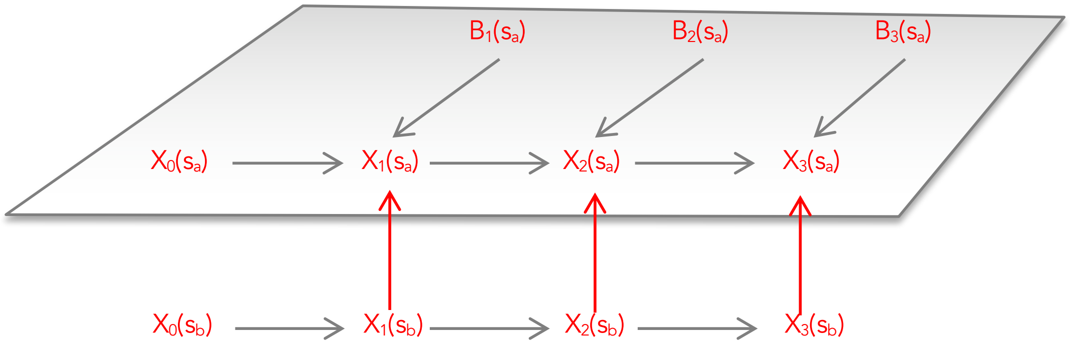
The strength relationship between activies willoften degrade with time,
e.g. even if a_1 is strong pre-requisite to a_2,
the knowledge gained in a_1 won't remain activated for ever.

So far we treated them separately, but one may infer the learner's state from both his behaviour and his previous state

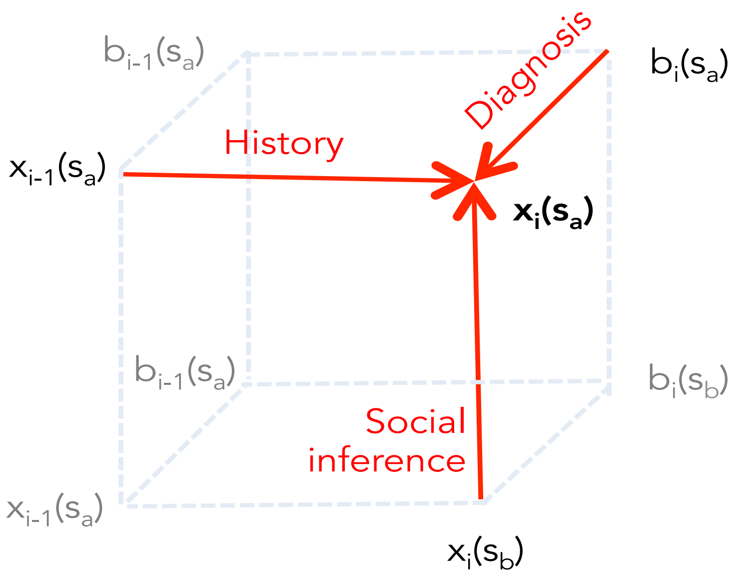


Bayesian Knowledge Tracing

One step further: one may infer the learner's state from his behaviour (depth), his previous state (horizontally) and the state of others (vertically)



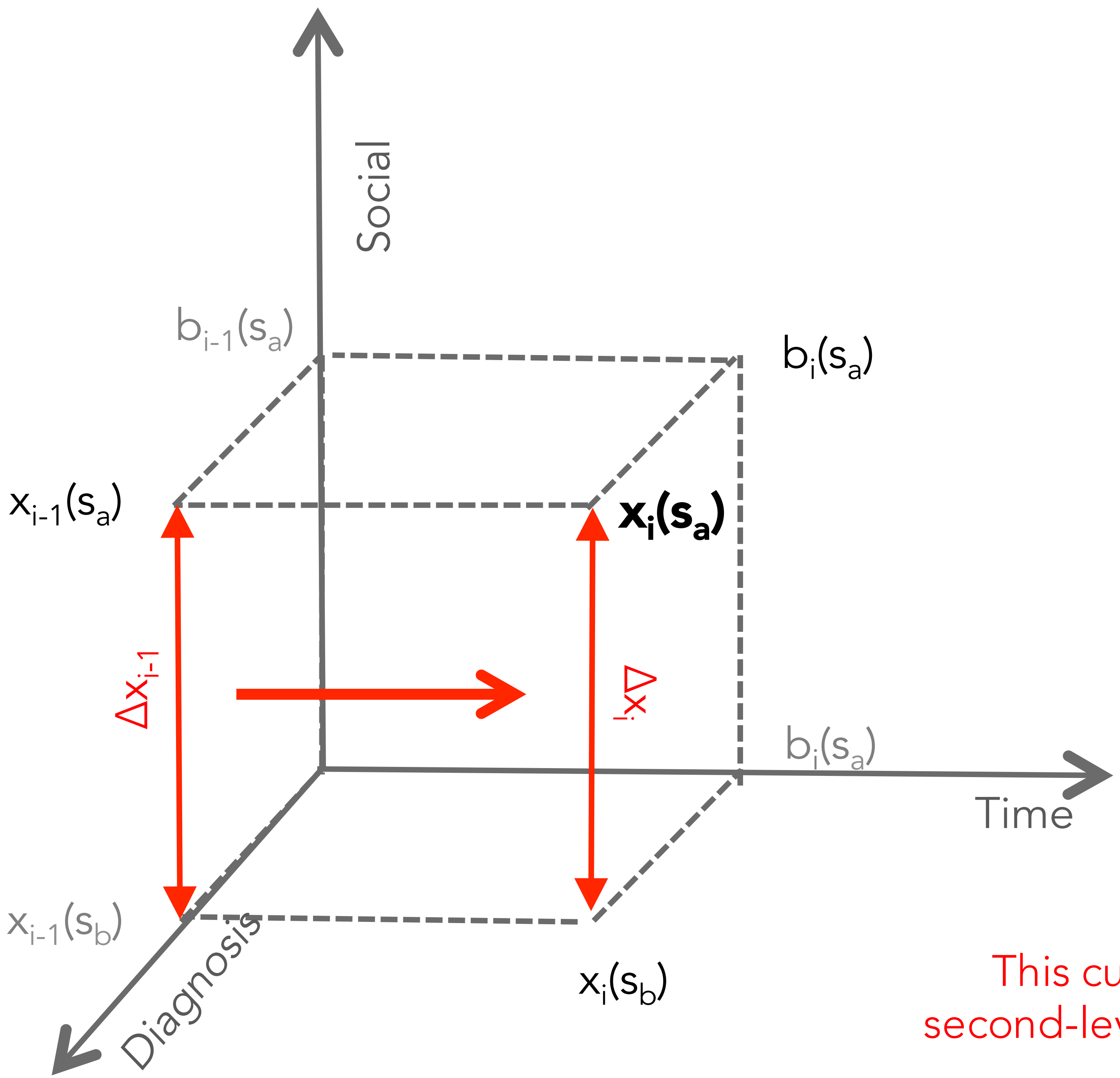
The learning analytics cube:
3 axes of inference



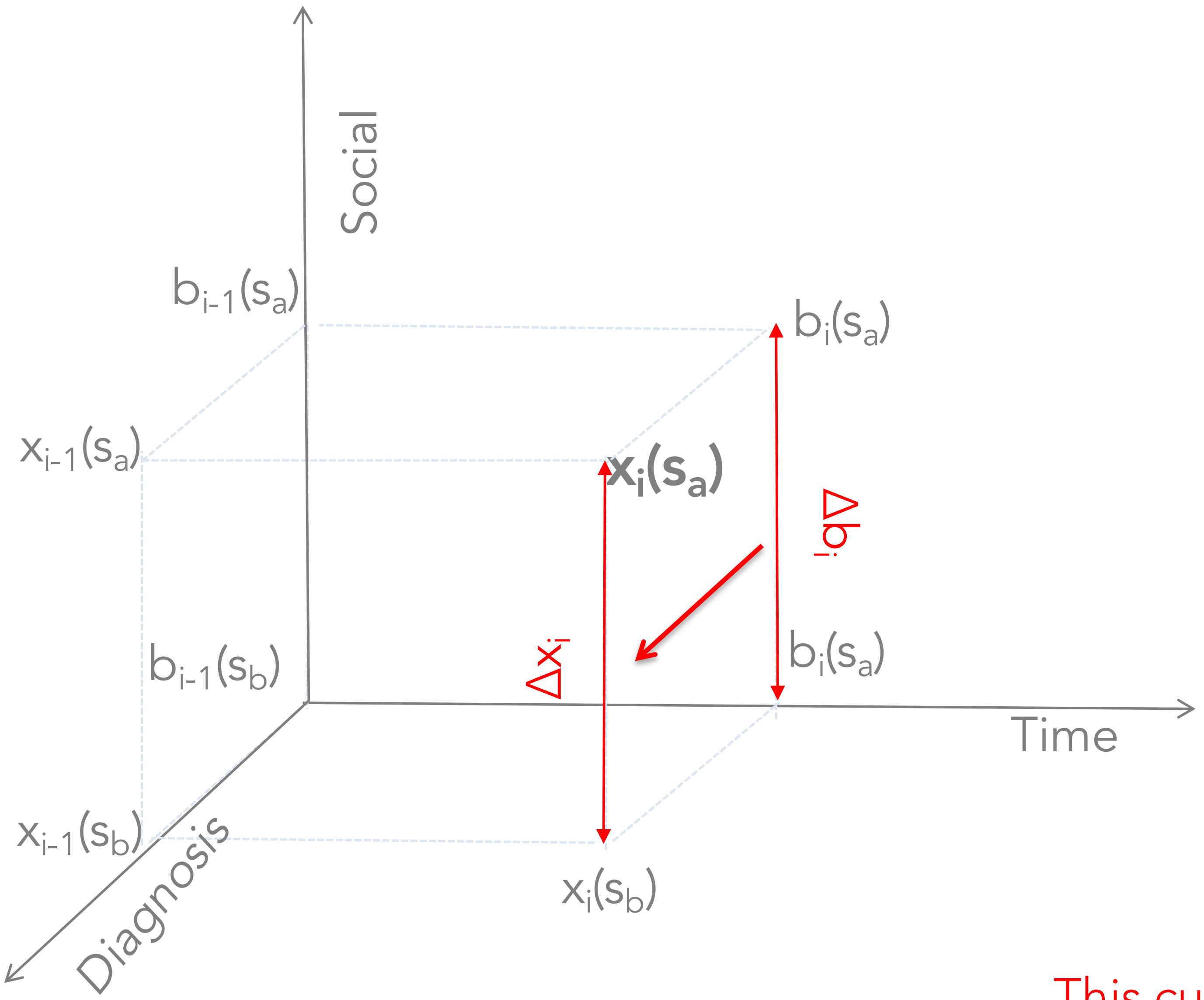
A. John does probably not understand SD deviation because he removed the central values of the distribution

B. John does probably not understand SD deviation because he did not understand what is a mean and the mean is a pre-requisite

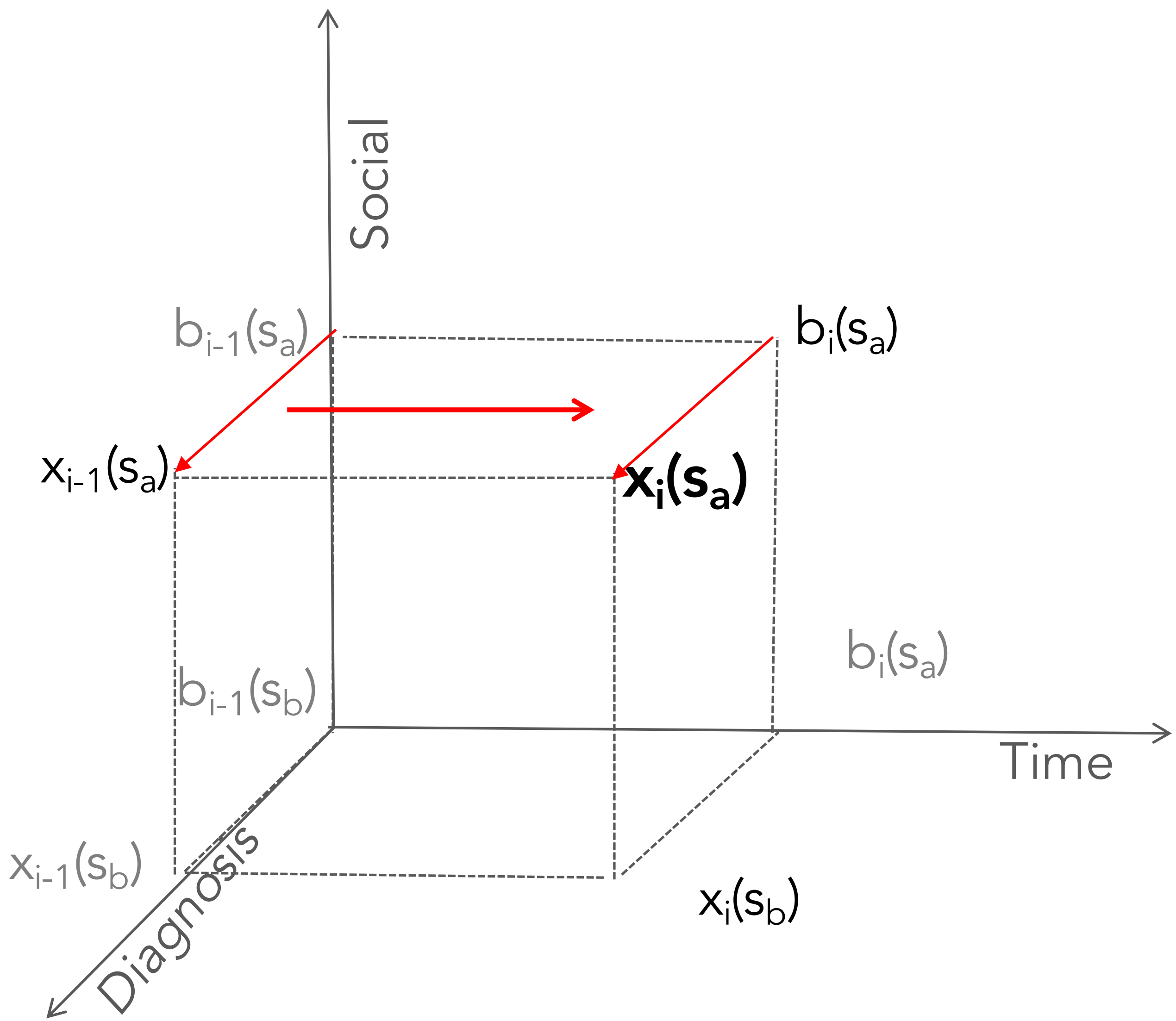
C. John does probably not understand SD deviation because most learners in that class failed and John is one of the weakest



This cube may allow
second-level inferences



This cube may allow
second-level inferences



This cube may allow
second-level inferences

So far we use common sense to describe the learner state

$$x_i(s) \in X_i(S) = \{\text{fine, active, lost, drop}\}$$

but educational research defined is much richer set of states

Measured
at time t

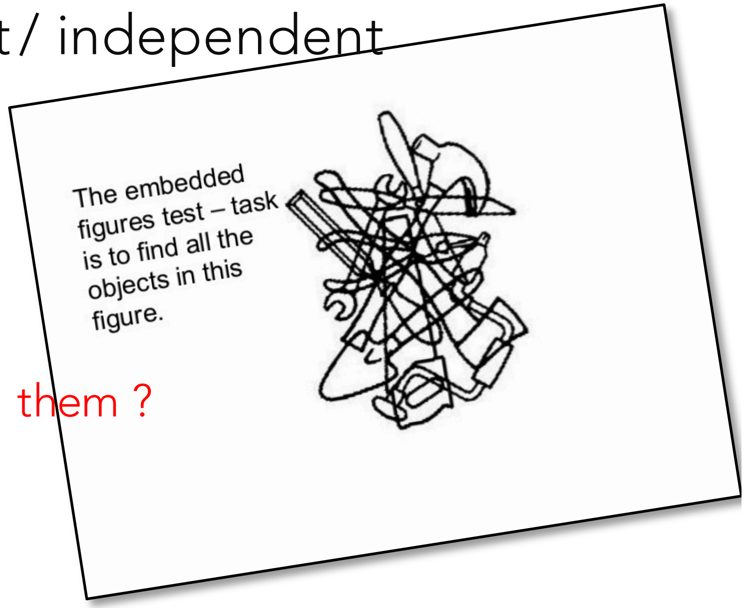
Stable
in time

State ≠ Trait

Learning styles

Cognitive styles

- Anxious / Self-confident
- Risk-aversive / Risk-seeking
- Aural / visual / kinesthetic
- Deep / Surface
- Field-dependent / independent



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
- Non-verbal learning disability, NVLD
- ...
- High-potential chidren
- ....

Labels help Sales

Learning Analytics