Risk assessment
Experts vs. Laypersons

Tor Stålhane – IDI / NTNU

Timo Malm – VTT / Tampere

Why did we run the experiments

We want developers to perform risk assessment during change impact analysis. Is this OK?

We need to know: Do experts

- Find the “correct” risk level more often than laypersons?
- Agree more than laypersons => do their scores have a lower variance?

What did we do – 1

The experiment consisted of assessing the risk of 18 cases – nine from moving machinery and nine from a robot tool cell.

Two series of experiments

- First series: expert assessors from Norway, Sweden and Finland
  - Robots: 17 experts
  - Moving machinery: 19 experts
- Second series:
  - Robots: 18 third year NTNU Software Engineering students
  - Moving machinery: 20 third year NTNU Software Engineering students

What did we do – 2

Each participant

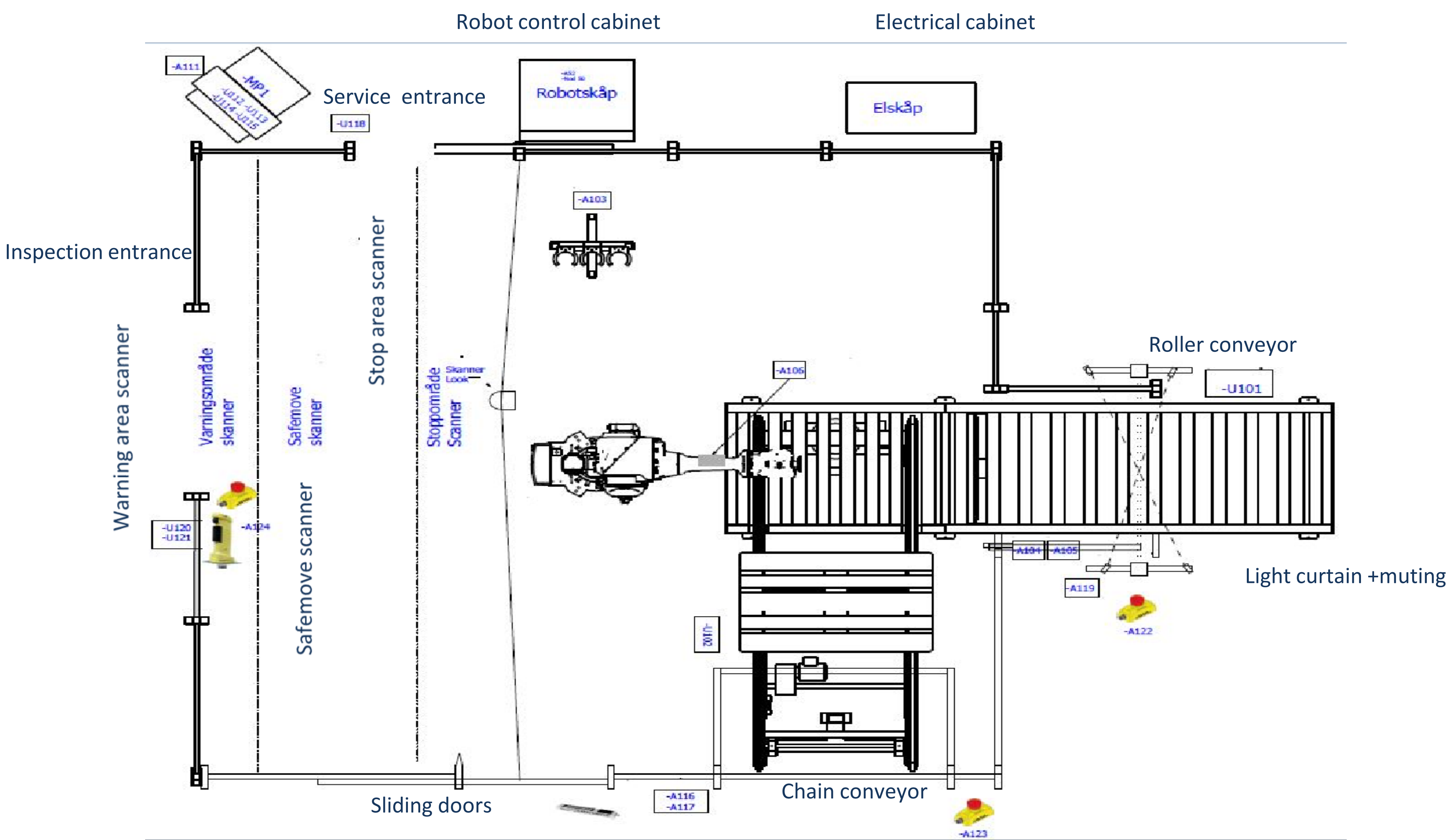
1. Got a copy of the experiment Excel sheets
2. Filled in all 18 cases using the methods defined by
   a) ISO 13849
   b) IEC 62061
3. Returned the Excel sheets via e-mail to
   a) VTT (experts)
   b) IDI (students)

Case 1 – Moving machinery

Tractor Loader

- Backhoe Traveling <40 km/h
- Unexpected brake apply.
- Machine stops very abruptly, and may skid.
- Steering remains functional, but is limited.
- Bystander may be
  - crushed between machine and hard surface.
  - run over.

Case 2 – Robot tool cell – 1

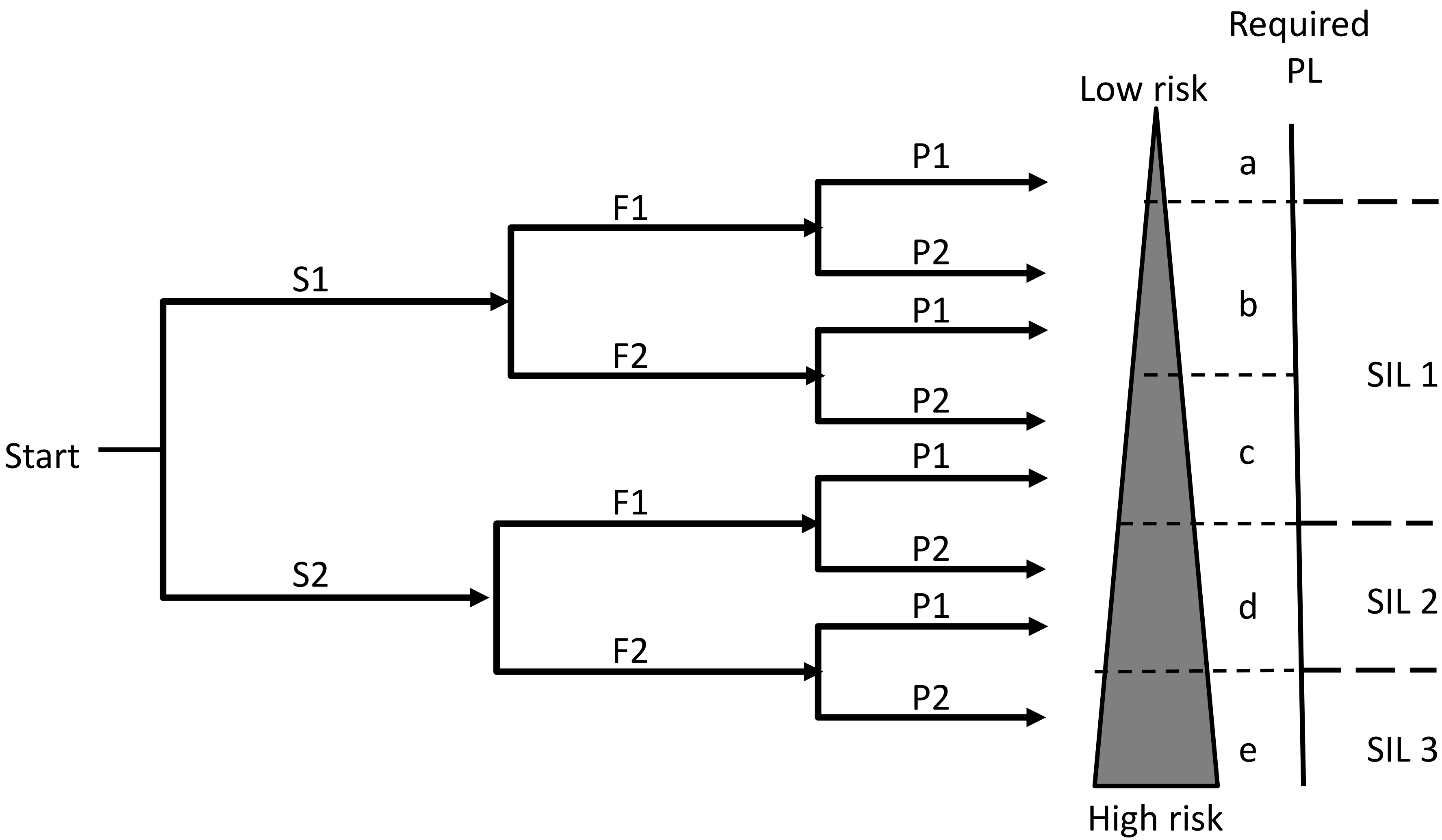


Case 2 – Robot tool cell – 2

Hazard: Moving elements

- Hazardous event: Robot or machine moves in unpredictable way or speed.
- Harm: Impact/ punch/ crushing
- Foreseeable sequence of events:
  - Unintentional impact on operating devices.
  - Workers unintentionally impact operating device, e.g., changing speed or range of robot or starting chain conveyor.
- Hazardous situation: The system stands near a passage/entrance in a factory. Many people pass by. Both visitors and different workers.

ISO 13849



Required PL
Low risk
Start
S1
S2
F1
F2
F1
F2
P1
P2
P1
P2
P1
P2
P1
P2
a
b
c
d
e
SIL 1
SIL 2
SIL 3
High risk

IEC 62061

Frequency and duration Fr		Probability of hzd. event, Pr		Avoidance Av	
<= 1 hour	5	Very high	5		
>1hour - <= day	5	Likely	4		
>1 day - <= 2 weeks	4	Possible	3	Impossible	5
>2 weeks - <= 1 year	3	Rarely	2	Possible	3
>1 year	2	Negligible	1	Likely	1

Consequences	Severity Se	Class Cl = Fr + Pr + Av				
		3 – 4	5 – 7	8 - 10	11 - 13	14 - 15
Death, losing an eye or an arm	4	SIL 2	SIL 2	SIL 2	SIL 3	SIL 3
Permanent, losing fingers	3		QM	SIL 1	SIL 2	SIL 3
Reversible, medical attention	2			QM	SIL 1	SIL 2
Reversible, first aid	1				QM	SIL 1

Human choices

Two important tendencies for human choices

- When assessing consequences of event =>
  Choose the worst case consequences
- When selecting from a Likert scale or a table of alternatives =>
  End-avoidance - it is safer to select something in the middle.

What should we expect

Frequency and duration Fr		Probability of hzd. event, Pr		Avoidance Av	
<= 1 hour	5	Very high	5		
>1hour - <= day	5	Likely	4		
>1 day - <= 2 weeks	4	Possible	3	Impossible	5
>2 weeks - <= 1 year	3	Rarely	2	Possible	3
>1 year	2	Negligible	1	Likely	1

Consequences	Severity Se	Class Cl = Fr + Pr + Av				
		3 – 4	5 – 7	8 - 10	11 - 13	14 - 15
Death, losing an eye or an arm	4	SIL 2	SIL 2	SIL 2	SIL 3	SIL 3
Permanent, losing fingers	3		QM	SIL 1	SIL 2	SIL 3
Reversible, medical attention	2			QM	SIL 1	SIL 2
Reversible, first aid	1				QM	SIL 1

What did we see – 1

ALL - EXPERT

Se	**3 - 4**	**5 - 7**	**8 - 10**	**11 - 13**	**14 - 15**	**Sum**
4	4	59	100	39	3	205
3		27	37	8		72
2		6	14	2		22
1		4	3	2	1	10
Sum	4	96	154	51	4	309

ALL - STUDENT

Se	**3 - 4**	**5 - 7**	**8 - 10**	**11 - 13**	**14 - 15**	**Sum**
4	6	39	43	10		98
3	6	26	50	13		95
2		15	28	20		63
1	1	13	19	16		49
Sum	13	93	140	59		305

What did we see – 2

Machinery – Experts

Case 3, SIL1 / Case 5, SIL3

Se	3 - 4	5 - 7	8 - 10	11 - 13	14 - 15
4	1 / 1	5 / 7	3 / 7	1 / 2	
3		5 / 0	2 / 2	1 / 0	
2					
1		1 / 0			
Sum	1 / 1	11 / 7	5 / 9	2 / 2	0 / 0

Case 1 – Moving machinery – 1

PL

e	0	0	1	0	**2**	1	2	6	3
d	9	6	4	12	9	**10**	2	**5**	2
c	8	**12**	11	**4**	7	7	**2**	5	**7**
b	1	0	**1**	2	1	0	4	1	1
a	0	1	2	1	0	1	7	1	5
0	**1**	0	0	0	0	0	1	1	0
std	-	c	b	c	e	d	c	d	c
Case	1	2	3	4	5	6	7	8	9

Experts

e	1	0	1	1	**2**	1	0	1	0
d	6	2	1	4	15	**7**	2	**6**	3
c	10	**11**	10	**8**	3	8	**3**	13	**16**
b	1	3	**3**	4	0	1	7	0	1
a	2	4	5	3	0	3	8	0	0
0	**0**	0	0	0	0	0	0	0	0
Std.	0	c	b	c	e	d	c	d	c
Case	1	2	3	4	5	6	7	8	9

Students

The number of answers to the mobile work machine cases according to the ISO 13849 method.

Case 2 – Robot Tool Cell – 1

PL									
e	1	1	1	0	**2**	5	0	0	3
d	3	12	7	9	3	**5**	10	4	**8**
c	2	**4**	9	**6**	3	3	**7**	1	5
b	0	0	0	0	6	3	0	1	0
a	1	0	**0**	0	3	1	0	**0**	1
0	0	0	0	0	0	0	0	0	0
std		c	a	c	e	d	c		d
Case	1	2	3	4	5	6	7	8	9

Experts

PL

e	0	1	0	0	1	0	2	0	2
d	0	2	1	5	4	3	7	1	8
c	3	3	9	6	10	2	1	2	5
b	0	6	4	4	2	9	1	0	2
a	0	6	4	3	1	4	7	0	1
0	3	0	0	0	0	0	0	3	0
std	0	c	a	c	e	d	c	a	d
Case	1	2	3	4	5	6	7	8	9

Students

The number of answers of the nine robot cases according to the ISO 13849 method.

Experts vs. Students; Paired t-test

Case 1

Paired t-test:
exp - 13849; stud - 13849
t-test of mean
difference = 0 vs. not = 0:
t-Value = -1,14 P-Value = 0,286

Paired t-test :
exp - 62061; stud - 62061
t-test of mean
difference = 0 vs. not = 0:
t-Value = -0,94 P-Value = 0,375

Case 2

Paired t-test :
exp - 143849; stud - 13849
t-test of mean
difference = 0 vs. not = 0:
t-Value = 0,35 P-Value = 0,737

Paired t-test :
exp - 62061; stud - 62061
t-test of mean
difference = 0 vs. not = 0:
t-Value = 1,35 P-Value = 0,214

No difference between laypersons and experts

Experts vs. Students: F-test

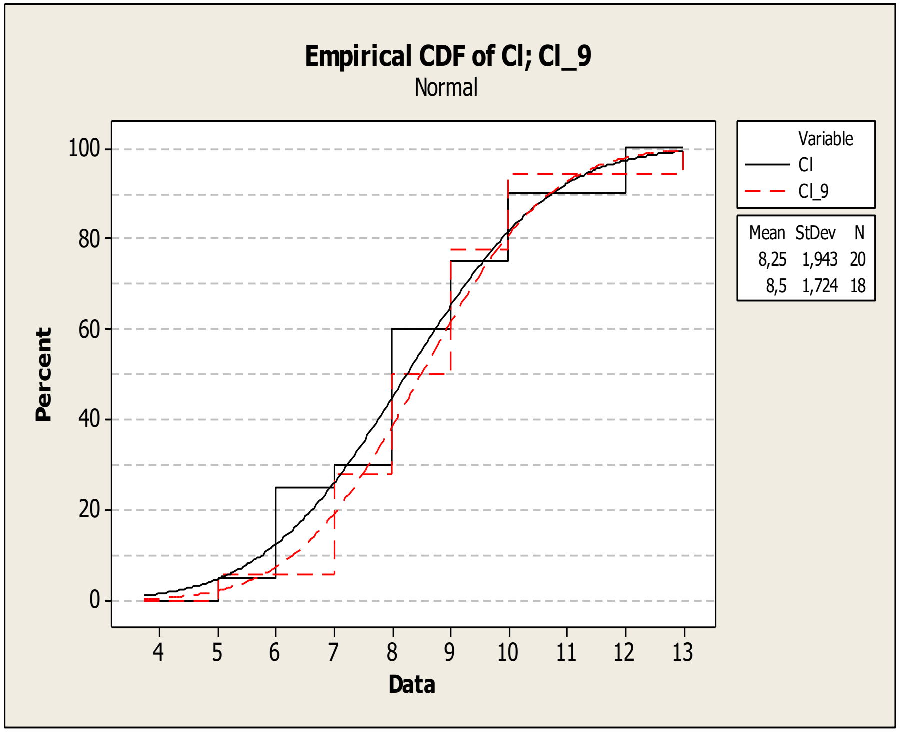
Case 1 – Variances

Case	P-value	Var 1	Var 2
1	0.16	0.51	0.26
2	0.97	0.80	0.79
3	0.49	1.01	0.73
4	0.26	0.73	0.43
5	0.00	0.22	0.99
6	0.84	0.70	0.77
7	0.46	0.82	0.58
8	0.80	1.33	1.19
9	0.23	1.39	0.79

Case 2 – Variances

Case	P-value	Var 1	Var 2
1	-	-	-
2	0.01	0.18	0.69
3	0.68	0.60	0.74
4	0.89	0.88	0.82
5	0.68	0.93	0.76
6	0.09	1.43	0.61
7	0.02	0.25	0.85
8	-	-	-
9	0.75	0.69	0.81

Parameter distributions



Distributions are compared using the Smirnov-Kolmogorov test.

Example below:

Critical value: 0.35

Cl	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7	Case 8	Case 9
$D_{13,13}$	0.21	0.15	0.22	0.23	0.17	0.24	0.22	0.34	0.12

Two alternatives for each parameter ISO 13849

How is the standard used?

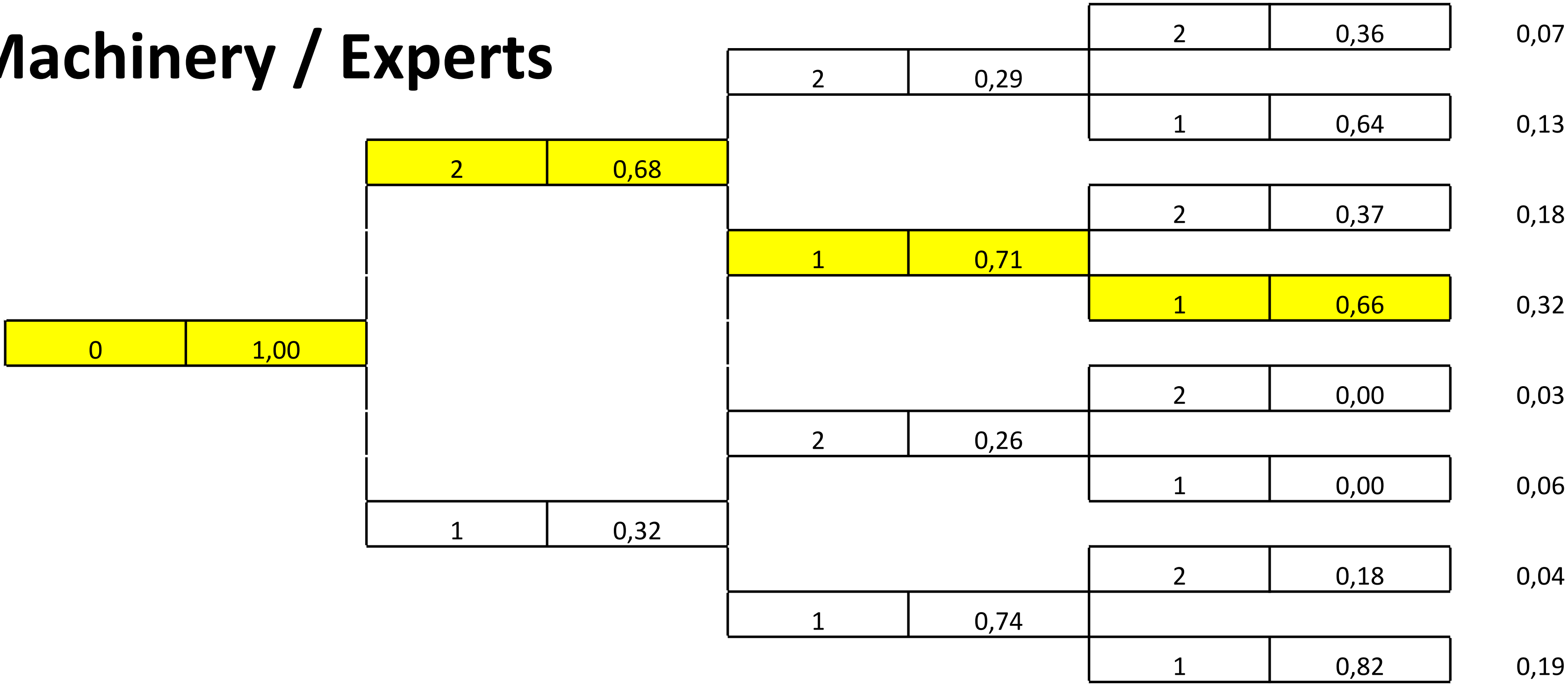
We will look at the results for

- All participants – experts and laypersons
- Moving machinery – experts and laypersons separated
- Two cases using only expert results. The two cases should give different results - PL e and PL b

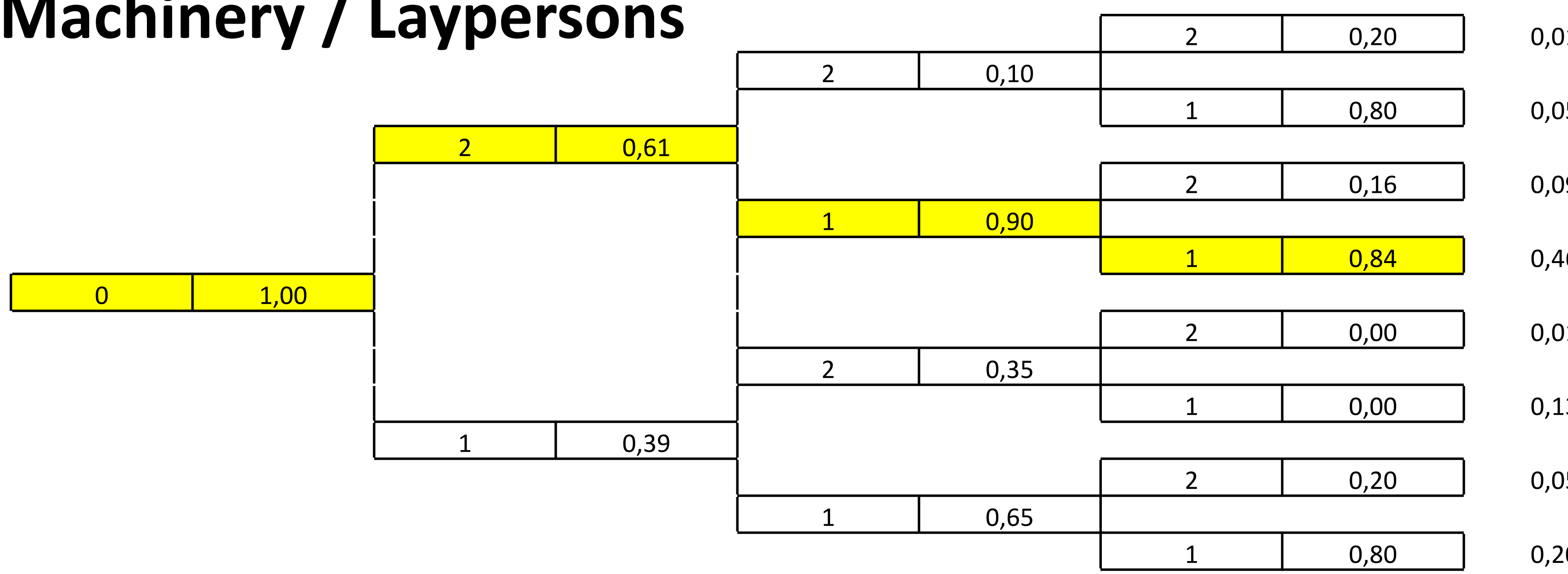
All participants – ISO 13849

Experiment	222	221	212	211	122	121	112	111	Sum
Robot - expert	4	12	27	22				1	65
Robot - student	5	6	15	12	12		22		72
Machinery – expert	5	9	13	22	2	4	3	14	72
Machinery – student	1	4	7	37	1	10	4	16	80
Sum	15	31	62	93	15	14	29	31	290
	201				89				290

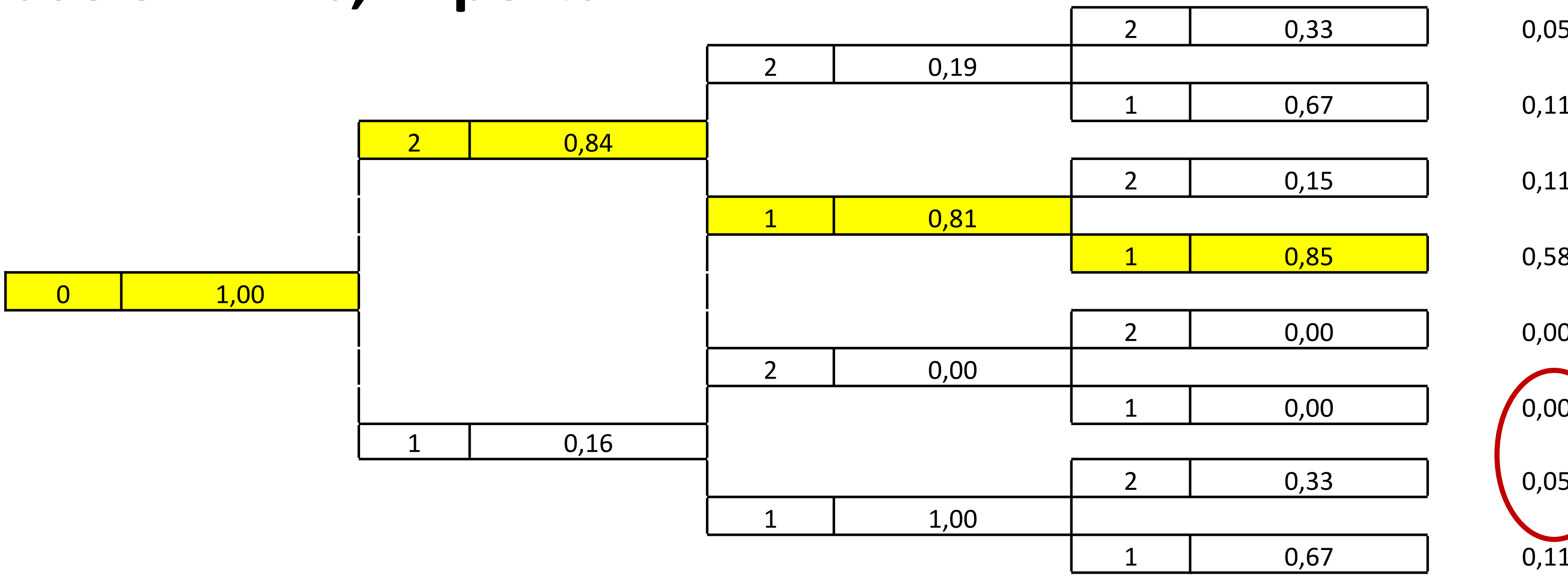
Machinery / Experts



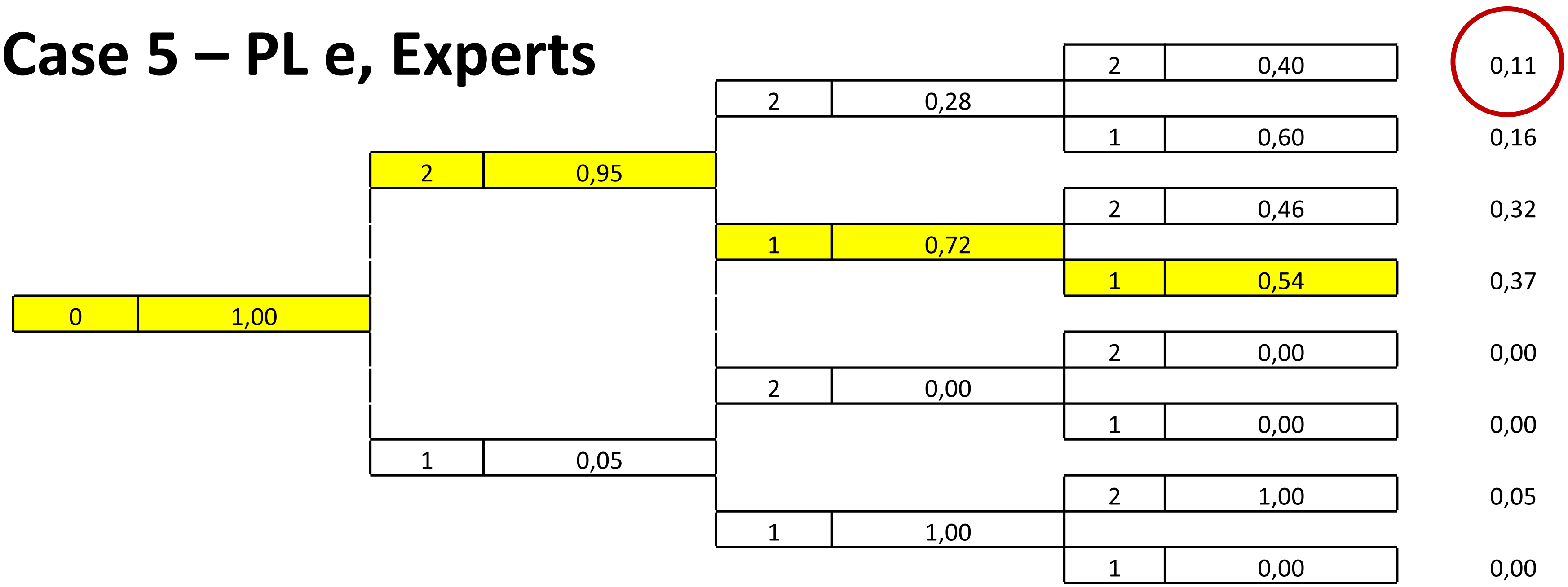
Machinery / Laypersons



Case 3 – PL b, Experts



Case 5 – PL e, Experts



A tentative explanation – 1

Human characteristics:

- Allow personal bias to affect judgments
- The more complex the conditions, the more assessors tend to ‘compensate’
- Subconscious use of other criteria than those specified, or apply their own weightings of existing criteria

A tentative explanation – 2

Based on the compensation idea, the participants – experts and students – reason as follows:

1. We assume the worst case => S = 2 (69%)
2. Compensation
   a) But it won't happen so often => F = 1 (77%)
   b) And it is mostly possible to avoid => P = 1 (60%) but sometimes P = 2 (40%)

Diagnostic vs. Informative factors

We can split all available information into two groups

- Diagnostic factors – few but important
- Informative factors – many and with varying quality.

Too many informative factors gives bad decisions because

- The extra information waters down the diagnostic factors
- People intuitively feel that they must use all available information

How to do a risk assessment

We need a description of

- The system and how it can fail – a few general failure modes will do
- The environment that the system will operate in – what can go wrong
- How the system interacts with its environment - how can the system's failures cause harm

We do ***NOT*** need

- Detailed knowledge about the inner working of the system
- A large amount of detailed
  - failure modes
  - environment descriptions

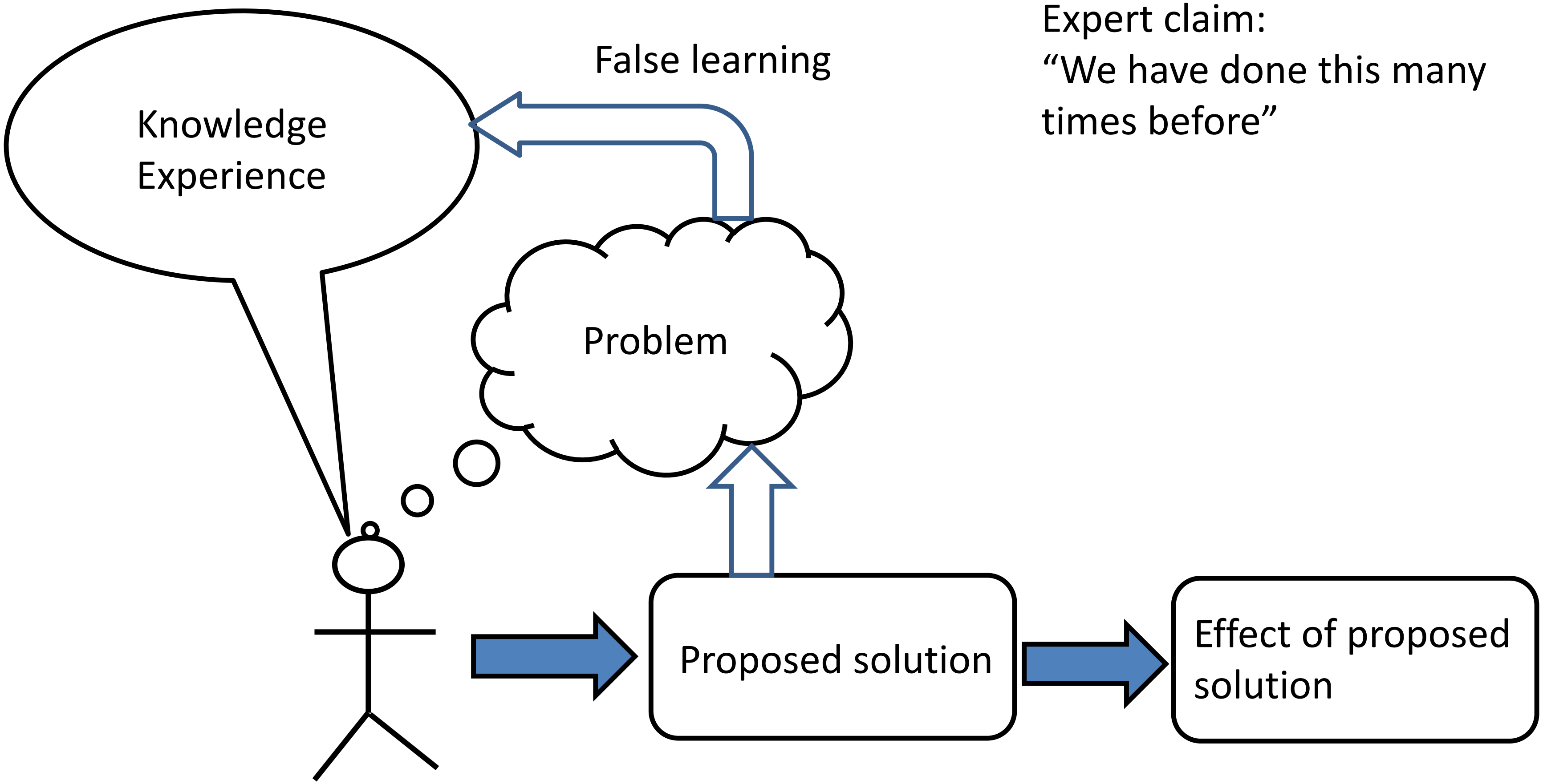
Main conclusion

Safety assessment experts and laypersons are first and foremost human.

The same heuristics, biases and weaknesses apply to both categories.

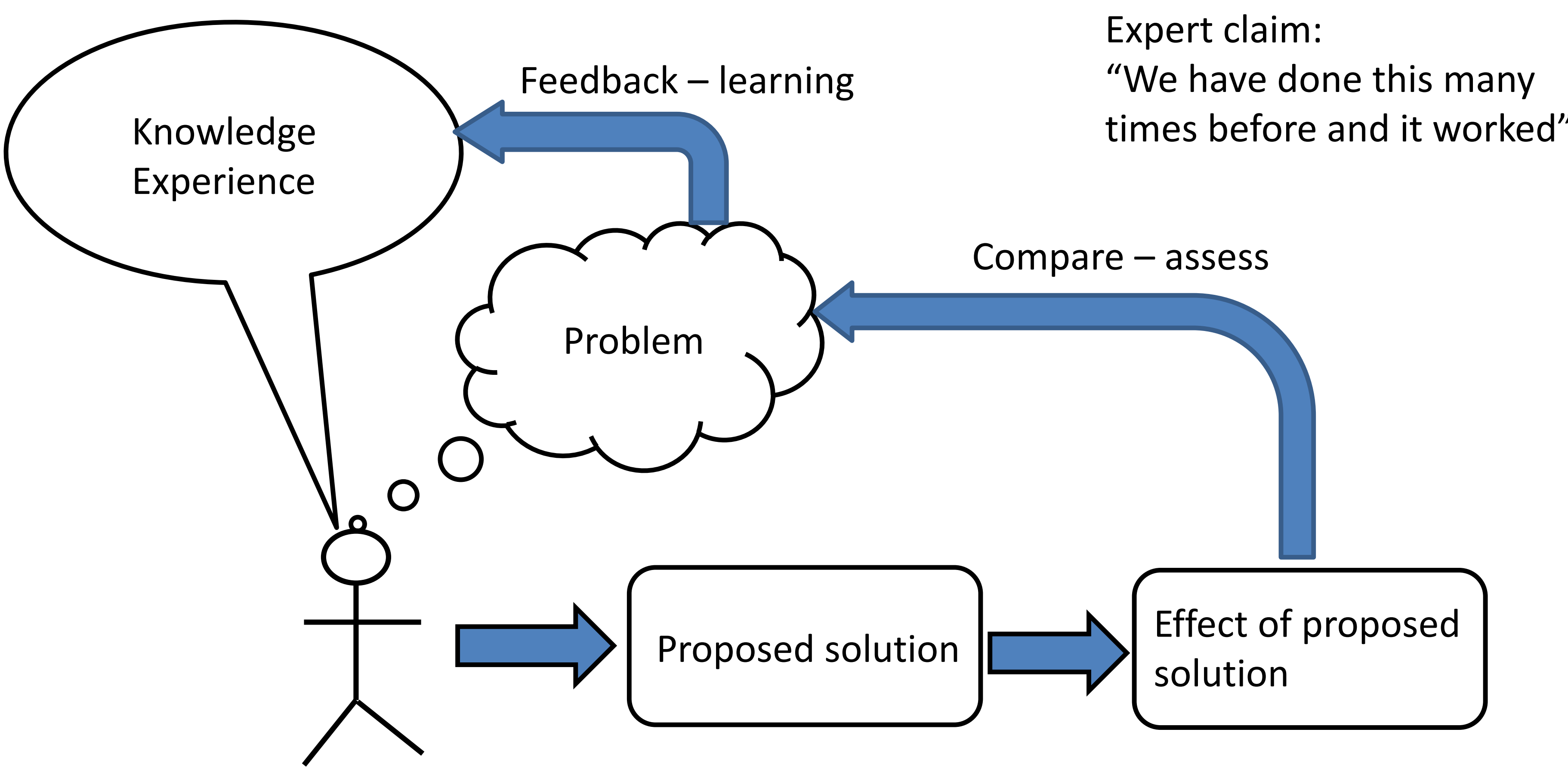
- The number of correct assessments are the same for experts and laypersons
- The parameter distributions are the same for laypersons and experts
- The variances are the same for experts and laypersons for all but three cases

No expert



Expert claim:
"We have done this many times before"

Real expert


Expert claim:
“We have done this many times before and it worked”
Feedback – learning
Knowledge
Experience
Compare – assess
Problem
Proposed solution
Effect of proposed solution

Where do we go from here – 1

Data problem

Most (all) data on hazardous situations that are collected stem from systems that have already been analysed and protected according to a safety standard.

What we need to get a better risk assessment is data related to

- Near misses
- How often the protection part of the system has been activated

Where do we go from here – 2

During assessment, the assessor should document the rational for each value.

This will have two effects:

- The rational will be available to others to discuss, agree or disagree.
- Writing the rational might help the assessors to overcome the basic, psychological reactions

The Delphi method might be a good solution