University of Nottingham
UK | CHINA | MALAYSIA

Exploring how Component Factors and their Uncertainty Affect Judgements of Risk in Cyber-Security

Zack Ellerby

Josie McCulloch

Melanie Wilson

Christian Wagner

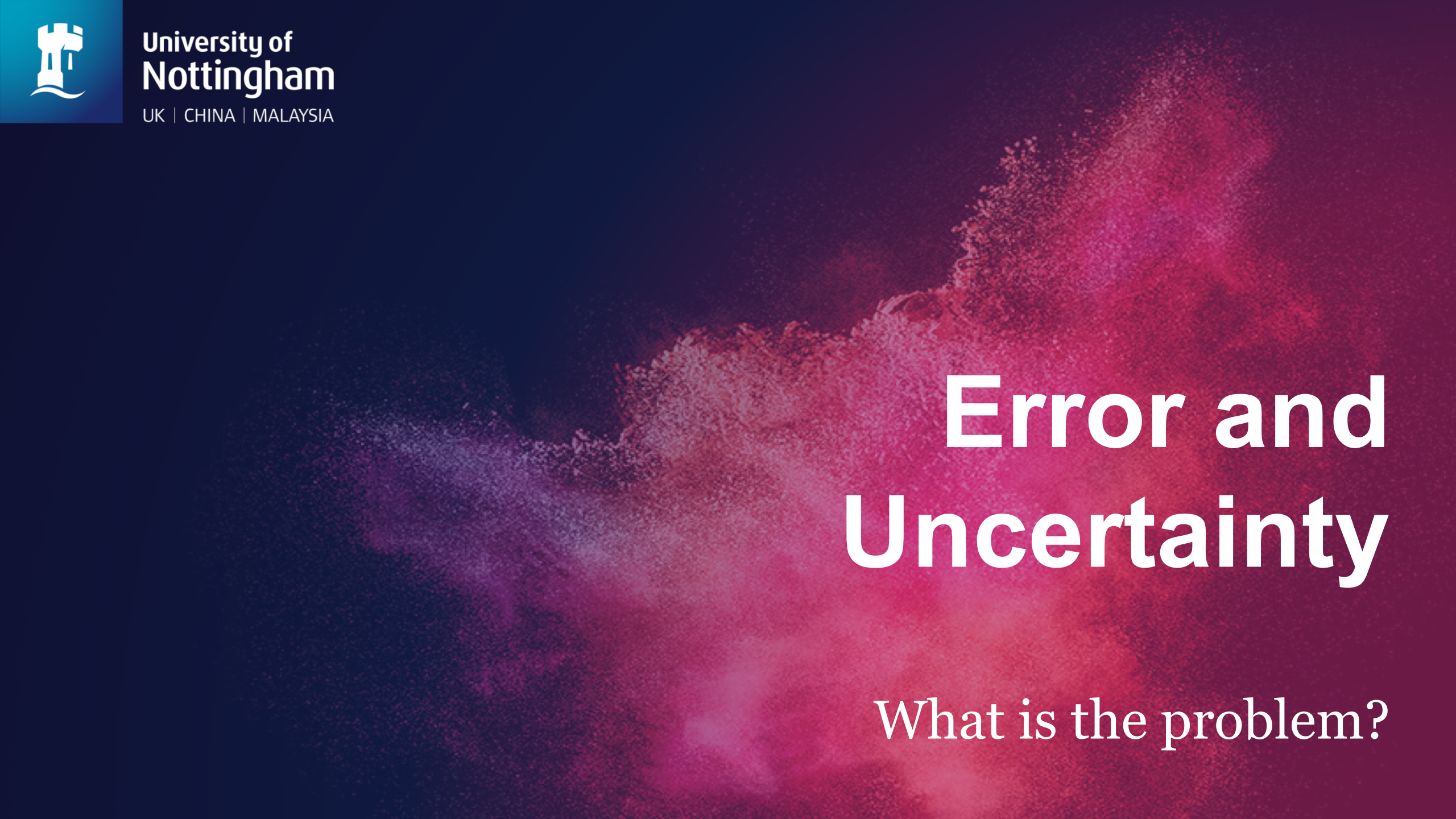
LUCID

Lab for Uncertainty in Data and Decision Making

University of Nottingham
UK | CHINA | MALAYSIA

Who are we?

- We are part of the Lab for Uncertainty In Data and Decision Making (LUCID) research group working at the University of Nottingham (UK).
- Within the School of Computer Science, but different backgrounds.
- This work was part of a collaborative project with Carnegie Mellon University and the UK's National Cyber Security Centre (NCSC).
- Our aim is to develop methods for effective capture and modelling of uncertainty in ratings – in the context of cyber-security.



Error and Uncertainty

What is the problem?

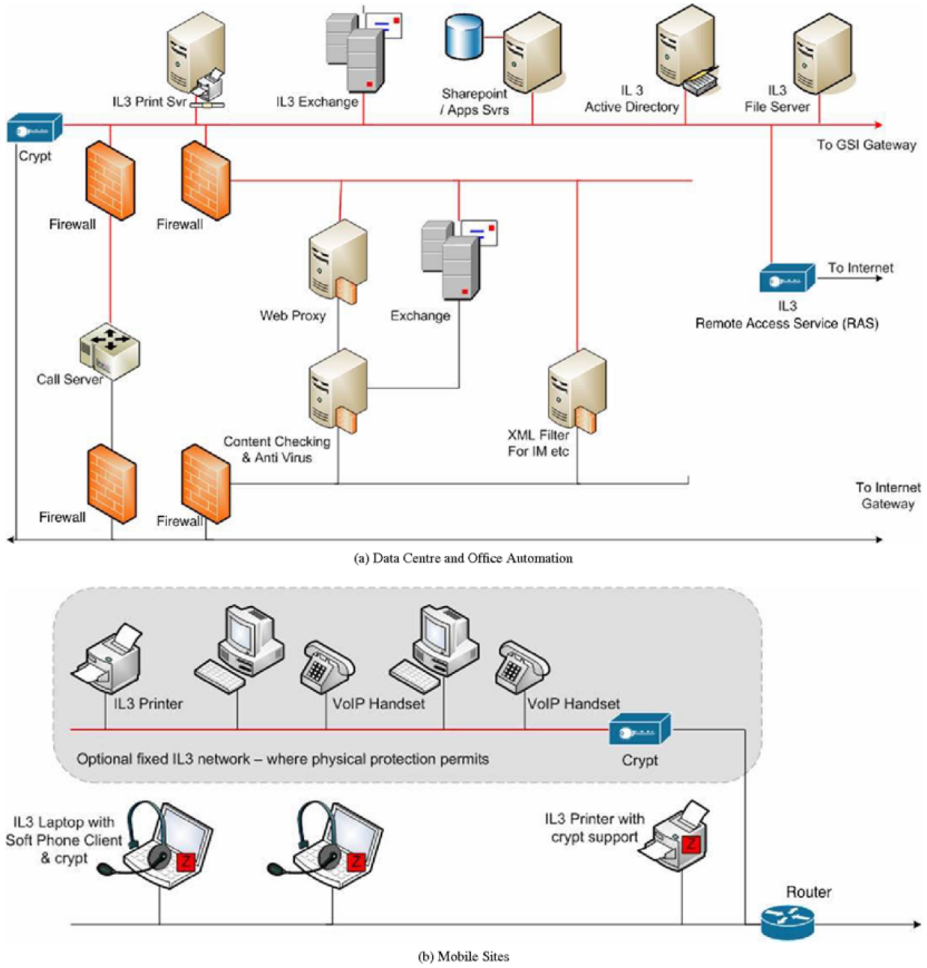
Uncertainty in subjective measurement

It is impossible to completely remove uncertainty from subjective judgements.

This can arise from a variety of sources:

- Limited or out of date information.
- Unfamiliarity with a technology or situation.
- Imprecision, ambiguity or vagueness in the measure, question or scenario.

Cyber systems are often complex and rapidly changing. This compounds the issue of uncertainty in assessments.



Uncertainty in subjective measurement

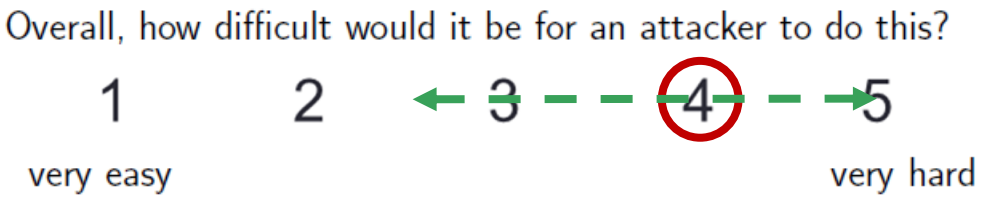
When forced to give a discrete response, uncertainty (or vagueness) is converted into error.

We want to capture uncertainty at the level of the response, before this happens.

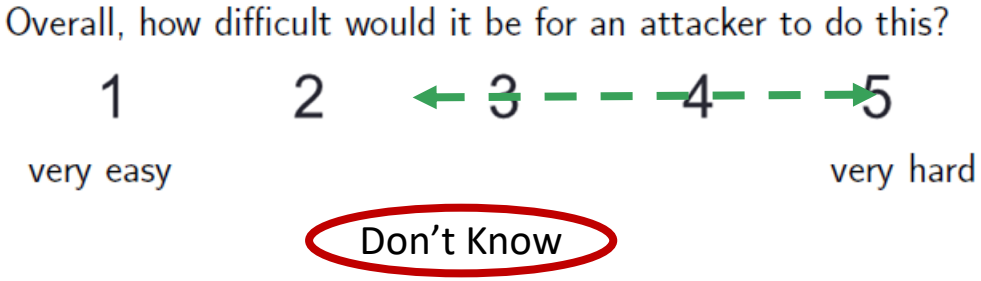
Some existing metrics allow responses of '*Don't know*' or '*Not defined*' - but uncertainty is not all or nothing.

We aim to quantify the uncertainty, alongside a quantitative value for the response itself.

Option 1: Best estimate



Option 2: Opt-out




University of
Nottingham
UK | CHINA | MALAYSIA
Intervals
Are they the solution?

What is our solution?

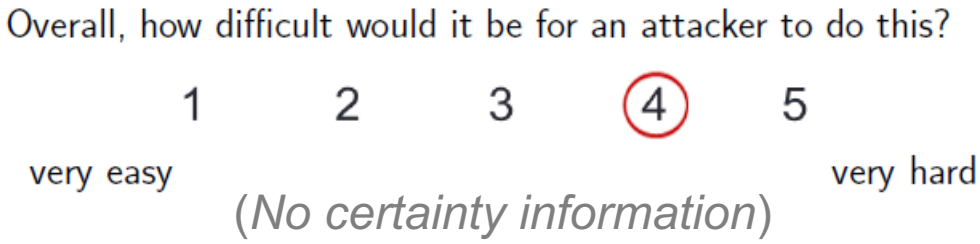
We propose the use of interval-valued responses, obtained in the form of ellipses.

These can capture response range, vagueness and/or uncertainty.

This approach is:

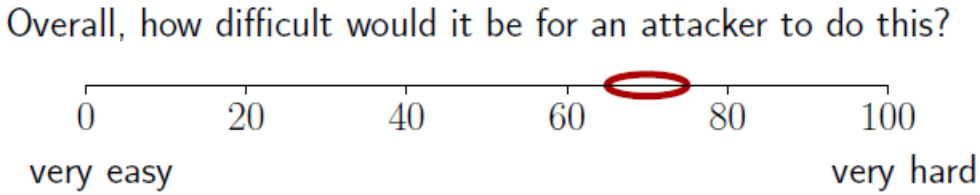
- Quick
- Coherent
- Intuitive
- Assumption free

Discrete (Ordinal)

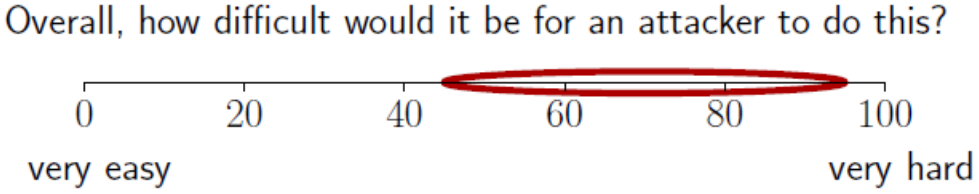


(*No certainty information*)

Interval-valued



(*More certain*)



(*More uncertain*)

University of Nottingham
UK | CHINA | MALAYSIA

Application

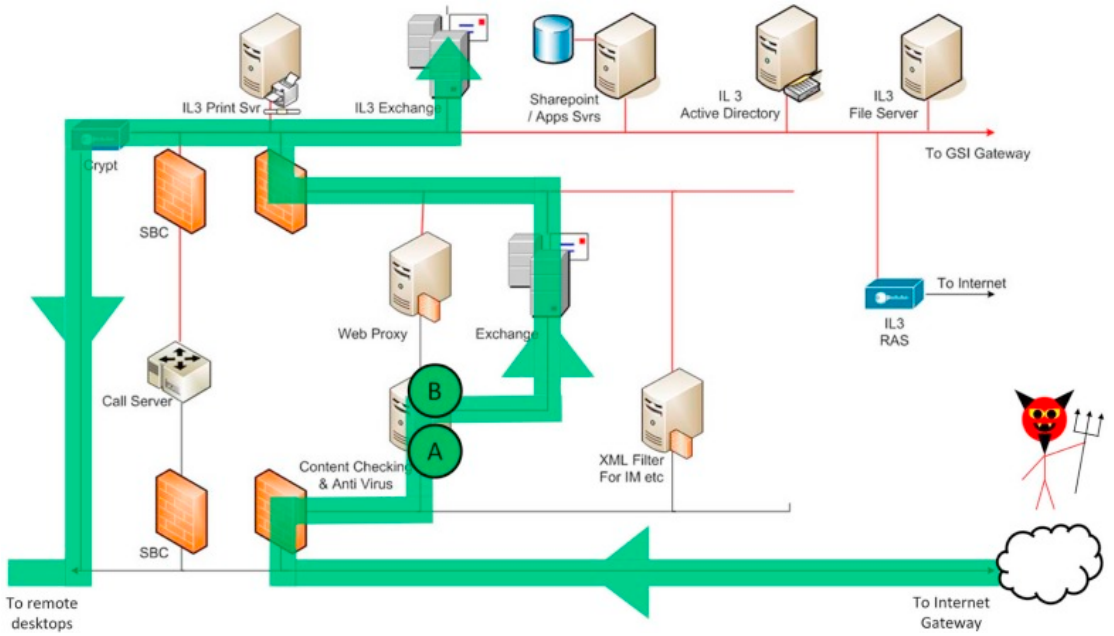
A validation study in CyS context.

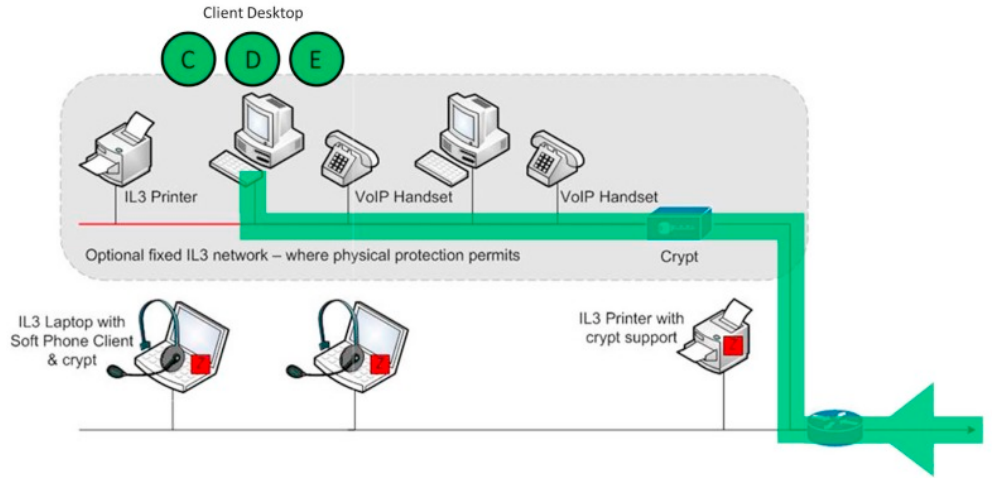
Are intervals useful in vulnerability assessments?

We ran an analysis on interval-valued vulnerability ratings, collected from 38 experts chosen by CESG (NCSC precursor).

Experts rated the overall vulnerability of a number of hops (*attack* or *evade*) within a hypothetical government system – e.g., *bypass gateway content checker*, *overcome client lockdown.*

They also rated each hop on a series of more specific attributes – e.g., *complexity*, *tool availability*, *technological maturity*.





Attack hop ratings

Overall difficulty - Overall, how difficult would it be for an attacker to do this?

1. **Complexity** - How complex is the target component (e.g. in terms of size of code, number of sub-components)?
2. **Interaction** - How much does the target component process/interact with any data input?
3. **Frequency** - How often would you say this type of attack is reported in the public domain?
4. **Availability of tool** - How likely is it that there will be a publicly available tool that could help with this attack?
5. **Inherent difficulty** - How inherently difficult is this type of attack? (i.e. how technically demanding would it be to do from scratch, with no tools to help.)
6. **Maturity** - How mature is this type of technology?
7. **Going unnoticed** - How easy is it to carry this attack out without being noticed?

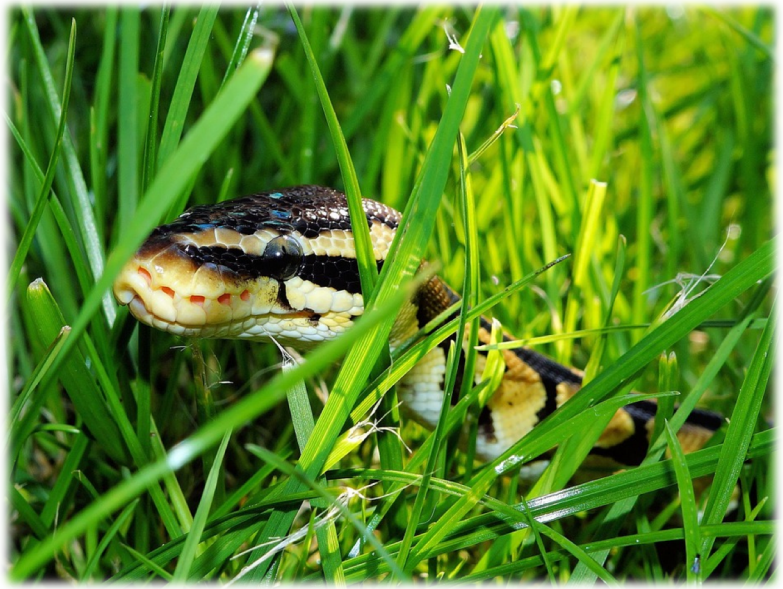
Evade hop ratings

Overall difficulty - Overall, how difficult would it be for an attacker to do this?

1. **Complexity** - How complex is the job of providing this kind of defence?

2. **Availability of information** - How likely is it that there will be publicly available information that could help with evading defence?

3. **Maturity** - How mature is this type of technology?



Are intervals useful in vulnerability assessments?

In this study, our primary aims were:

- Identify which attributes substantially influence overall hop vulnerability, and quantify these contributions.
  - By establishing the best predictive models.

- Determine whether expert uncertainty, captured through interval widths, provides added predictive value over only the position of the ratings.

What is the best mix of factors, and does this include uncertainties?

Attack	attribute	var
	Complexity	c
	Interaction	t
	Frequency	f
	Tool availability	a
	Inherent difficulty	d
	Maturity	r
	Going unnoticed	g
	Overall difficulty	**o**

Evade	attribute	var
	Complexity	c
	Info availability	a
	Maturity	r
	Overall difficulty	**o**

University of Nottingham
UK | CHINA | MALAYSIA

Analysis and Results

A validation study in CyS context.

Analysis

We used linear mixed effects modelling (an extension of linear regression) – to examine the effects of both midpoints and widths of each attribute rating together.

But we have separate models for the two hop types: *Attack* and *Evade*.

We also predict overall rating, and uncertainty around this, in separate models.

This gives a total of four models – e.g.

$$\gamma^z_{i,j} = \beta^z_0 + E^z_m + E^z_w + E^z_{mw} + \mu^z_i + \mu^z_j + \epsilon^z_{i,j}.$$

Shared baseline rating

Differences in baseline ratings between expert and hop

Error

$$E^z_m = \beta^z_1 x^{cm}_{i,j} + \beta^z_2 x^{am}_{i,j} + \beta^z_3 x^{rm}_{i,j}$$

$$E^z_w = \beta^z_4 x^{cw}_{i,j} + \beta^z_5 x^{aw}_{i,j} + \beta^z_6 x^{rw}_{i,j}$$

$$E^z_{mw} = \beta^z_7 (x^{cm}_{i,j} \cdot x^{cw}_{i,j}) + \beta^z_8 (x^{am}_{i,j} \cdot x^{aw}_{i,j}) + \beta^z_9 (x^{rm}_{i,j} \cdot x^{rw}_{i,j})$$

Analysis

Seeing as the models can have so many input variables, and we don't know if all are important, we use a variable reduction process.

Simply, we begin with all variables, then:

- Remove the variable with the least effect.
- Test if this significantly worsens the model.
- If not, keep it out.
- Repeat…

$$\gamma_{i,j}^{Aoz} = \beta_0^z + A_m^z + A_w^z + A_{mw}^z + \mu_i^z + \mu_j^z + \epsilon_{i,j}^z$$

$$A_m^z = \beta_1^z x_{i,j}^{cm} + \beta_2^z x_{i,j}^{tm} + \beta_3^z x_{i,j}^{fm} + \beta_4^z x_{i,j}^{am} + \beta_5^z x_{i,j}^{dm} + \beta_6^z x_{i,j}^{rm} + \beta_7^z x_{i,j}^{gm}$$

$$A_w^z = \beta_8^z x_{i,j}^{cw} + \beta_9^z x_{i,j}^{tw} + \beta_{10}^z x_{i,j}^{fw} + \beta_{11}^z x_{i,j}^{aw} + \beta_{12}^z x_{i,j}^{dw} + \beta_{13}^z x_{i,j}^{rw} + \beta_{14}^z x_{i,j}^{gw}$$

$$A_{mw}^z = \beta_{15}^z (x_{i,j}^{cm} \cdot x_{i,j}^{cw}) + \beta_{16}^z (x_{i,j}^{tm} \cdot x_{i,j}^{tw}) + \beta_{17}^z (x_{i,j}^{fm} \cdot x_{i,j}^{fw}) + \beta_{18}^z (x_{i,j}^{am} \cdot x_{i,j}^{aw}) + \beta_{19}^z (x_{i,j}^{dm} \cdot x_{i,j}^{dw}) + \beta_{20}^z (x_{i,j}^{rm} \cdot x_{i,j}^{rw}) + \beta_{21}^z (x_{i,j}^{gm} \cdot x_{i,j}^{gw})$$

Etc.

Results – Final Models

Effects on attack hop overall difficulty ratings.

Less difficult if:

- Frequently reported
- Tools available
- Immature technology

More difficult if:

- Higher inherent difficulty*
- Uncertain about maturity
- Easy to go unnoticed (if difficult to conduct may also be difficult to detect)

Fixed Effects Estimates	β	SE	t	p
Intercept : $(_0)$	.012	.066	.175	.861
Frequency m : $(x^{fm}_{i,j})$	-.223	.044	-5.065	$<$ .001
Availability Tool m : $(x^{am}_{i,j})$	-.201	.044	-4.574	$<$ .001
Inherent Difficulty m : $(x^{dm}_{i,j})$	.357	.030	11.890	$<$ .001
Maturity m : $(x^{rm}_{i,j})$	.126	.030	4.159	$<$ .001
Going Unnoticed. m : $(x^{gm}_{i,j})$	.142	.027	5.194	$<$ .001
Maturity w : $(x^{rw}_{i,j})$	.071	.027	2.612	.009
Availability Tool $m \cdot w$: $(x^{am}_{i,j} \cdot x^{aw}_{i,j})$	.077	.036	2.168	.031
Random Effects Estimates	μ			
Expert intercept $(_i)$	.183			
Hop intercept $(_j)$	.204			
Residual $\epsilon_{i,j}$	.502			

N = 532, DF = 524, AIC = 896.7, BIC=943.6

- Effect of tool availability is greater when certain, and reduced when uncertain.

Results – Final Models

Effects on evade hop overall difficulty ratings.

Less difficult if:

- Information available*
- Immature technology

More difficult if:

- Uncertain about info availability

- Certain high complexity
- Uncertain low complexity

Fixed Effects Estimates	β	SE	t	p
Intercept : $(_0)$	-.023	.133	-.173	.863
Availability Information m : $(x^{am}_{i,j})$	-.240	.049	-4.895	$<$ .001
Maturity m : $(x^{rm}_{i,j})$	.177	.051	3.459	$<$ .001
Availability Information w : $(x^{aw}_{i,j})$	.142	.049	2.878	.004
Complexity $m \cdot w$: $(x^{cm}_{i,j} \cdot x^{cw}_{i,j})$	-.105	.053	-1.993	.047
Random Effects Estimates	μ			
Expert intercept $(_i)$	.457			
Hop intercept $(_j)$	.340			
Residual $\epsilon_{i,j}$	.772			

N = 418, DF = 413, AIC = 1081.8, BIC=1114.0

Results – Final Models

Effects on attack hop uncertainty in overall ratings.

More **certain** if:

- Frequently reported
- Mature technology

More **uncertain** if:

- Tools available
- Unsure about frequency
- Unsure about tool availability
- Unsure about inherent difficulty*
- Unsure about going unnoticed

Fixed Effects Estimates	β	SE	t	p
Intercept : $(_0)$	-.031	.036	-.857	.392
Frequency m : $(x^{fm}_{i,j})$	-.116	.045	-2.614	.009
Availability Tool m : $(x^{am}_{i,j})$	.131	.045	2.934	.003
Maturity m : $(x^{rm}_{i,j})$	-.093	.031	-3.013	.003
Frequency w : $(x^{fw}_{i,j})$	.141	.035	4.034	$<$ .001
Availability Tool w : $(x^{aw}_{i,j})$	.095	.039	2.420	.016
Inherent Difficulty w : $(x^{dw}_{i,j})$	.406	.037	10.959	$<$ .001
Going Unnoticed w : $(x^{gw}_{i,j})$	.268	.036	7.399	$<$ .001
Maturity $m \cdot w$: $(x^{rm}_{i,j} \cdot x^{rw}_{i,j})$	-.122	.035	-3.484	$<$ .001
Going Unnoticed $m \cdot w$: $(x^{gm}_{i,j} \cdot x^{gw}_{i,j})$	-.080	.035	-2.270	.024
Random Effects Estimates	μ			
Expert intercept $(_i)$	.127			
Hop intercept $(_j)$	.000			
Residual $\epsilon_{i,j}$	.609			

N = 532, DF = 522, AIC = 1066.3, BIC=1121.7

- The effect of maturity was driven by when this was uncertain.
- The effect of uncertainty in going unnoticed was greater when this was difficult.

Results – Final Models

Effects on evade hop uncertainty in overall ratings.

More **uncertain** if:

- Unsure about complexity
- Unsure about info availability*
- Unsure about maturity

Fixed Effects Estimates	β	SE	t	p
Intercept : $(_0)$	-.000	.058	-.000	$>$ .999
Complexity w : $(x^{cw}_{i,j})$	.241	.046	5.200	$<$ .001
Availability Information w : $(x^{aw}_{i,j})$	.440	.045	9.683	$<$ .001
Maturity w : $(x^{rw}_{i,j})$	.134	.045	2.982	.003
Random Effects Estimates	μ			
Expert intercept $(_i)$	.070			
Hop intercept $(_j)$	.159			
Residual $\epsilon_{i,j}$	.643			

N = 418, DF = 414, AIC = 863.0, BIC=891.2

Conclusions

What did we learn?

Summary and Conclusions

Key findings:

- Not all rated factors were found to contribute significantly to overall ratings, or their associated uncertainty – even though these were pre-selected by experts.
    - For instance, neither component complexity, nor interaction with input data had any consistent effect on attack difficulty.

- The inherent difficulty of attacking a hop (to do from scratch), and availability of information to help evade a hop were found to have the strongest influences on overall difficulties.

- Uncertainty around these same factors had the strongest influence on overall uncertainty.

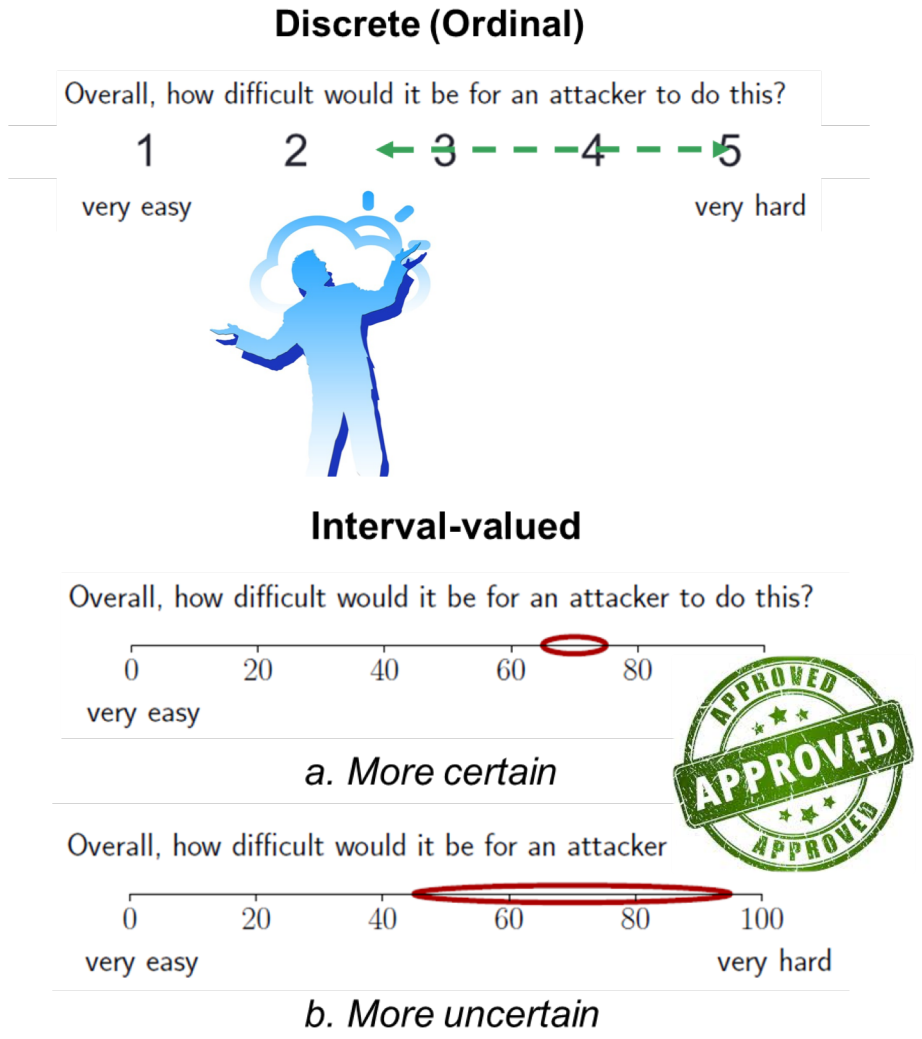
- But other factors were also found to be significant.

Summary and Conclusions

- ‘Crossover’ was found:
  - ➢ Some non-uncertainty attribute ratings influenced overall uncertainty.
  - ➢ Uncertainty around some attribute ratings influenced overall hop difficulty ratings.

- This finding indicates that quantitative capture of degrees of uncertainty can provide substantial added value in vulnerability assessments.

- This can be achieved quickly, coherently and effectively through the use of an interval-valued response format.




University of
Nottingham
UK | CHINA | MALAYSIA
Any
Questions?
Thanks for listening.


University of
Nottingham
UK | CHINA | MALAYSIA
DECSYS
What does it do?

University of Nottingham
UK | CHINA | MALAYSIA

What is DECSYS?

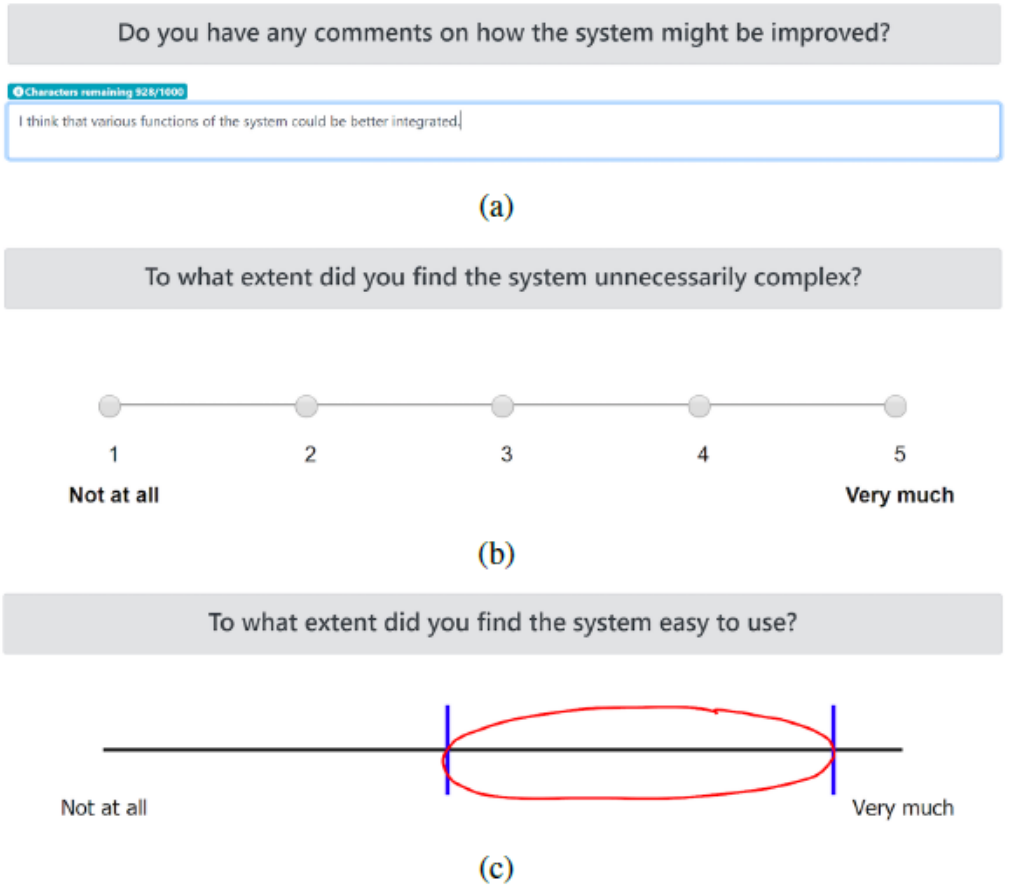
DECSYS is the '***D**iscrete and **E**llipse-based response **C**apture **SYS**tem*'.

This software allows survey designers to capture interval-valued responses.

It is designed to be used with touch-screen and stylus-based devices (e.g. Microsoft Surface).

So, responses are similar to paper surveys, but much easier to collect and collate.

Initially developed for Cyber Security applications (i.e. expert vulnerability assessments).



This work was part funded by the UK's National Cyber Security Centre (NCSC) and the EPSRC EP/P011918/1 grant.

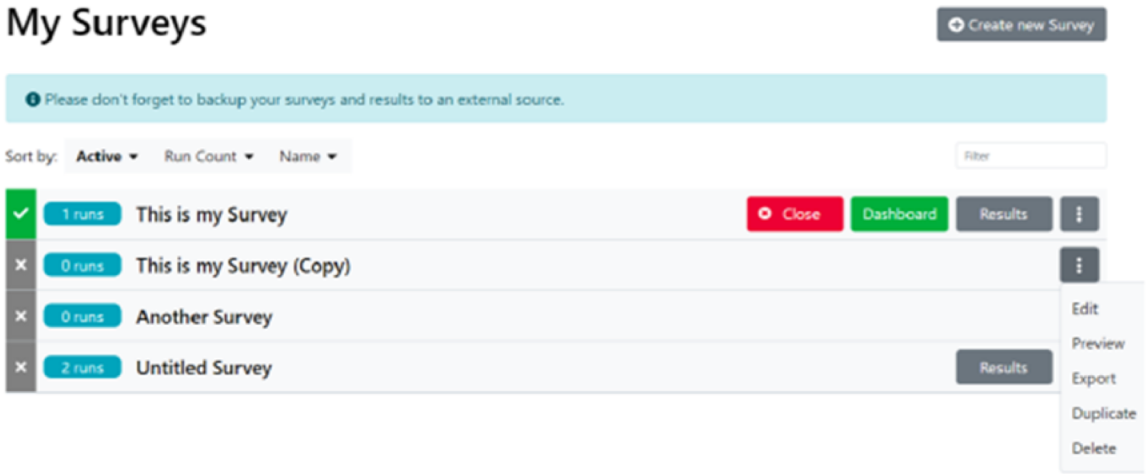
What can DECSYS do?

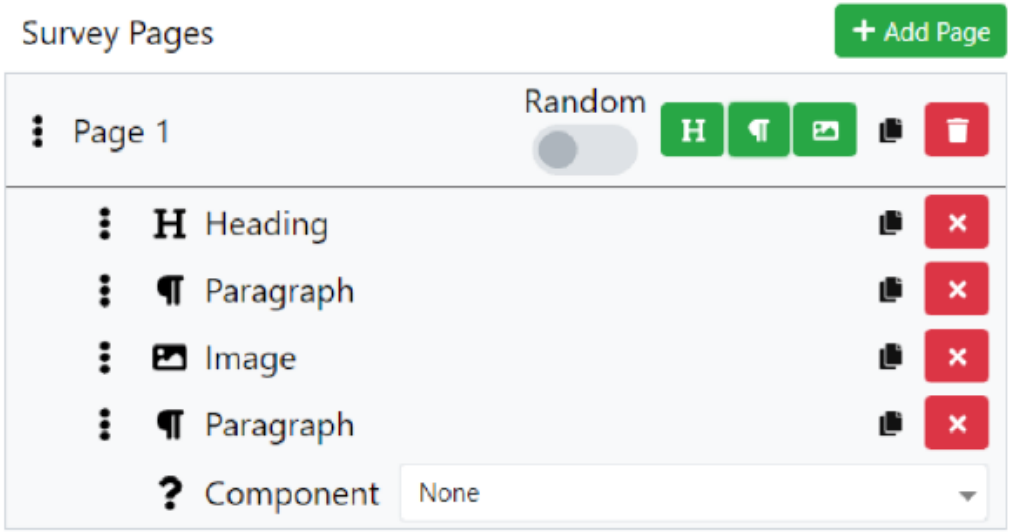
DECSYS supports both creation and administration of surveys.

It is designed to be *versatile*.

- It supports multiple response formats:
  - ✓ Conventional (tick, discrete, free-text)
  - ✓ Interval-valued (ellipse)

- It provides many customisable options:
  - ✓ Tailored survey content
  - ✓ Tailored survey formats
  - ✓ Stimulus presentation (images)
  - ✓ Question order (randomisation)





How can DECSYS be used?

DECSYS supports both creation and administration of surveys.

It is designed to be *versatile*.

- It supports multiple distribution methods:

  ✓ Locally - '*Workshop-mode*'
  (secure setting - ideal for companies)

  ✓ To a wide audience - '*Online-mode*'
  (still in development)

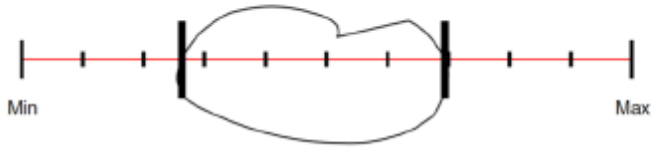
Is DECSYS user friendly?

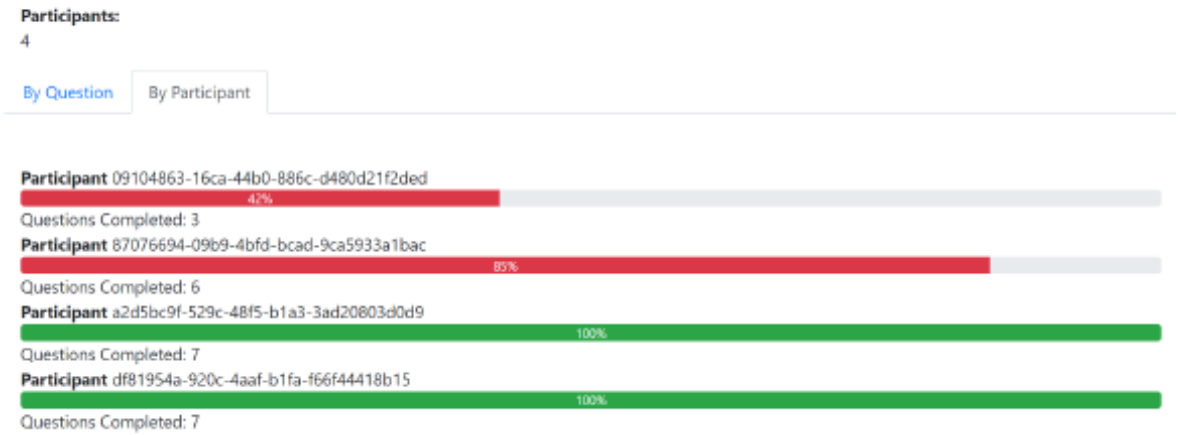
DECSYS supports both creation and administration of surveys.

It is designed to be *easy to use*.

- It offers visual aids for experimenters:
  - ✓ Building surveys
    (Page & full survey previews)
  - ✓ Running surveys
    (Live progress dashboard)

- Results can be easily exported.
  (standardised .json, or .csv format)





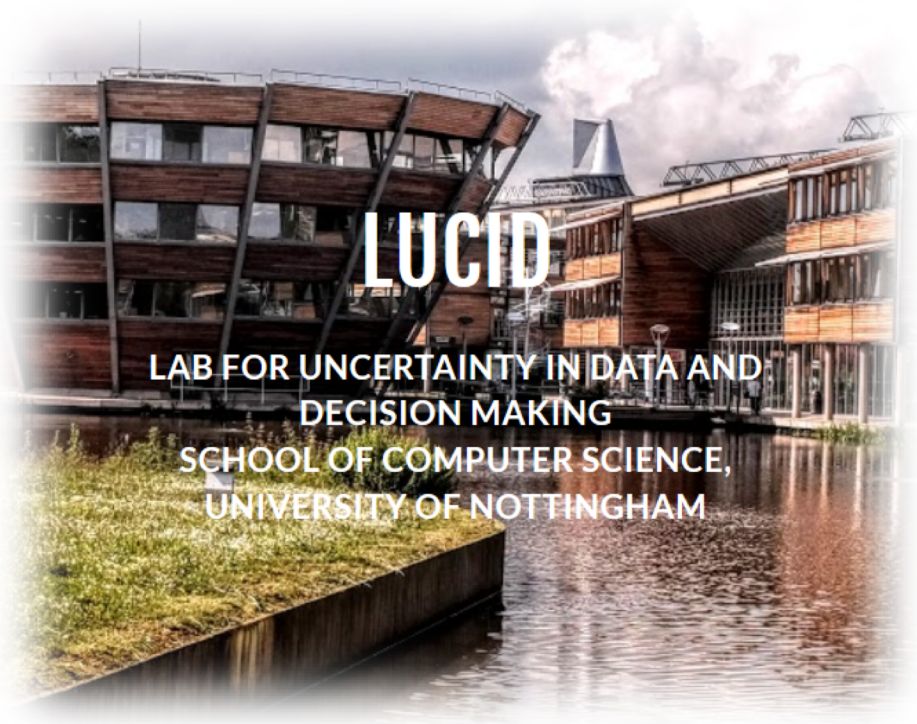
How can you get DECSYS?

DECSYS is open-source, and freely available for academic use.

An initial version is available via:
http://www.lucidresearch.org/decsys.html

Development is ongoing.

- Features are being added and refined.
- We encourage third-party development.
- Feedback and suggestions are welcome.



Fuzzy logic/IAA

Another approach

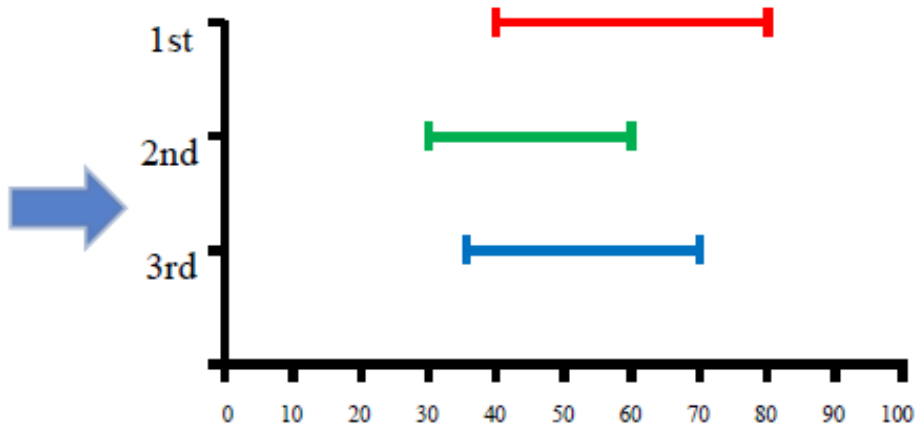
- Example: suppose four experts have been surveyed three times on the same subject

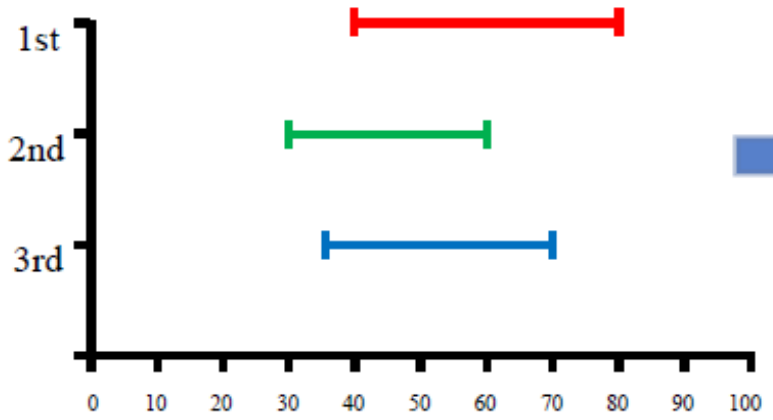
	Expert A		Expert B		Expert C		Expert D	
1st Answer	40	80	40	85	20	80	25	75
2nd Answer	30	60	50	80	30	85	35	75
3rd Answer	35	70	45	95	25	75	30	70

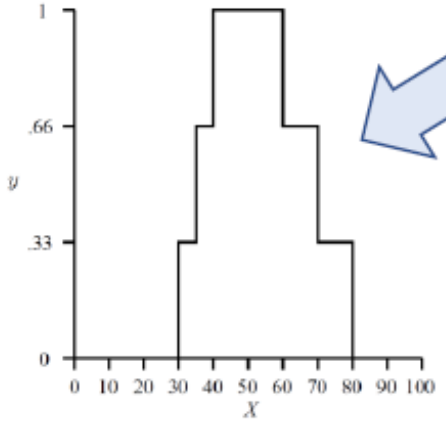
- Their answers are intervals in the range [0,100]

	Expert A		Expert B		Expert C		Expert D	
1st Answer	40	80	40	85	20	80	25	75
2nd Answer	30	60	50	80	30	85	35	75
3rd Answer	35	70	45	95	25	75	30	70

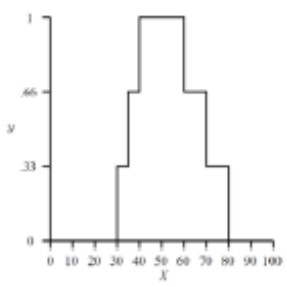
	Expert A	
1st Answer	40	80
2nd Answer	30	60
3rd Answer	35	70

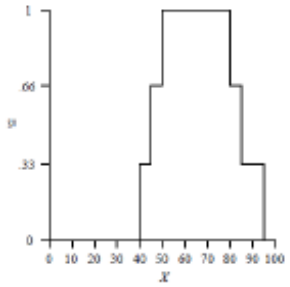


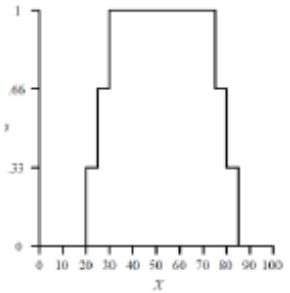


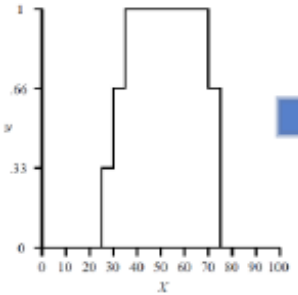


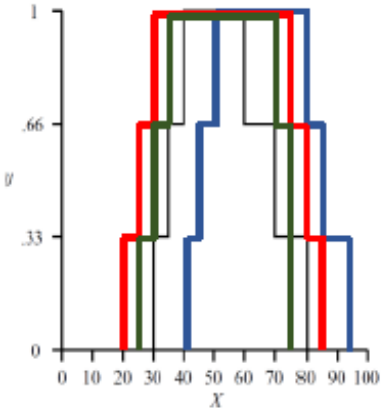
They do not usually look so nice…





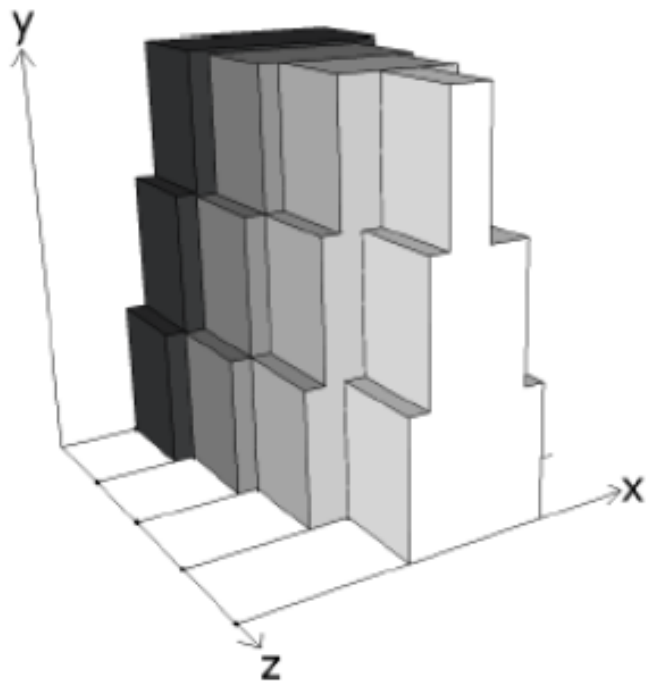




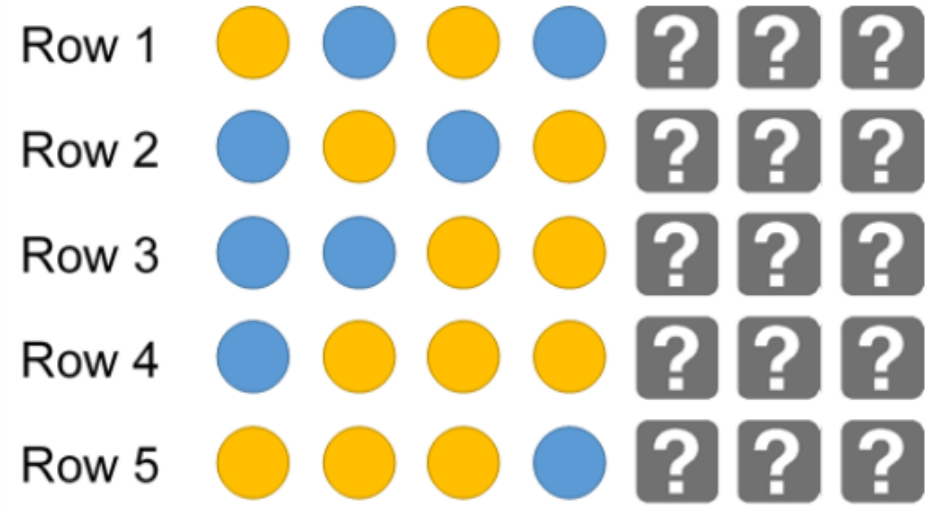


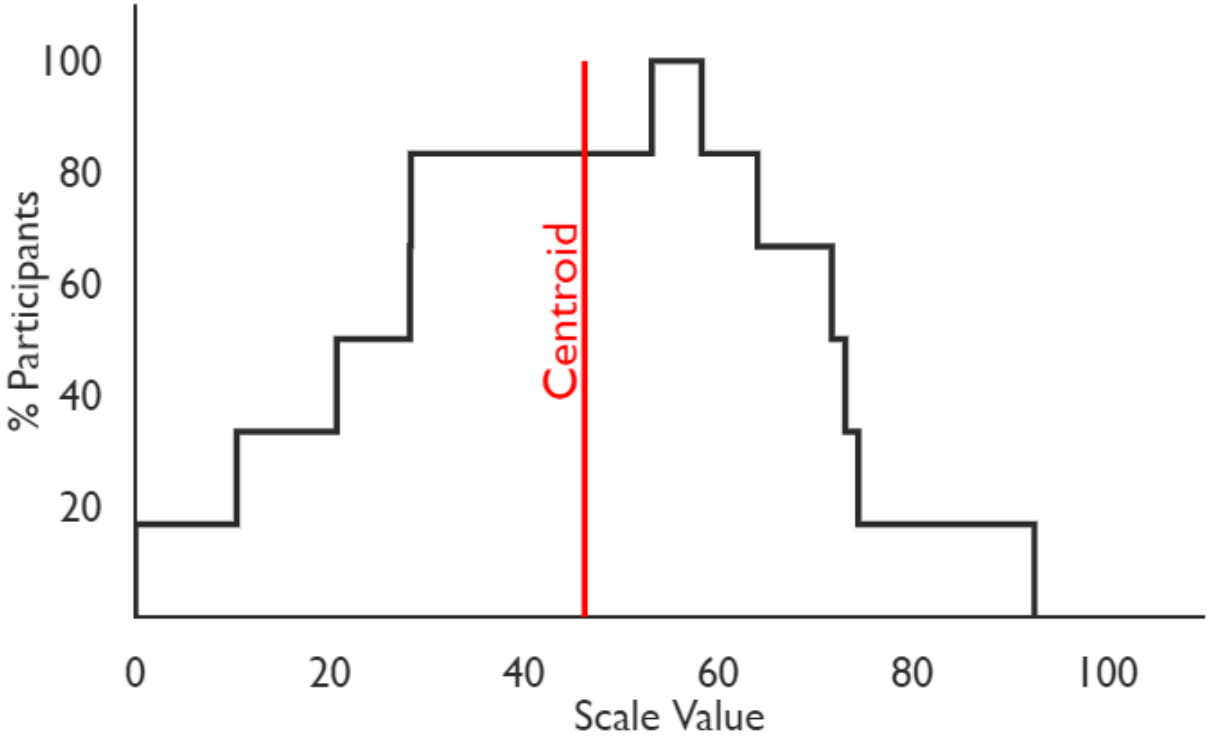
Resulting in a form of a *model (a type-2 fuzzy set)* that:

- Represents inter- and intra-expert variation in two distinct dimensions
- Makes no assumptions about the distribution of the data
- Discards minimal information
  - No outlier removal
- Suitable for inference, similarity, decision support, etc.

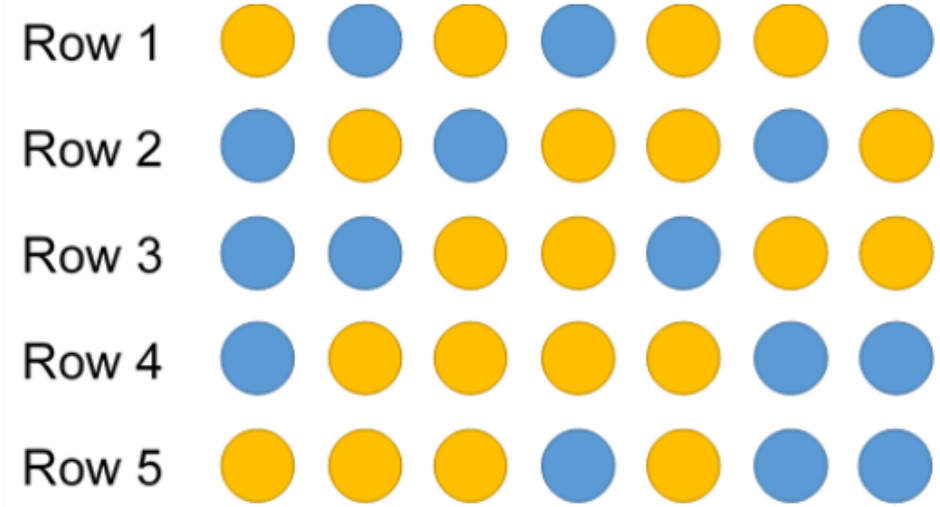


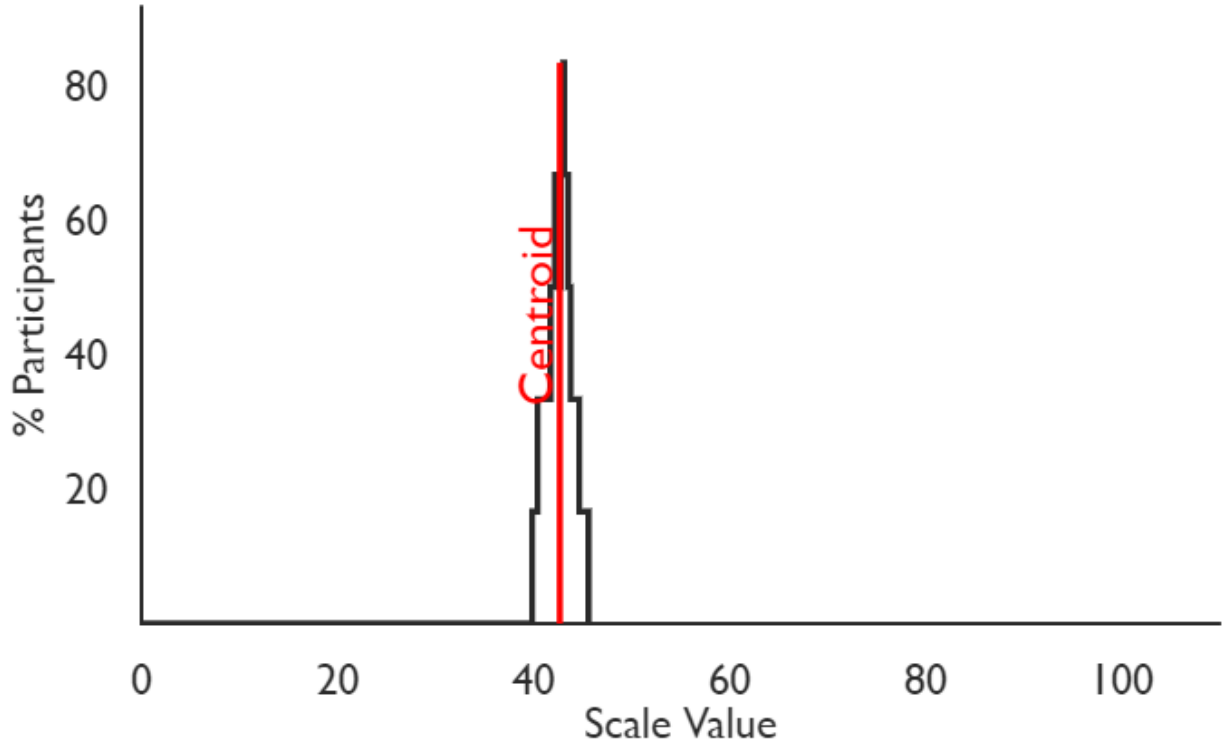
IAA examples



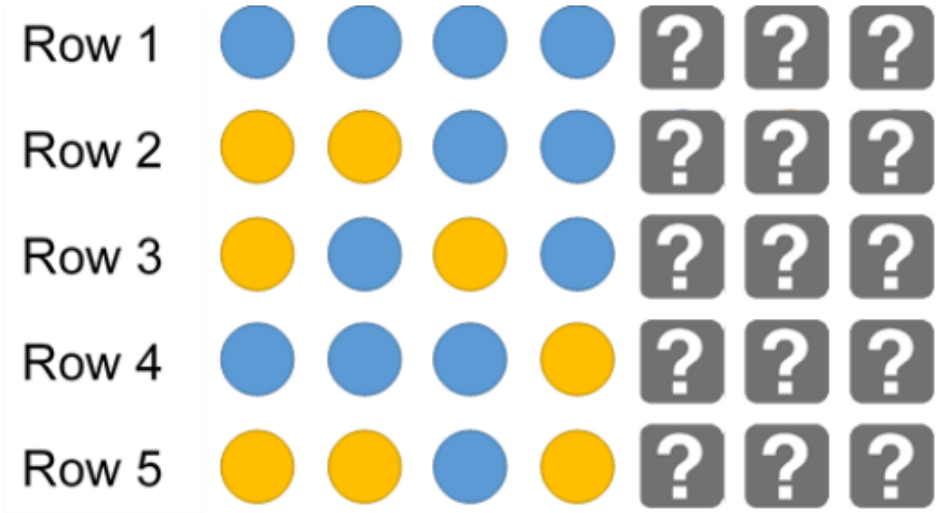


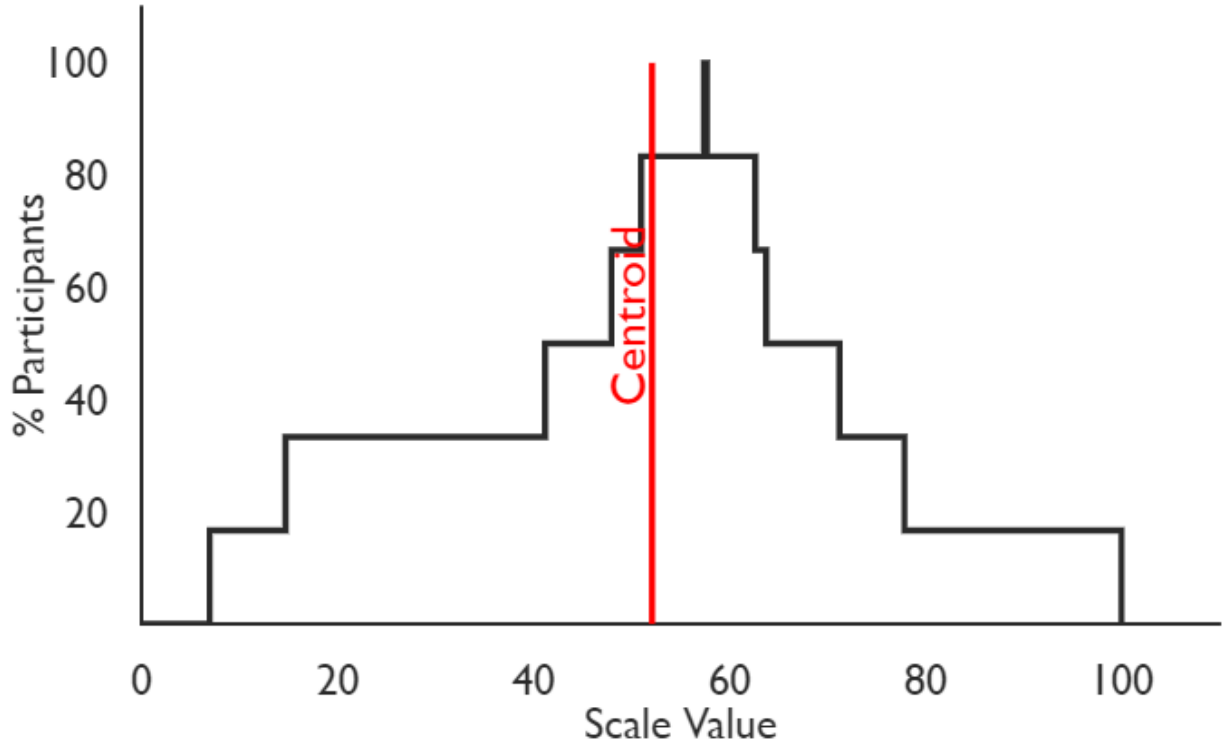
IAA examples





IAA examples





IAA examples