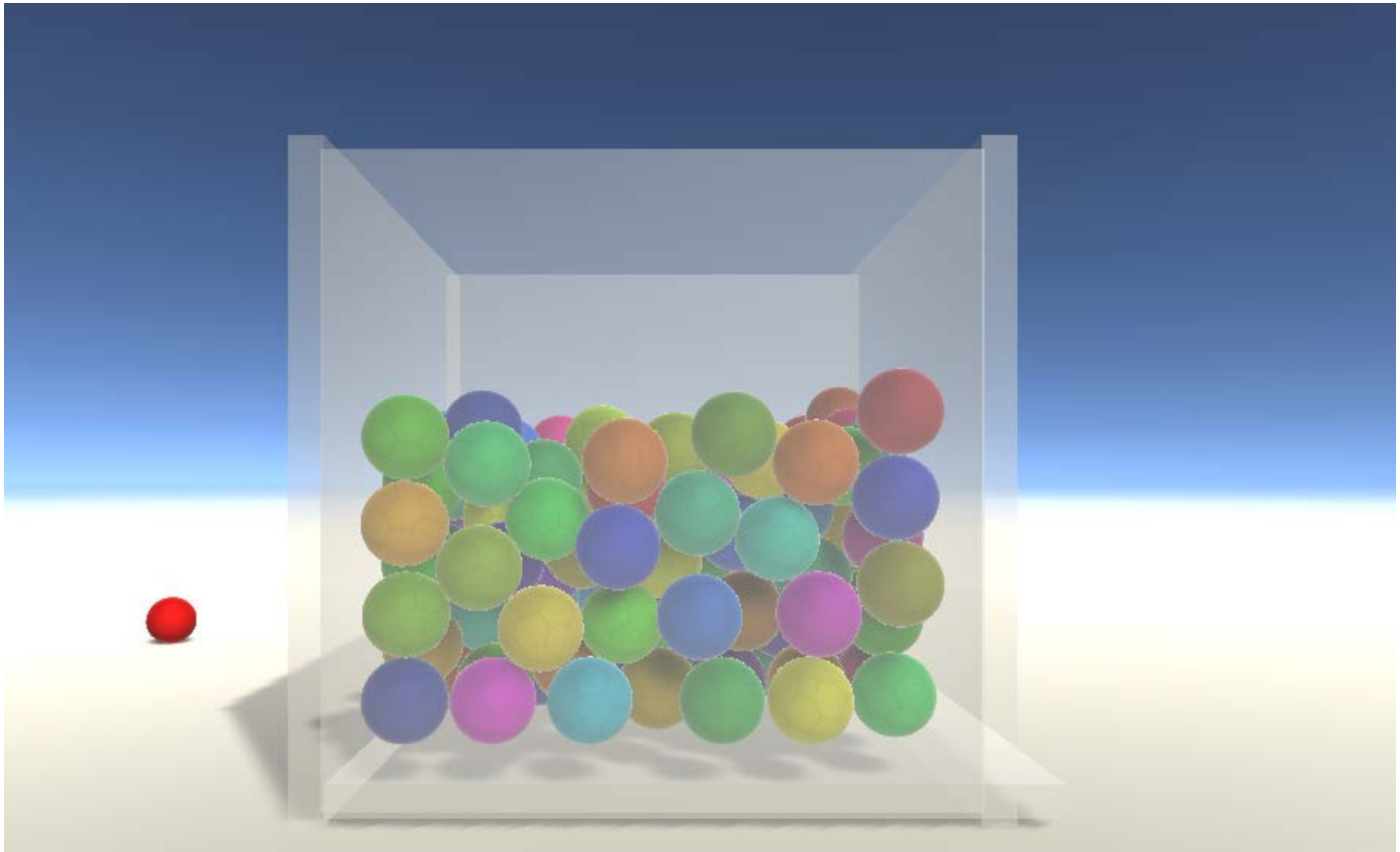
Using games to train or test our ability to express epistemic uncertainty

Ullrika Sahlin

Lund University, Sweden



1. Daniel Kahneman was awarded the Nobel price in Economy in A) 1978 or B) 2002
2. Ingmar Bergman made the movie A) The Holy Grail or B) The Seventh Seal

State how certain you are in this answer

5 If you have no clue if the answer is A or B

6
7
8
9 If you think you know, but are not fully certain

10 If you are certain that you are right

Gage and Spiegelhalter 2016

1. Daniel Kahneman was awarded the Nobel price in Economy in A) 1978 or B) 2002
2. Ingmar Bergman made the movie A) The Holy Grail or B) The Seventh Seal

Your confidence	5	6	7	8	9	10
Score if you were right	0	9	16	21	24	25
Score if you were wrong	0	-11	-24	-39	-56	-75

Gage and Spiegelhalter 2016

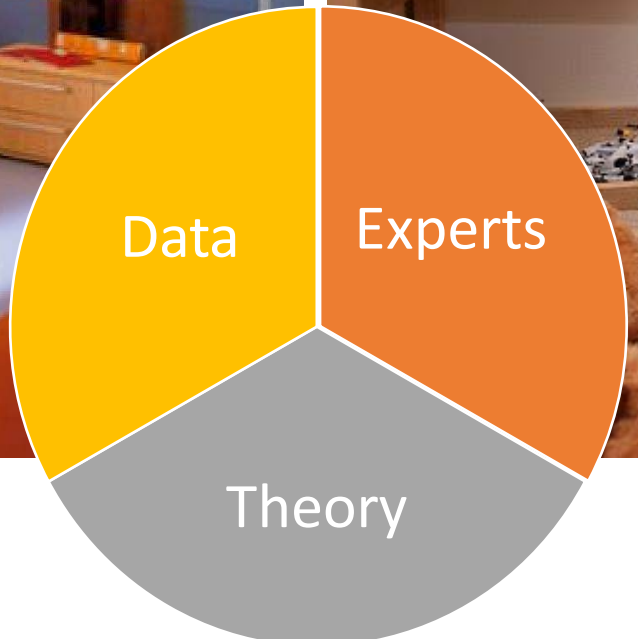
Science

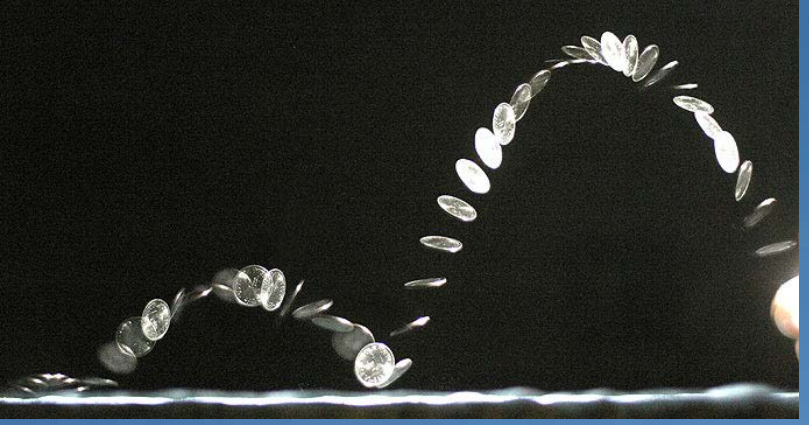
IKEA

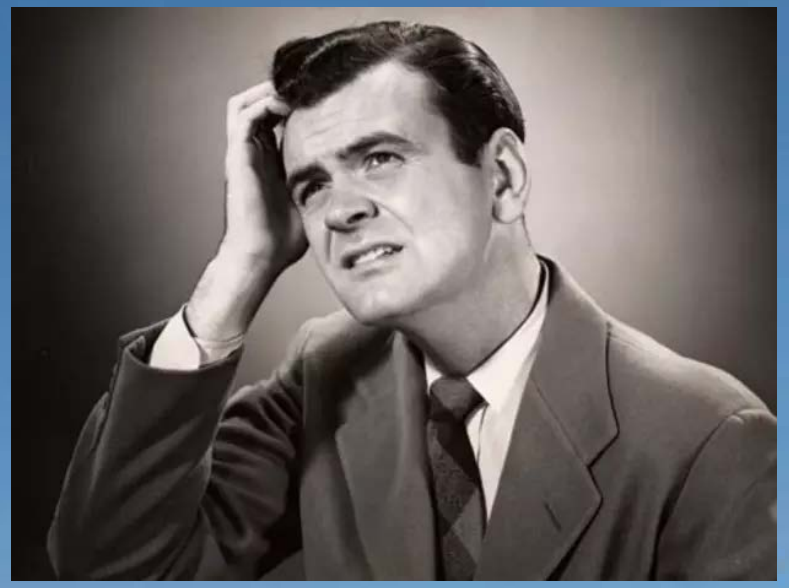
Science

Scientific assessments

Epistemic uncertainty



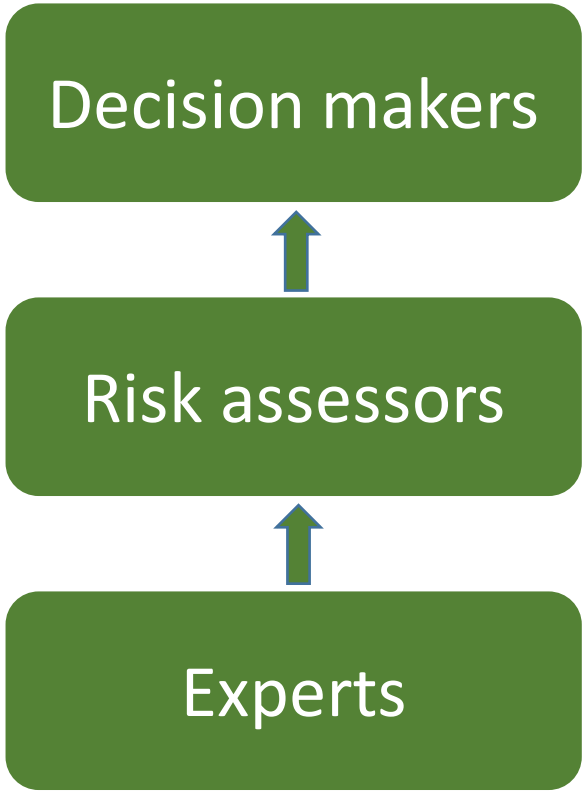




Getty images

Epistemic uncertainty in scientific assessments

- Skills in methods for uncertainty analysis
- Skills in communication of epistemic uncertainty
- Epistemic uncertainty is personal
- Who's uncertainty?
- Skills in being uncertain



Lundaloppet predictive challenge

Guess your time and express your uncertainty in your running time in which way you want!

After the race on and the actual running times are known, we will evaluate the performances of all predictions.

Lundaloppet predictive challenge

- Few respondents use probability to express uncertainty
- Plain intervals are popular
- Many answers are very precise (and often wrong)
- Some characterize and motivate their uncertainty with a story combined with quantitative estimates
- Some guesses turned out to be cases for black swan events

- A fun and appreciated event
- Comparison difficult across different ways to express uncertainty

50'10'' to 50'40''
(actual time 47)

There is a 50% chance that I will run, and if I do I will run with an average of 30 minutes and a standard deviation of 5 (actual time 27)

When I used to be in a good shape I ran on 25-30 minutes, but today I am not in shape and may take 30-35 minutes. In addition, I may need to bring my 8 year old daughter and then I will run on 40-50 minutes (actual time 36)

Mr C guessed his time but did not run because he forgot to sign up for the race before the deadline

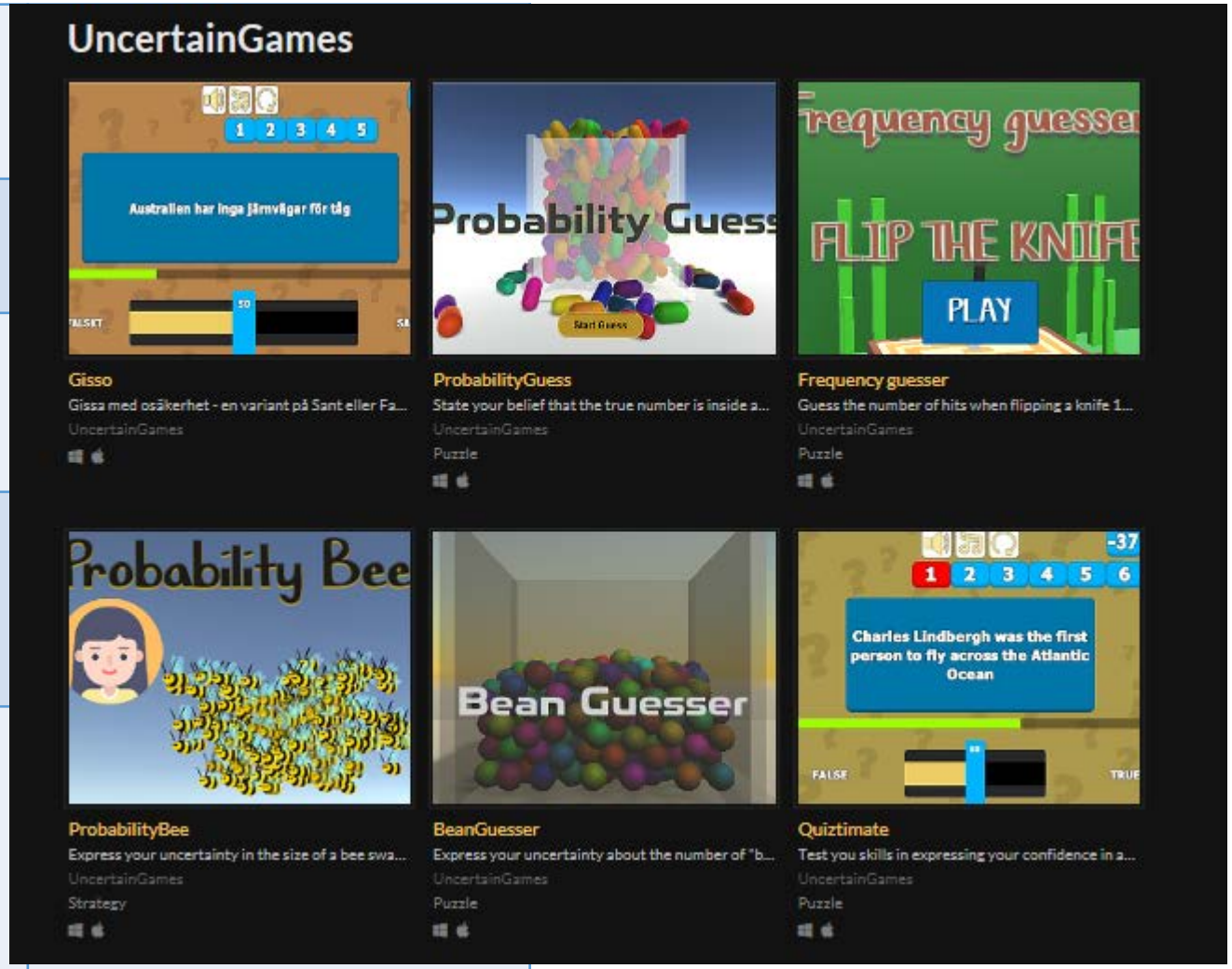
The ability to express and understand epistemic uncertainty **Uncertaincy**

- Are there individual differences in the ability to express and understand epistemic uncertainty?
- Cognitive skills: numeracy, probabilistic reasoning
- Behavioral factors: risk aversion, uncertainty aversion
- Is the ability to express and understand epistemic uncertainty a cognitive ability separate from other abilities?
- Can “uncertaincy” explain other skills and human behavior?
- Can “uncertaincy” improve testing of
  - communication messages of epistemic uncertainty
  - implementation of methods for uncertainty analysis
- How can we measure “uncertaincy”?

Games

https://uncertaingames.itch.io/

Quiztimate (Gisso in SWE)	probability in % for uncertainty in a TRUE/FALSE proposition
Bean Guesser	a probability interval for a number
Probability Guess	a probability for a number to fall inside a given interval
Probability Bee	bet on bins likely to contain the number (roulette method, almost a PDF)
Frequency guesser	use a lower bound to express uncertainty in a frequency with a possibility to explore the system by sampling



Scoring rules and feedback

- A scoring rule is a numerical score used to assess the quality of a probabilistic estimation by where the true outcome (event or value) eventually is revealed
- A scoring rule is proper if the expected score is the highest when the estimation is the same as the true outcome
- A proper scoring rule motivates the player to give an honest, carefully judged answer, reflecting her uncertainty
- Using improper scoring rules creates situations where the player may find strategies which maximise the score in more ways than to give an honest answer
- Feedback on performance and learning is important to increase understanding and stimulate improvement

Games

Quiztimate (Gisso in SWE)	probability in % for uncertainty in a TRUE/FALSE proposition	Quadratic scoring rule for probability
Bean Guesser	a probability interval for a number	Scoring rules for quantiles with a penalty for probability
Probability Guess	a probability for a number to fall inside a given interval	Log scoring rule for probability
Probability Bee	bet on bins likely to contain the number (roulette method, almost a PDF)	The number of bets in the accurate bin
Frequency guesser	use a lower bound to express uncertainty in a frequency with a possibility to explore the system by testing	Imprecise Brier score for lower probability

Quiztimate - Confidence in a TRUE/FALSE statement

Ten TRUE/FALSE

Time pressure

Inspired by exercise Teaching Probability by Gage and Spiegelhalter, 2016.

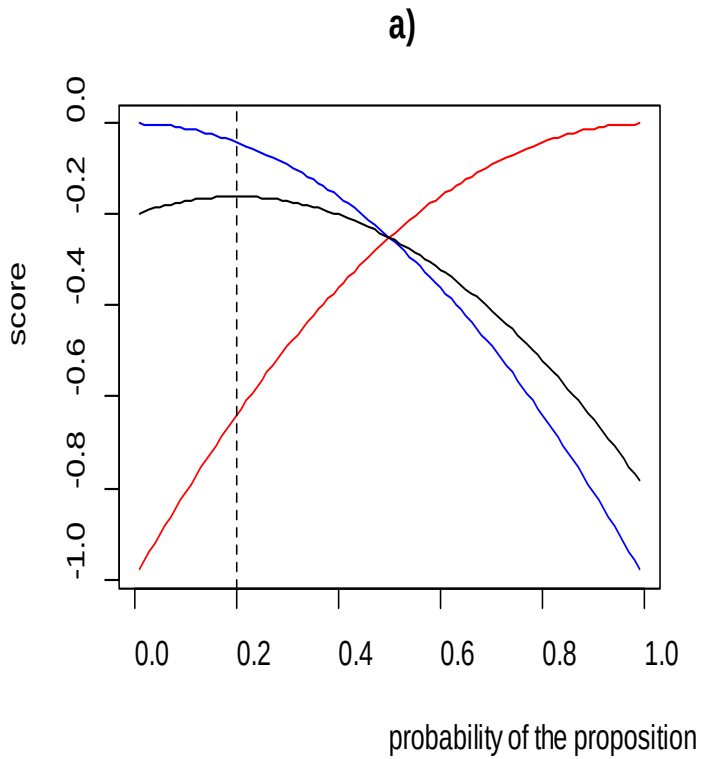
Quiztimate - Confidence in a TRUE/FALSE statement

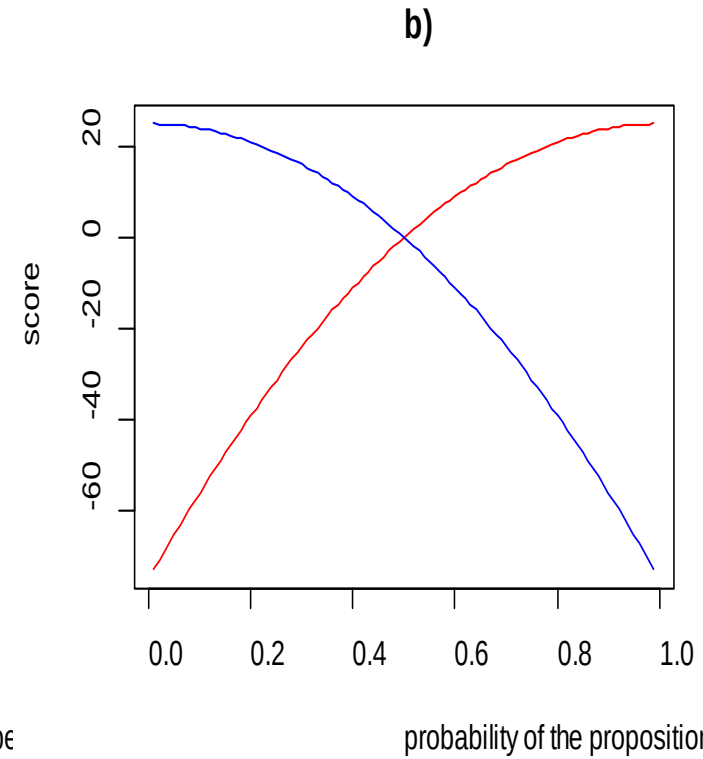
Confidence is measured by p, which is the probability that the proposition is TRUE

Scoring Rule:

$-(1-p)^2$ if TRUE

$-p^2$ if FALSE





Quiztimate

Feedback: Accuracy + Score

“Good. You are not Overconfident” if score >= 0“

”You are overconfident” if score < 0

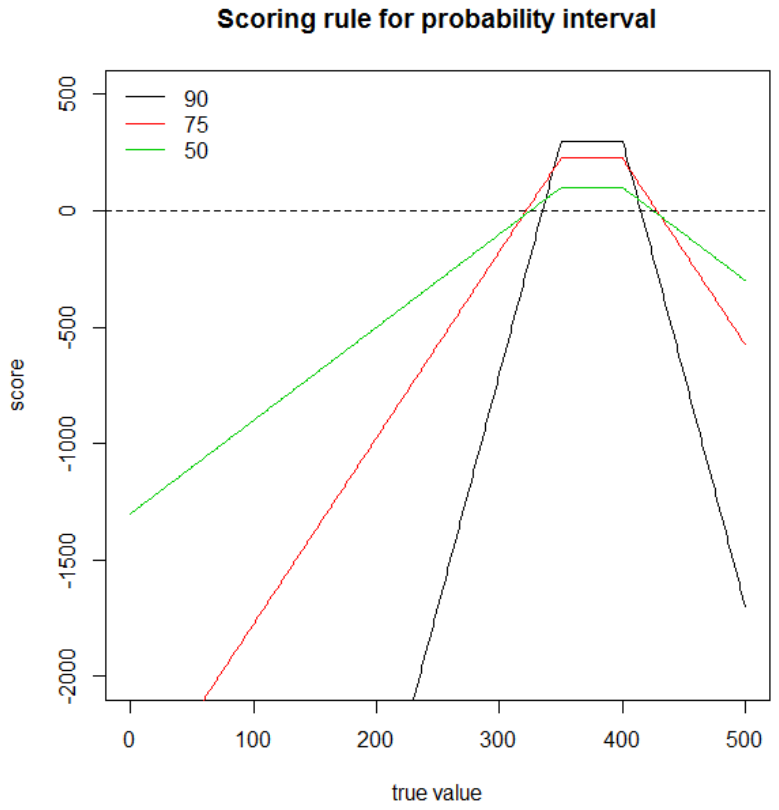
Bean Guesser - Probability interval to express uncertainty in a number

The player assign a probability interval by providing a lower (x) and upper bound (y) and a probability (p < 0.5) for the true value being between these two bounds.

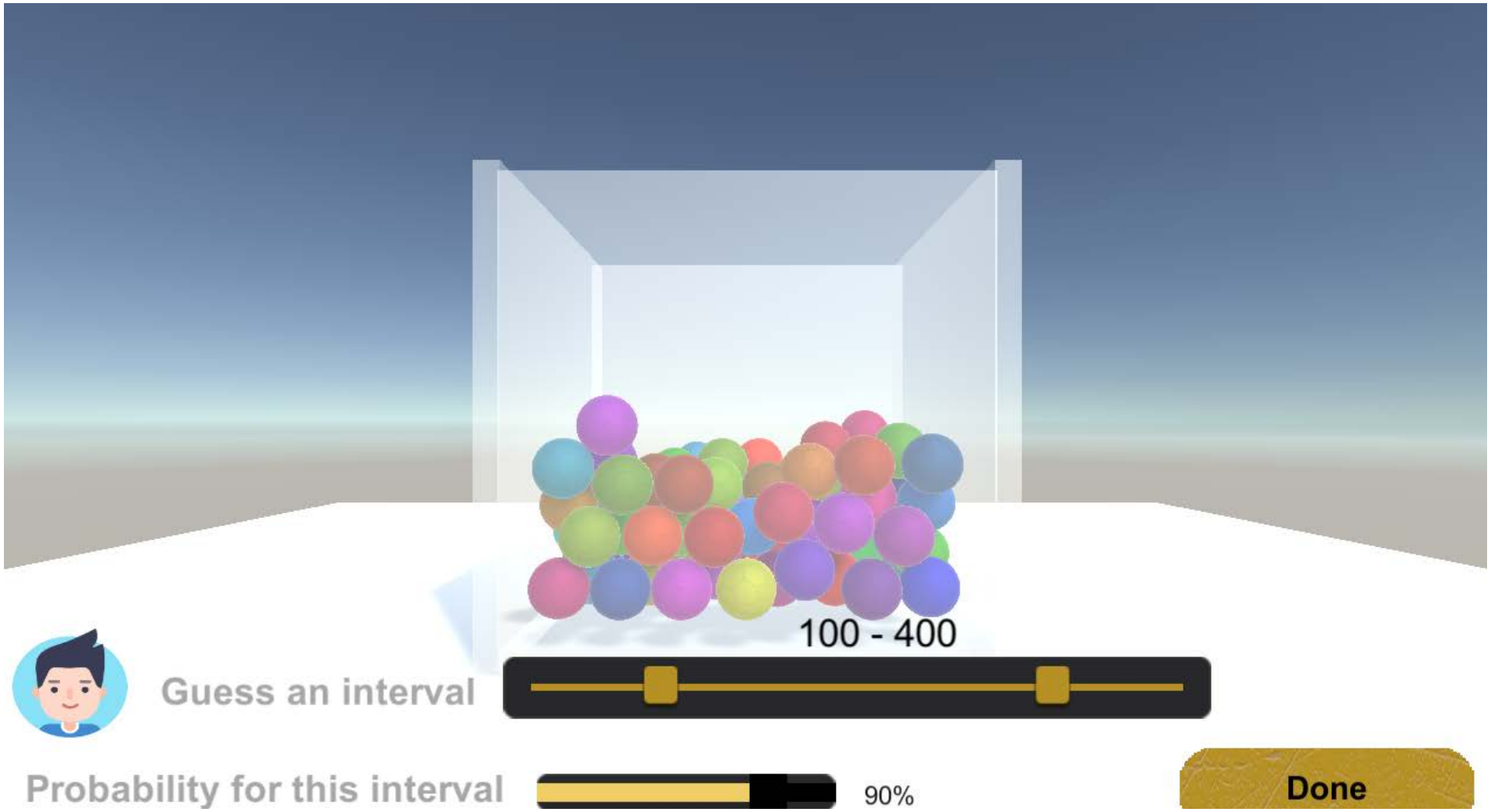
Scoring rule for quantiles modified from Gneiting and Raftery (2007):

$$-100 + 500\alpha(1-\alpha) - \left((y-x) + \frac{2}{\alpha}(x-t)I\{t<x\} + \frac{2}{\alpha}(t-y)I\{t>y\}\right)$$

where $\alpha = 1 - p$.

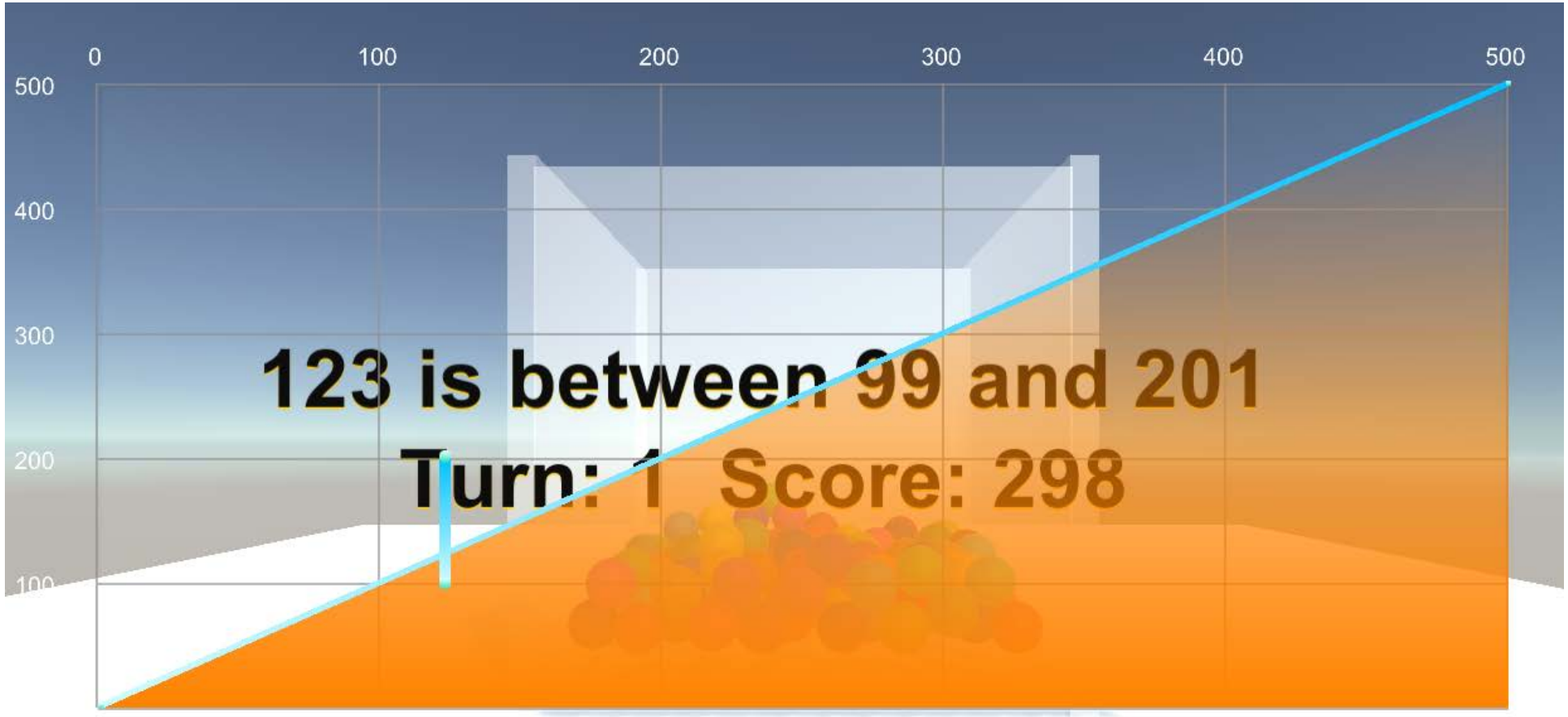


Bean Guesser - Probability interval to express uncertainty in a number



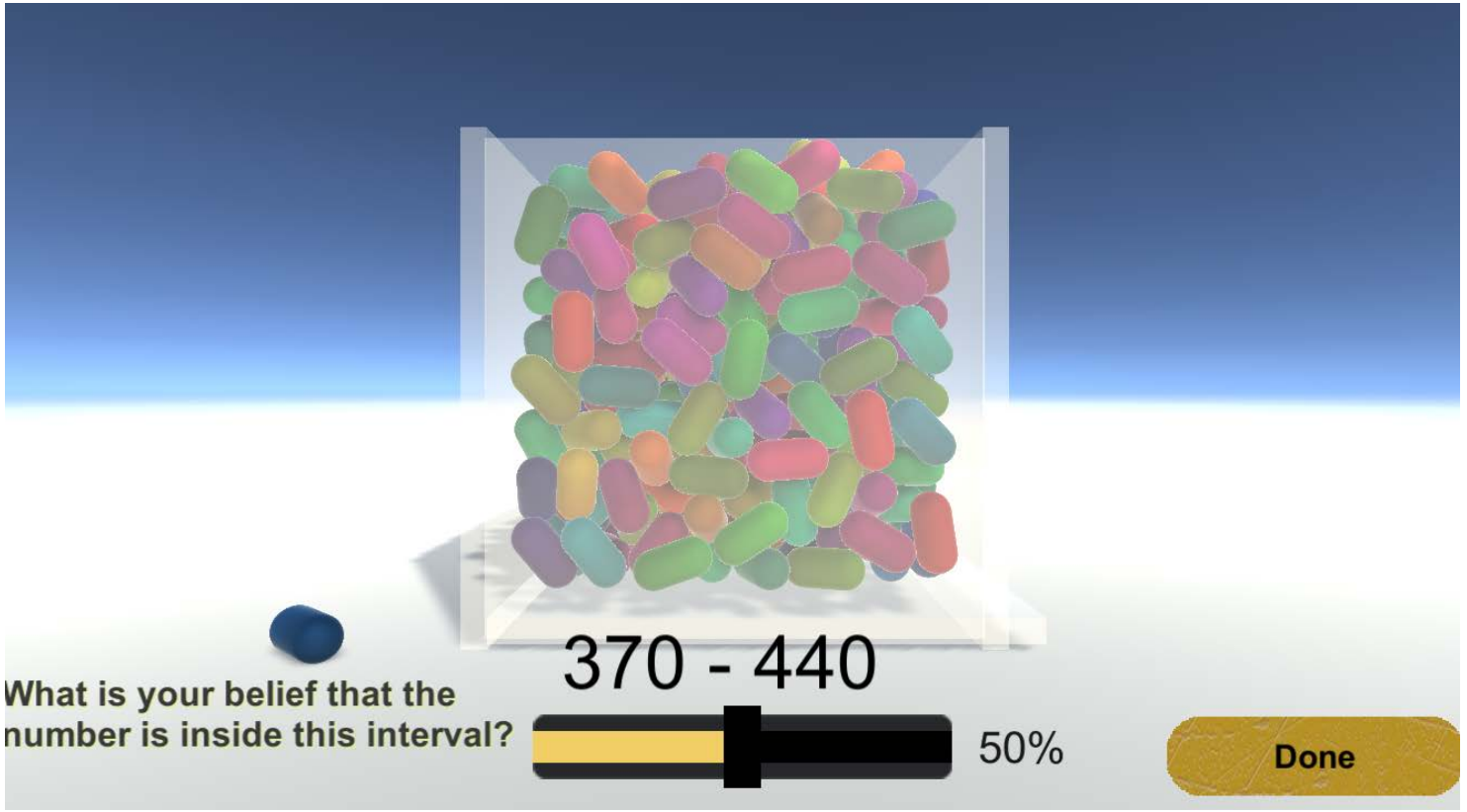
Bean Guesser

Feedback – calibration curve

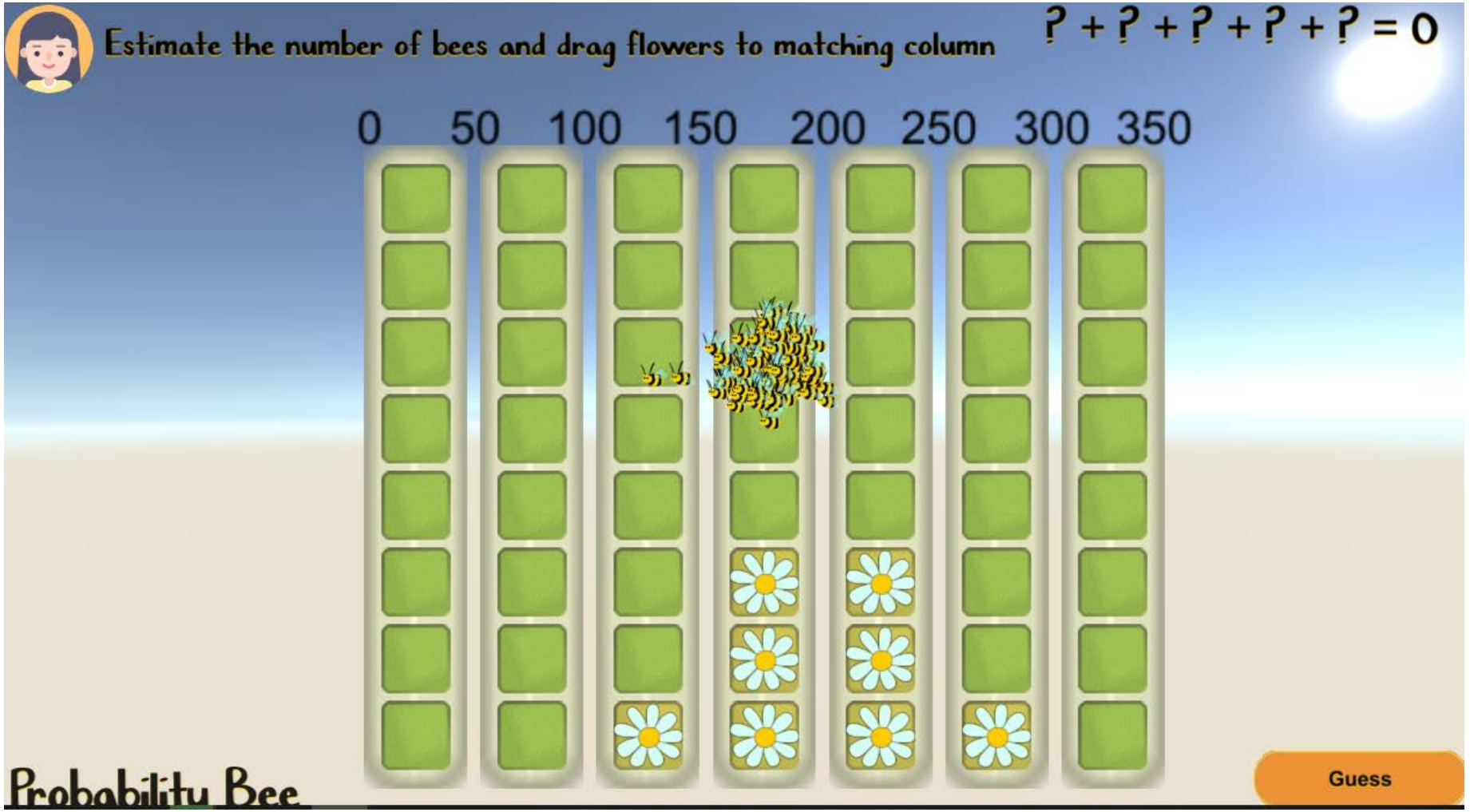


Probability Guess

The interval is given and one is express uncertainty that the number falls inside the interval

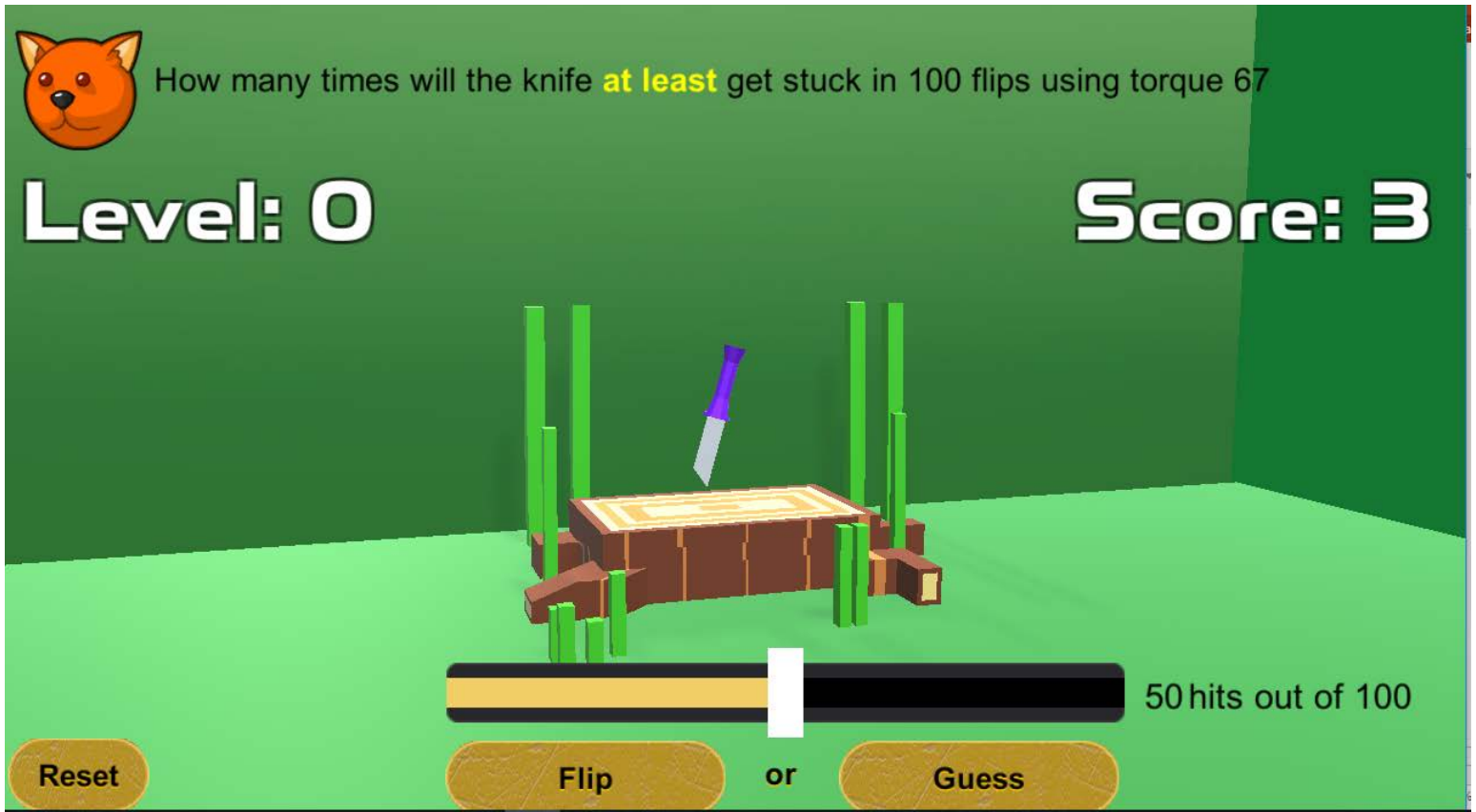


Probability Bee



Frequency Guesser – lower bound on a frequency

- Uncertainty in a frequency
- Ask for a lower bound on ”k” in the natural frequency ”k out of N”

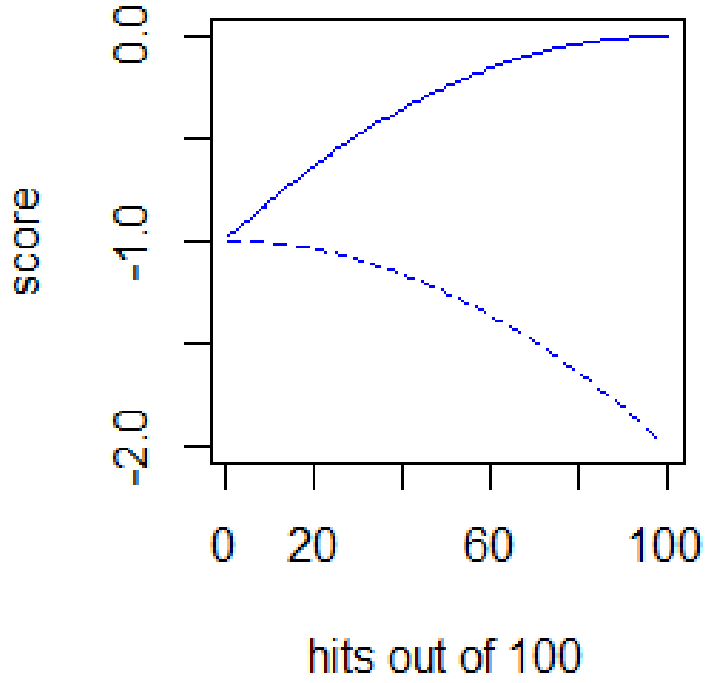


Frequency Guesser – lower bound on a frequency

- Uncertainty in a frequency
- Ask for a lower bound on ”k” in the natural frequency ”k out of N”

- Imprecise Brier score (Seidenfeld et al 2012)
- Lower bound:

$$g = -(1-x)^2 I\{p \geq x\} - (1+x^2) I\{p < x\}$$



What's next?

- Verify that the games are valid and reliable measuring instruments peoples ability to express their (epistemic) uncertainty
- Measure ”uncertaincy” in forthcoming studies
- Test if ”uncertaincy” exists
- If it does, test if ”uncertaincy” can improve our understanding of human behaviour and cognitive abilites
- Develop ways to help us improve our ”uncertaincy”

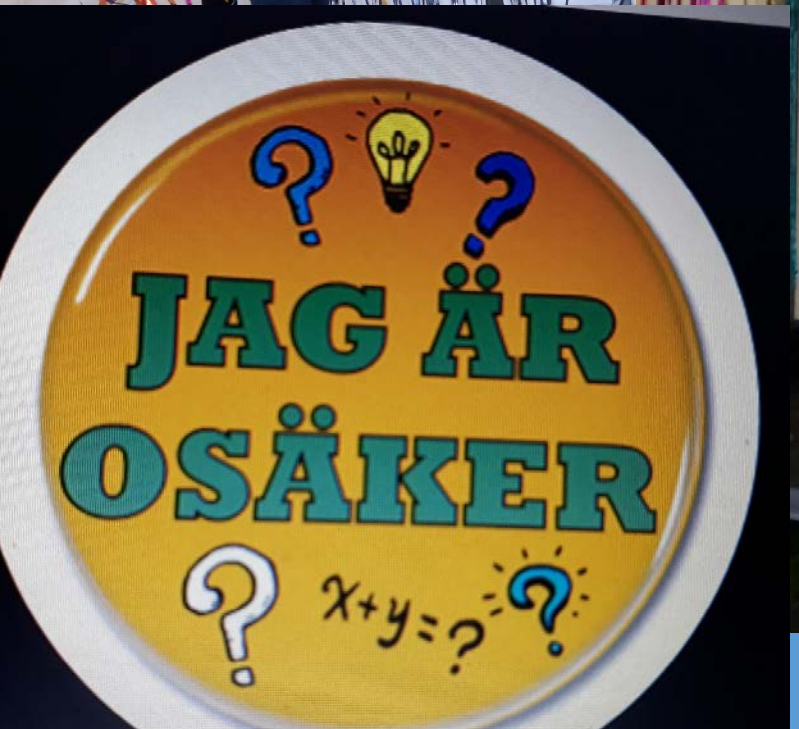
FORMAS

eit Climate-KIC

Den Nya Verkligheten

OSÄKERHETSSHOWEN

Ullrika Sahlin & Rasmus Bå



Spiegelhalter's tweet
https://twitter.com/EFSA_EU/status/1042123576713900033

There is no single 'true' uncertainty

It is my, yours or our uncertainty

Don't apologize for being uncertain

Just because you don't know everything does not mean you don't know anything

Reclaim uncertainty – be confident about your uncertainty!

References

- Gage, J., and D. J. Spiegelhalter. 2016. *Teaching probability*, *Cambridge Mathematics*. Cambridge: Cambridge University Press.
- Gneiting, T., and A. E. Raftery. 2007. "Strictly proper scoring rules, prediction, and estimation." *Journal of the American Statistical Association* 102 (477):359-378. doi: 10.1198/016214506000001437.
- Seidenfeld, T., M. J. Schervish, and J. B. Kadane. 2012. "Forecasting with imprecise probabilities." *International Journal of Approximate Reasoning* 53 (8):1248-1261. doi: 10.1016/j.ijar.2012.06.018.