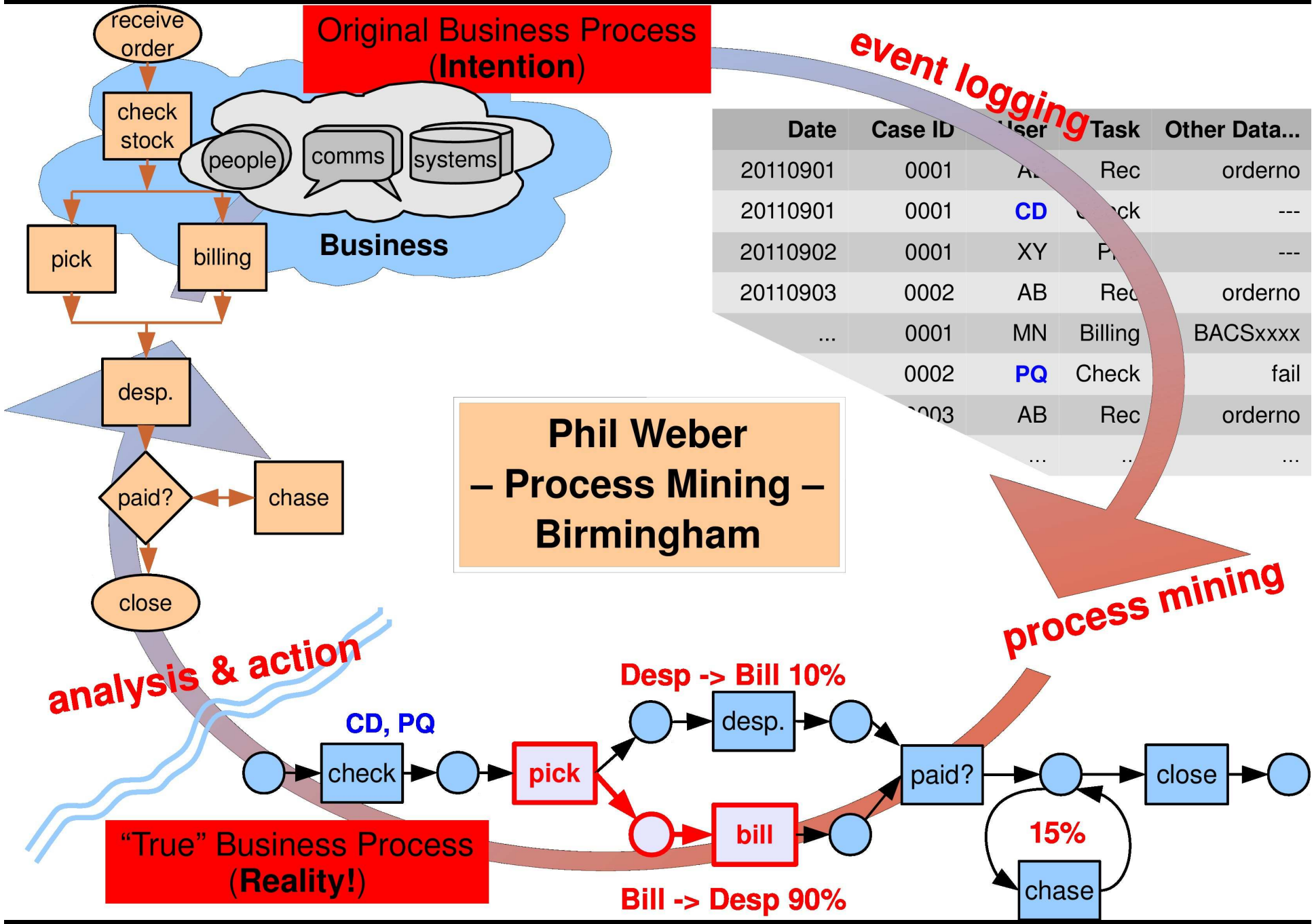
Real-Time Detection of Process Change Using Process Mining

Phil Weber
Behzad Bordbar, Peter Tiňo

University of Birmingham

EPSRC
Pioneering research and skills



Date	Case ID	User	Task	Other Data...
20110901	0001	A…	Rec	orderno
20110901	0001	CD	C…ck	---
20110902	0001	XY	P…	---
20110903	0002	AB	Rec	orderno
...	0001	MN	Billing	BACSxxxx
	0002	PQ	Check	fail
	…003	AB	Rec	orderno
	…	…	…	…

Changing Process

- Expected?
- Problem?

How to detect?

Pre-requisites

1. Notion of Real-time
2. Measures of Models
3. Detect Change
4. Statistical Significance

Notion of Real-time

Predictability
Results + timeframe

Accuracy vs Time

Accuracy:

$$Pr\left[d(P_M, P_{M_K}) > \epsilon\right] < \delta,$$

Mining Time:

$$t\left(\phi(L_k)\right) = A + k \cdot c < \eta.$$

Upper bound time
Lower bound data

→

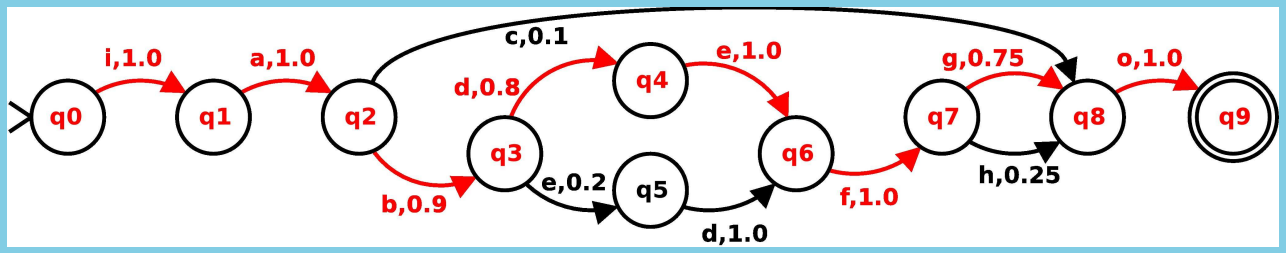
Requirement
Realism

Pre-requisites

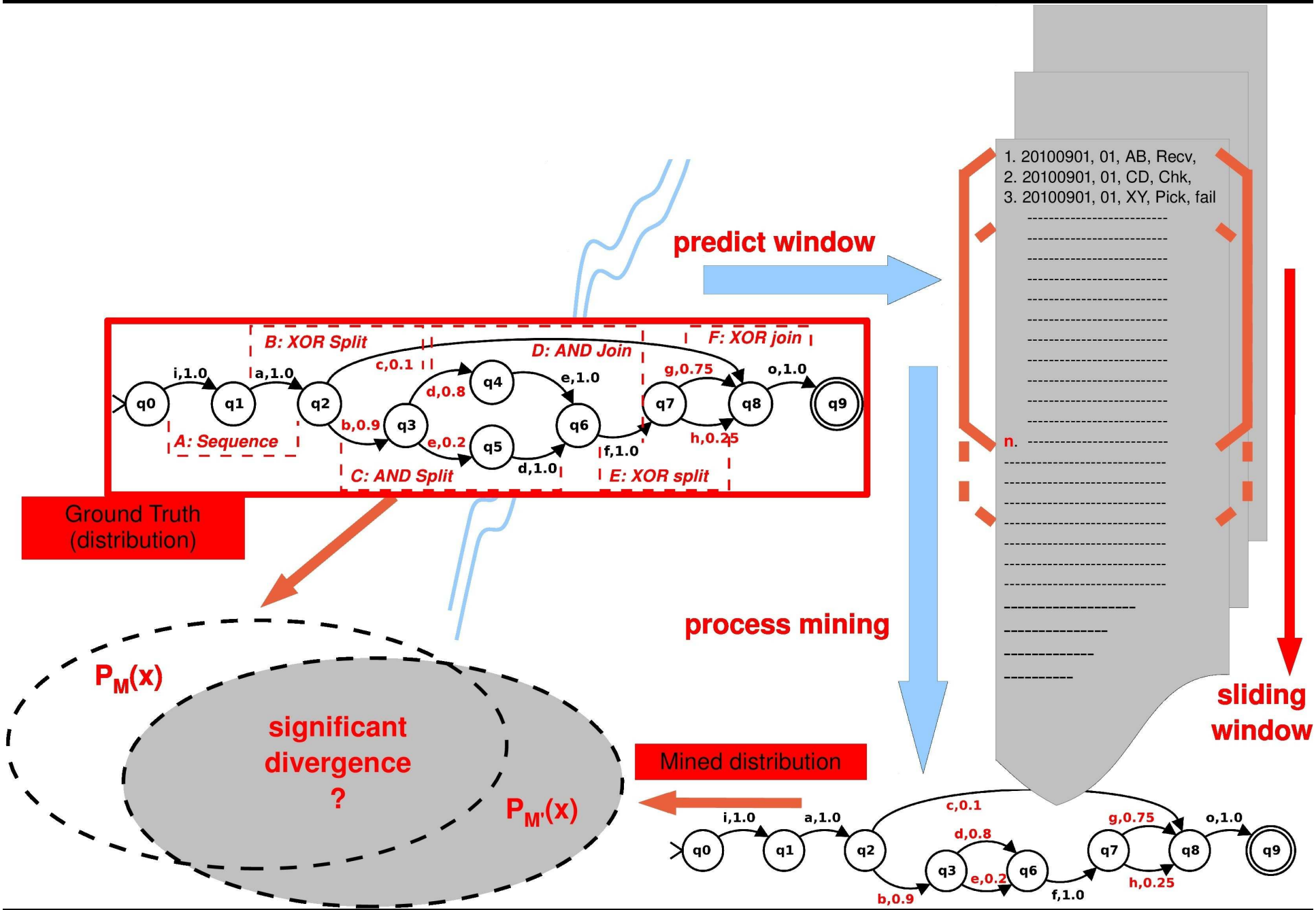
1. Notion of Real-time
2. Measures of Models
3. Detect Change
4. Statistical Significance

Compare Models & Detect Change

Probabilistic Representation



- Activities - symbols
- Traces - strings
- Distribution over strings
- PDFA


predict window
1. 20100901, 01, AB, Recv,
2. 20100901, 01, CD, Chk,
3. 20100901, 01, XY, Pick, fail
n.
B: XOR Split
D: AND Join
F: XOR join
A: Sequence
C: AND Split
E: XOR split
q0
q1
q2
q3
q4
q5
q6
q7
q8
q9
i,1.0
a,1.0
c,0.1
d,0.8
e,1.0
g,0.75
o,1.0
b,0.9
e,0.2
f,1.0
h,0.25
d,1.0
Ground Truth
(distribution)
process mining
sliding
window
$P_M(x)$
significant
divergence
?
$P_{M'}(x)$
Mined distribution

Pre-requisites

1. Notion of Real-time
2. Measures of Models
3. Detect Change
4. Statistical Significance

Statistical Significance

Sample from Distribution - Chi^2

String/Arc Probability

- Hoeffding
- Hypothesis

jpg

Statistical Tests

Chi^2 Test:

$$\chi_s^2 = \sum_{i=1}^{i=k} \frac{(n(x_i) - n \cdot p(x_i))^2}{n \cdot p(x_i)}, \text{ and}$$

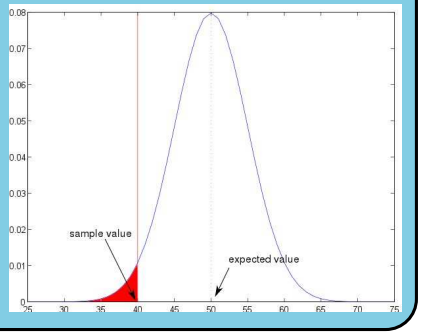
$$p = Pr(\chi_{k-1}^2 \geq \chi_s) = \int_{\chi_s}^{\infty} f(\chi_{k-1}^2) d(\chi_{k-1}^2).$$

Hoeffding Bound:

$$\forall a \in \Sigma, \left| \delta(q_1, a, q_1') - \frac{f(q_2, a)}{n(q_2)} \right| < \sqrt{\frac{1}{2n(q_2)} \ln\left(\frac{2}{\alpha}\right)}.$$

jpg

Hypothesis Test



Experimentation

Design & Predict

Simulate - Ground Truth

Mine "Sliding Window"

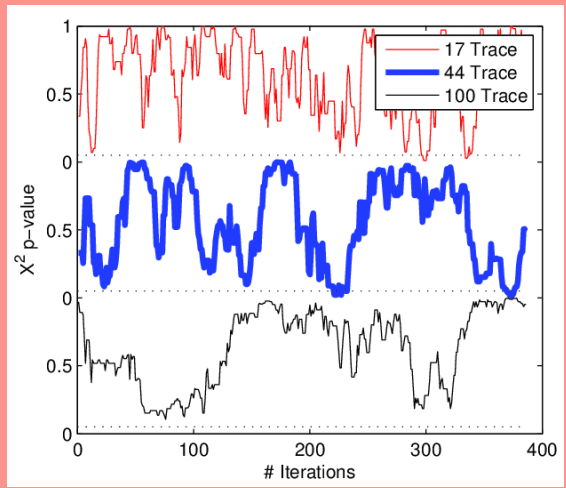
Compare Distributions

Statistical Tests

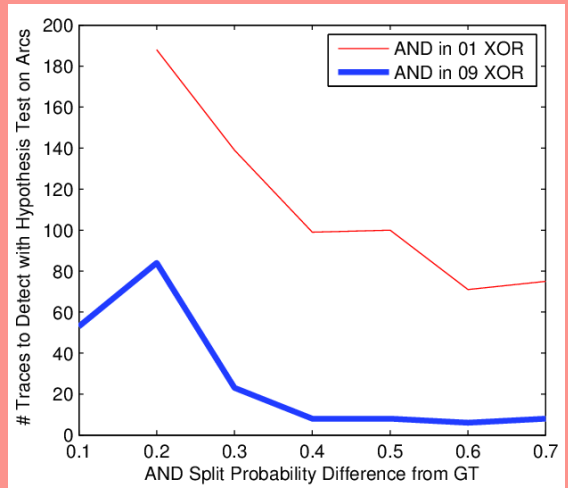
- vary probabilities *in* structures
- vary probabilities *of* structures
- vary how much data

Results

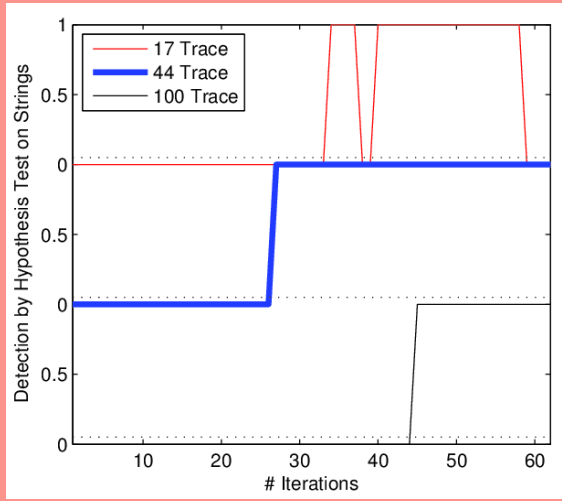
Chi^2 p-value - Unchanging Process



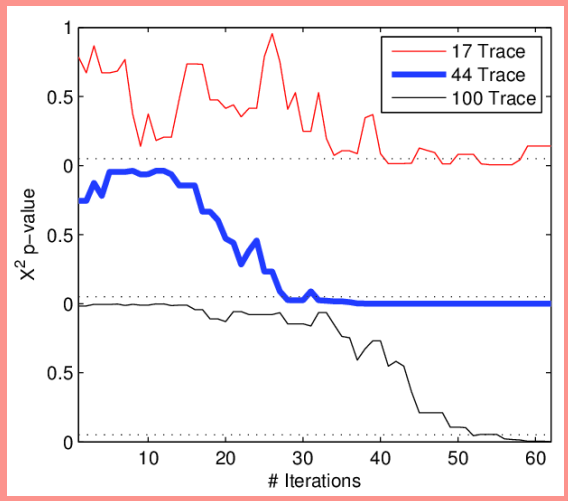
AND Split - Vary Probabilities



XOR Change - Hypothesis Testing



XOR Change - Chi^2



Questions & Thoughts?

Phil Weber

http://www.cs.bham.ac.uk/~pxw869

p.weber@cs.bham.ac.uk