The Centre for the Evaluation of Complexity Across the Nexus

DIAGNOSTIC EVALUATION WITH SIMULATED PROBABILITIES

cecan

Centre for the Evaluation of Complexity Across the Nexus

CECAN Webinar:

Diagnostic Evaluation with Simulated Probabilities

Tuesday 27th April 2021, 13:00 – 14:00 BST

Presenters: Barbara Befani & Corinna Elsenbroich

Welcome to our **CECAN Webinar.**

All participants are muted. Only the Presenters can speak. The webinar will start at **13:00 BST.**

Barbara and Corinna will speak for around 45 minutes and will answer questions at the end.

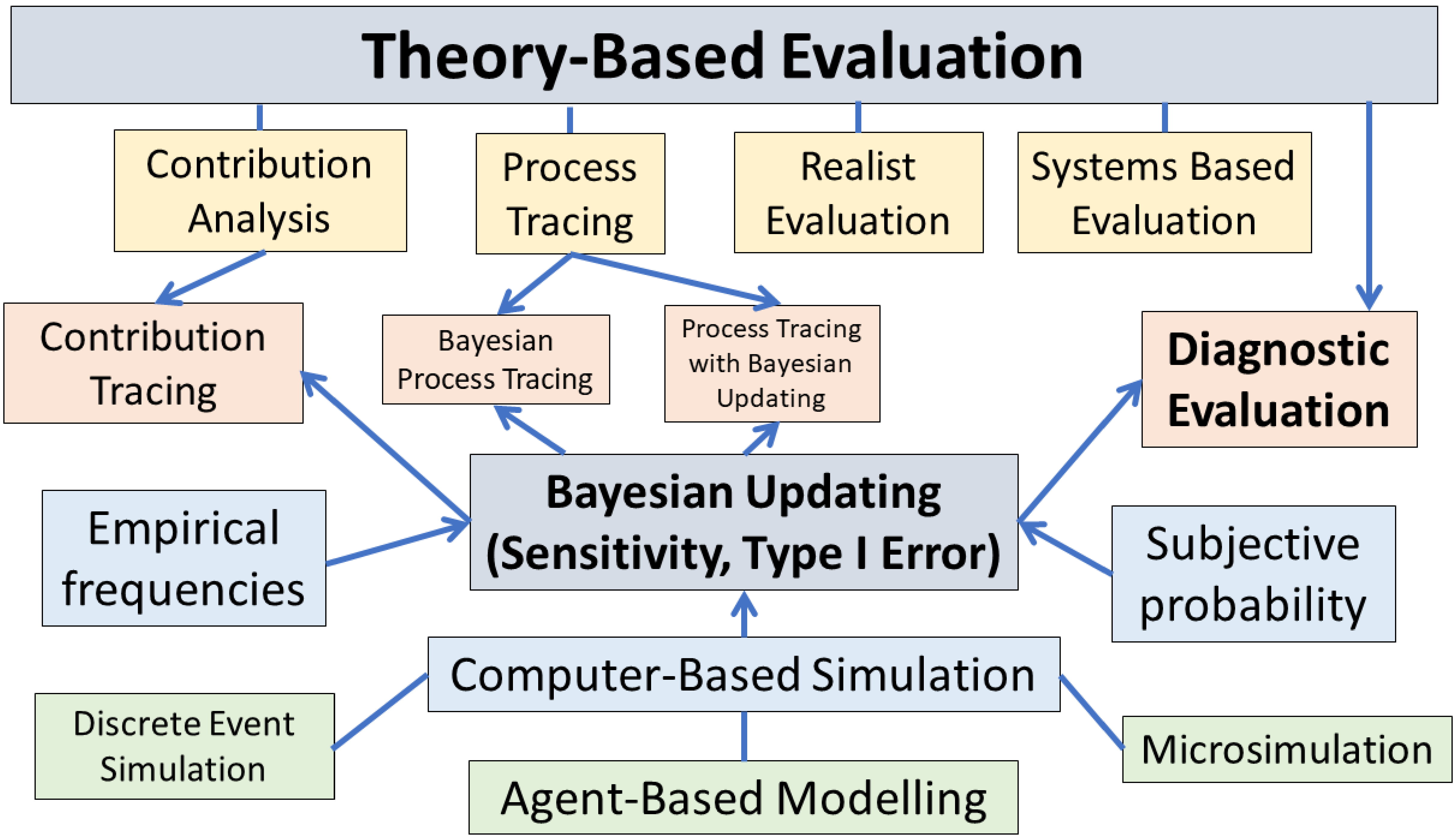
Please submit your questions at any point during the webinar via the question box in the Zoom webinar control panel.

Today's webinar will be recorded and made available on the CECAN website.

E Mail: cecan@surrey.ac.uk Web: www.cecan.ac.uk

www.facebook.com/CECANEXUS Twitter: @cecanexus

MOTIVATION FOR THE PAPER



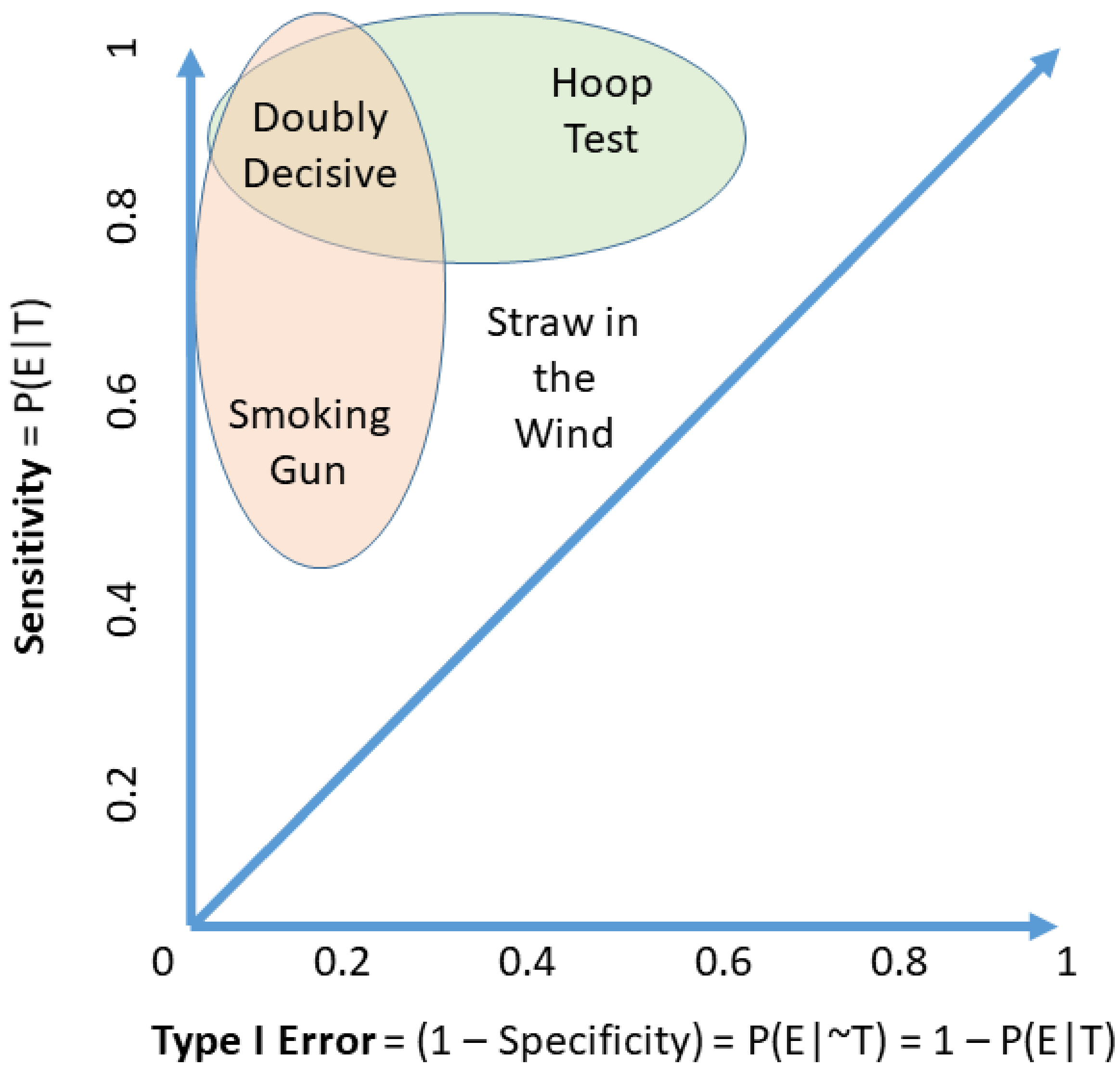
“DISCIPLINING” THEORY-BASED EVALUATION

- In Contribution Analysis, Realist Evaluation, some forms of Systems-Based evaluation:
- Often a **loose connection** between theory and data
- Lack of transparency on the dialogue between theory and evidence
- Particularly on how empirical observations change the theory
- In Process Tracing there’s a formal assessment of the weight of evidence for a certain theory
- *Smoking Gun, Hoop Test, Doubly-Decisive, Straw-in-the-Wind*
- But it’s rudimentary for some aspects; what is assessed:
- Direction: strengthening, weakening
- **Probative value**: strong or weak (**binary**, not fine-grained)
- Advocate the adoption of a formal Bayesian approach grounded on the Confusion Matrix

THE ~~CLARITY~~ CONFUSION MATRIX

		Theory (ontological reality)		
		The proposition / statement / theory is TRUE	**The proposition / statement / theory is FALSE**	
Empirical observation O leading us to believe that the proposition / statement / theory is true (observable reality)	**Evidence (O) is OBSERVED**	True Positive (TP)	False Positive (FP)	**Positive Predictive Value** = TP / (TP + FP)
	Evidence (O) is NOT OBSERVED	False Negative (FN)	True Negative (TN)	False omission rate = FN / (FN + TN)
		True positives rate (TPR) = **Sensitivity** = 1 – Type II error = TP / (TP + FN)	False positives rate (FPR) = 1 – Specificity = **Type I error** = FP / (FP + TN)	**Likelihood ratio** = TPR / FPR = Sensitivity / Type I error
		False negatives rate (FNR) = **Type II error** = 1 – Sensitivity = FN / (TP + FN)	True negatives rate (TNR) = **Specificity** = 1 – Type I error = TN / (FP + TN)	

THE BENEFITS OF A FORMAL BAYESIAN APPROACH



- Reality is nuanced
- Instead of saying "conclusive" (for confirmation or disconfirmation)
- Measure Sensitivity and Specificity
- Any real number from 0 to 1
  - (doesn't have to, can also be a scale of qualitative confidence levels)
- Measure the power to strengthen or weaken the theory
- For each theory-observation combination

BAYESIAN UPDATING WANTS PROBABILITIES

- Not a strict requirement (you can work with qualitative ranges)
- But that's the "natural" way of working with the Bayes formula
- **How to estimate those probabilities?**
- Let's forget the prior for now (we can set it at 0.5 and assume ignorance)
- **Sensitivity**: Probability of making a specific observation E under the assumption that the theory is true – **P (E|T)**
- **Type I Error**: Probability of making a specific observation E under the assumption that the theory is NOT true – **P (E|~T)**

ESTIMATING BAYES FORMULA PROBABILITIES

- Traditionally, there are two strategies:
- **Empirical frequencies** (mostly not available in evaluation)
- **Subjective Probability** (elicitation of expert judgement / opinions)
- <u>Why not do it with computer simulation?</u>

- If we manage to set computer models so that they represent different theoretical assumptions, we can run them until they produce estimates of quantities we are supposed to empirically observe

- Then, once we observe the quantity in reality, we can “reason backward” and identify the model settings that are most likely to produce that (which is supposed to represent the “real” theoretical parameters)

GENERATING PROBABILITIES

- We "know" the probability of throwing a 6 with a die as 1/6 because it has 6 sides.
- However, if the die is weighted the assessment is wrong.
- Then we have to throw again and again and again to "learn" the probability.

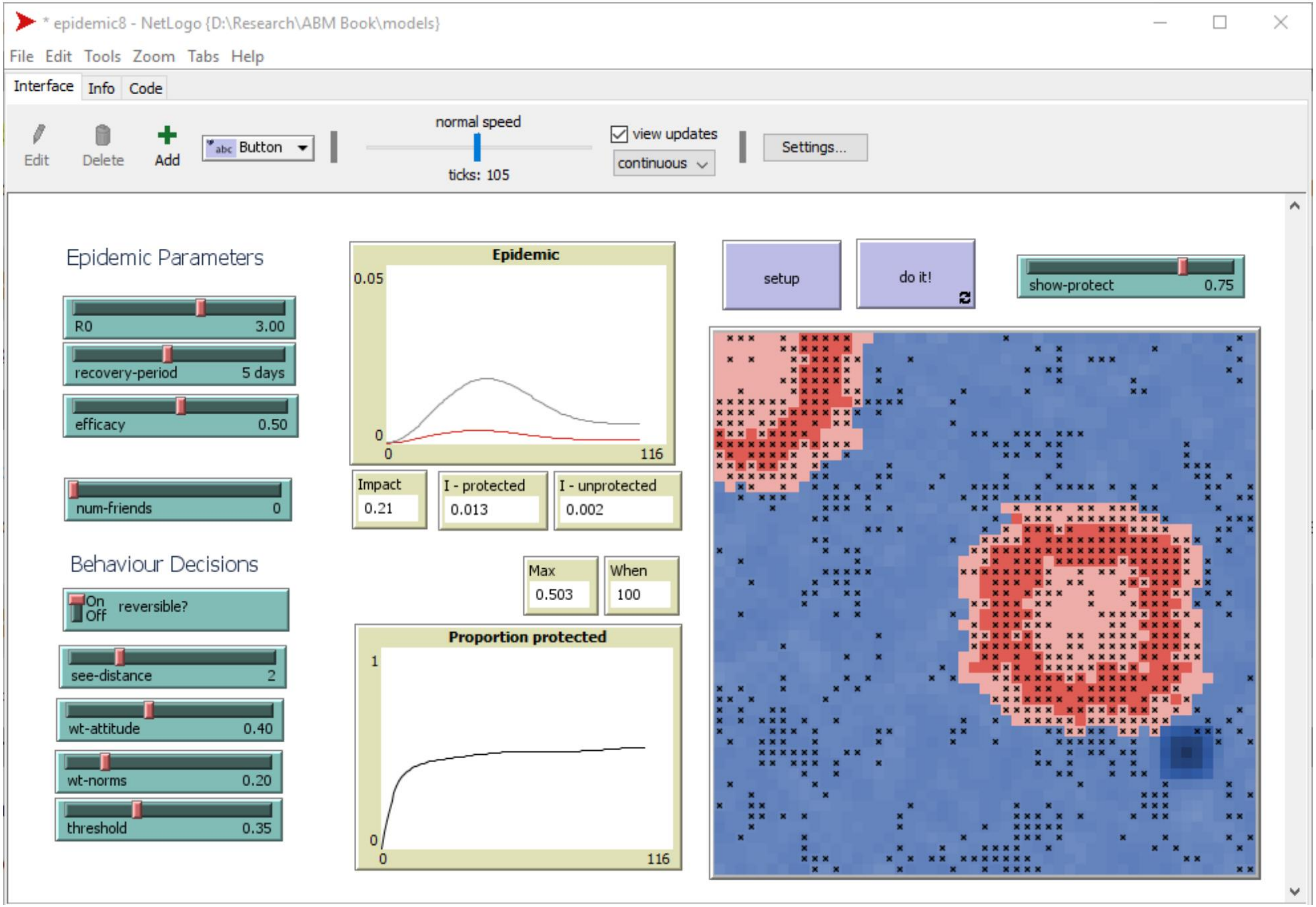
In the real world we constantly have to **learn** - but often cannot repeat.

AGENT-BASED MODELLING

Agent-based modelling is a computational method that enables a researcher to create, analyse, and experiment with models composed of agents that interact within an environment.

Gilbert 2008, pg 2

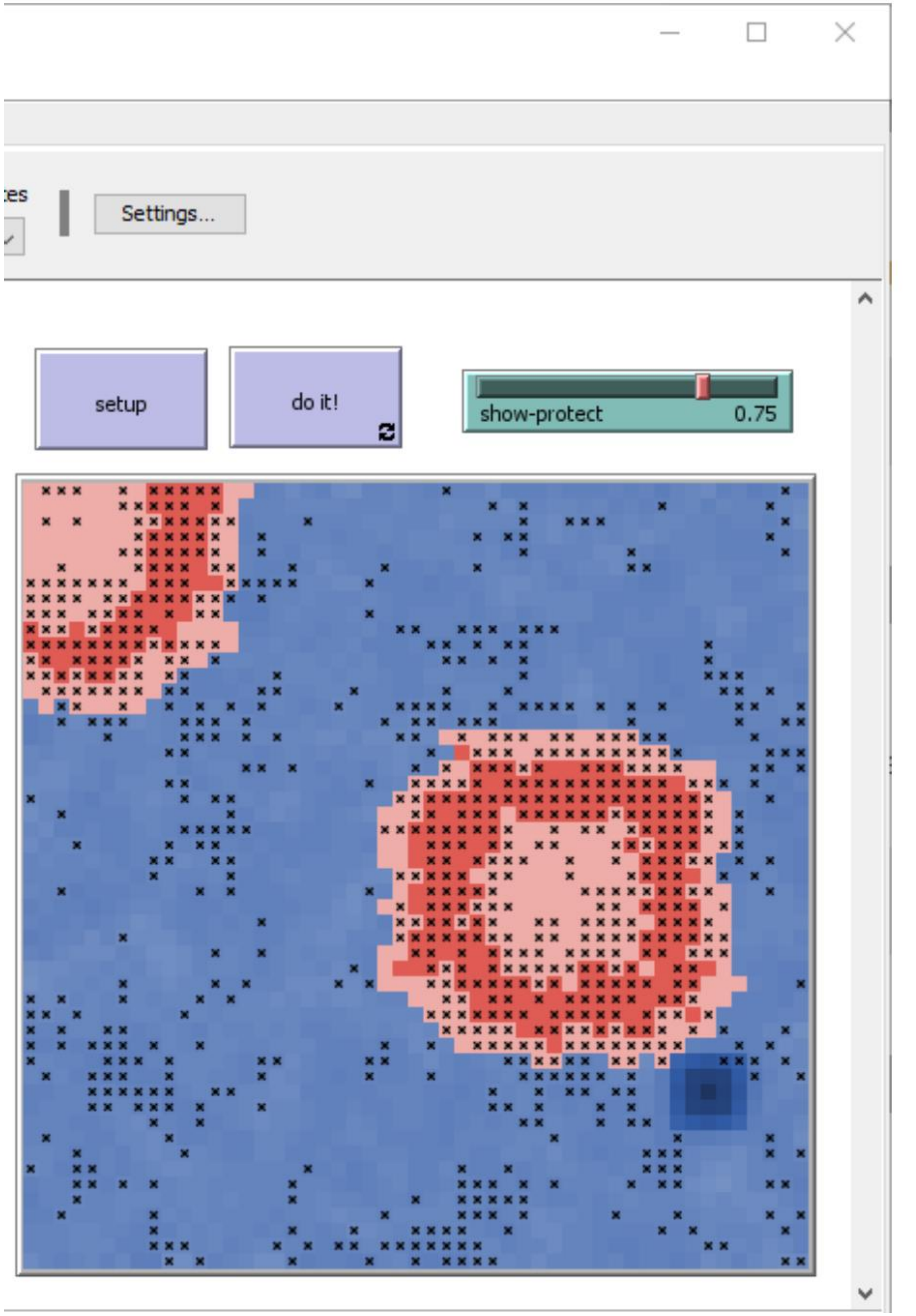
THE TELL ME MODEL



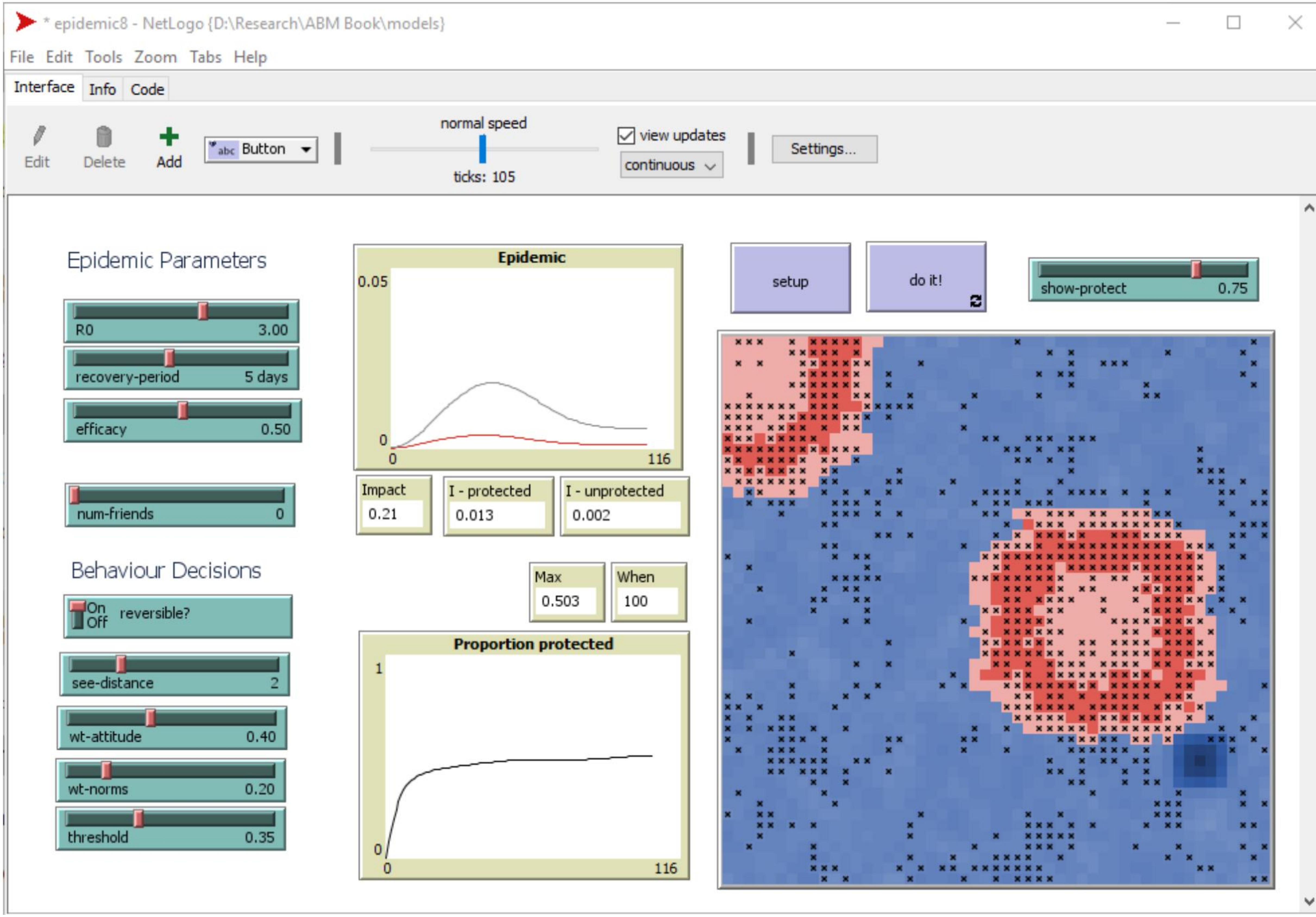
https://research.criticalconnections.com.au/ABMBook/ABMtutorial.pdf

THE TELL ME MODEL

- Epidemiology model
- Agents react to levels of perceived risk of being infected
- They assess risk through
  - Others they know being infected
  - How far away the infection is
- They adjust their behaviour according to
  - The perceived risk
  - The social norms that surround them



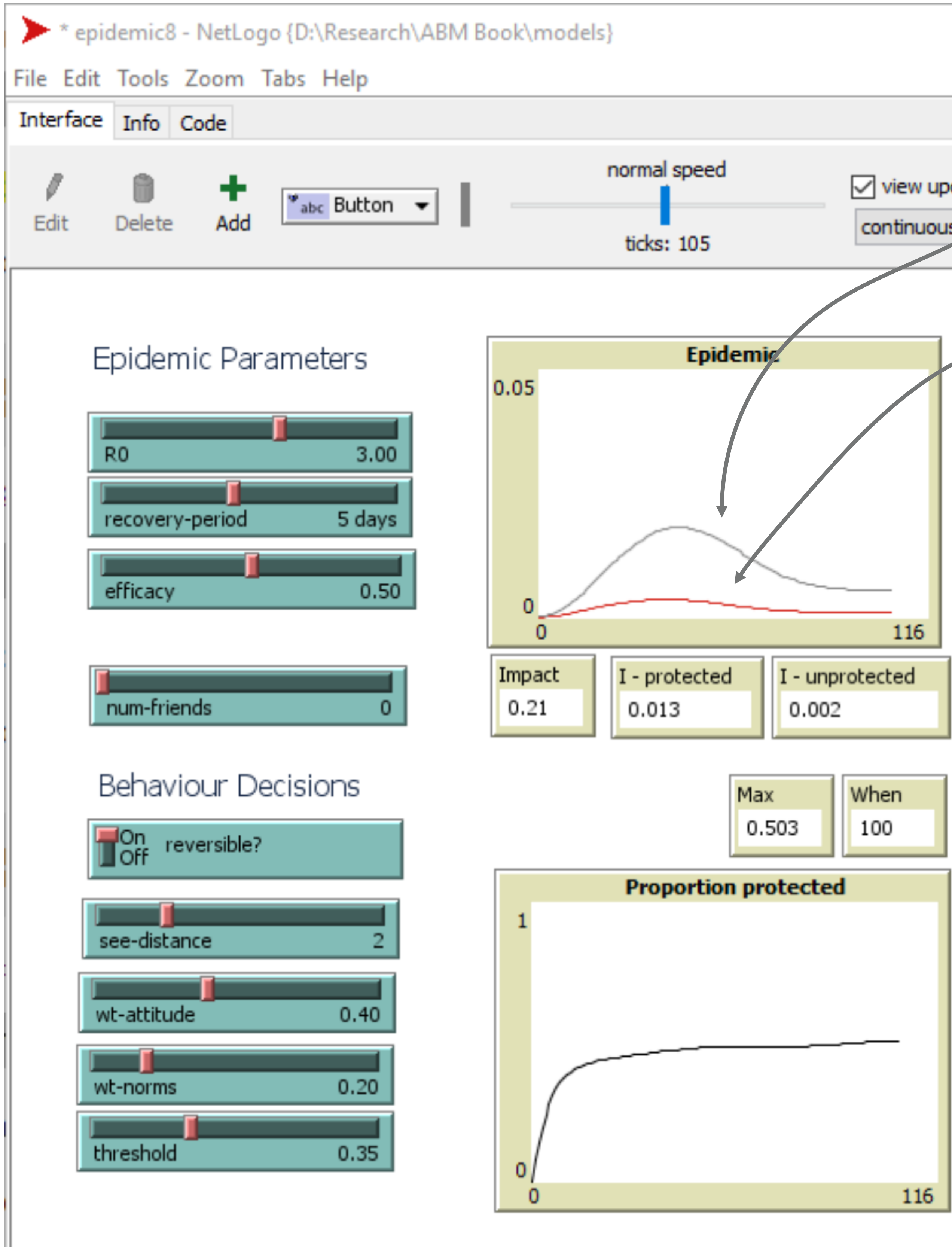
THE TELL ME MODEL



https://research.criticalconnections.com.au/ABMBook/ABMtutorial.pdf

THE TELL ME MODEL

https://research.criticalconnections.com.au/ABMBook/ABMtutorial.pdf



How many get sick?

How many die?

- ➤ Epidemic description
- ➤ Social Networks
- ➤ Protective behaviours
  - ➤ can be **reversible**, e.g. hand washing, masks etc
  - ➤ or **irreversible** e.g. vaccination.

THE TELL ME MODEL

https://research.criticalconnections.com.au/ABMBook/ABMtutorial.pdf

If E is the efficacy of the protective behaviour, P_r the protected proportion of people, and q the susceptible fraction becoming infected, the probability of a susceptible person becoming infected is given by:

$$P = \begin{cases} q & P_r = 0 \\ q\dfrac{1}{1 - P_r E} & \text{unprotected} \\ q\dfrac{1 - E}{1 - P_r E} & \text{protected} \end{cases}$$

- We use the model over a set of parameter variations for efficacy of protective behaviours
- The output data is a set of frequencies of particular outcomes, in relation to input settings
- This dataset is now used to infer backwards, about the likelihood of which kind of setting we are in in the "real world"

ABM AND THE REAL WORLD

Real World Evaluation of Epidemic Response

30% of the population were infected.

ABM AND THE REAL WORLD

Real World Evaluation of Epidemic Response

30% of the population were infected.

ABM of Epidemiological Behaviour

ABM AND THE REAL WORLD

Real World Evaluation of Epidemic Response

30% of the population were infected.

ABM of Epidemiological Behaviour

Protective efficacy high

Protective efficacy medium

Protective efficacy low

ABM AND THE REAL WORLD

Real World Evaluation of Epidemic Response

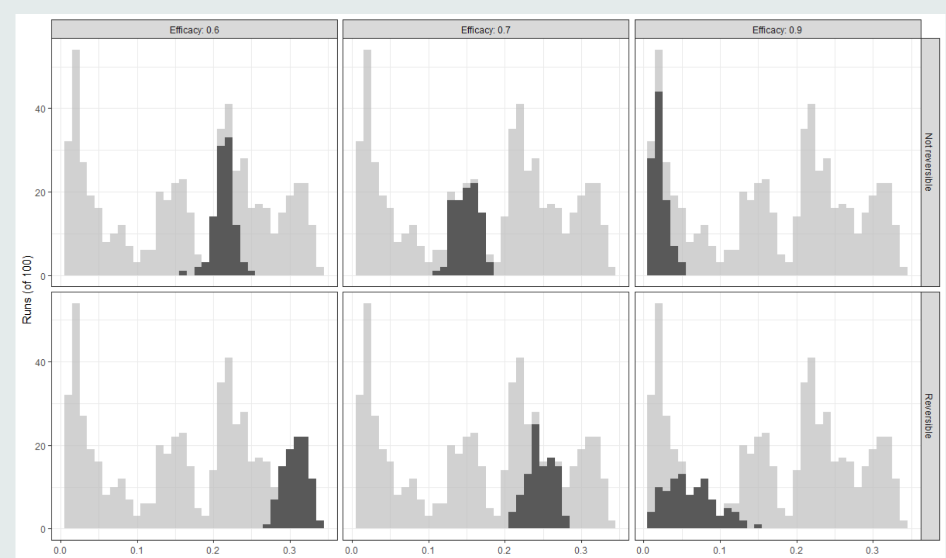
30% of the population were infected.

ABM of Epidemiological Behaviour

Protective efficacy high

Protective efficacy medium

Protective efficacy low



ABM AND THE REAL WORLD

Real World Evaluation of Epidemic Response

30% of the population were infected.

ABM of Epidemiological Behaviour

Protective efficacy high	T1 low and Sensitivity low
Protective efficacy medium	T1 medium and Sensitivity low
Protective efficacy low	T1 and Sensitivity for 30% is high

ABM AND THE REAL WORLD

Real World Evaluation of Epidemic Response

Good corroboration that protective behaviour has medium efficacy.

30% of the population were infected.

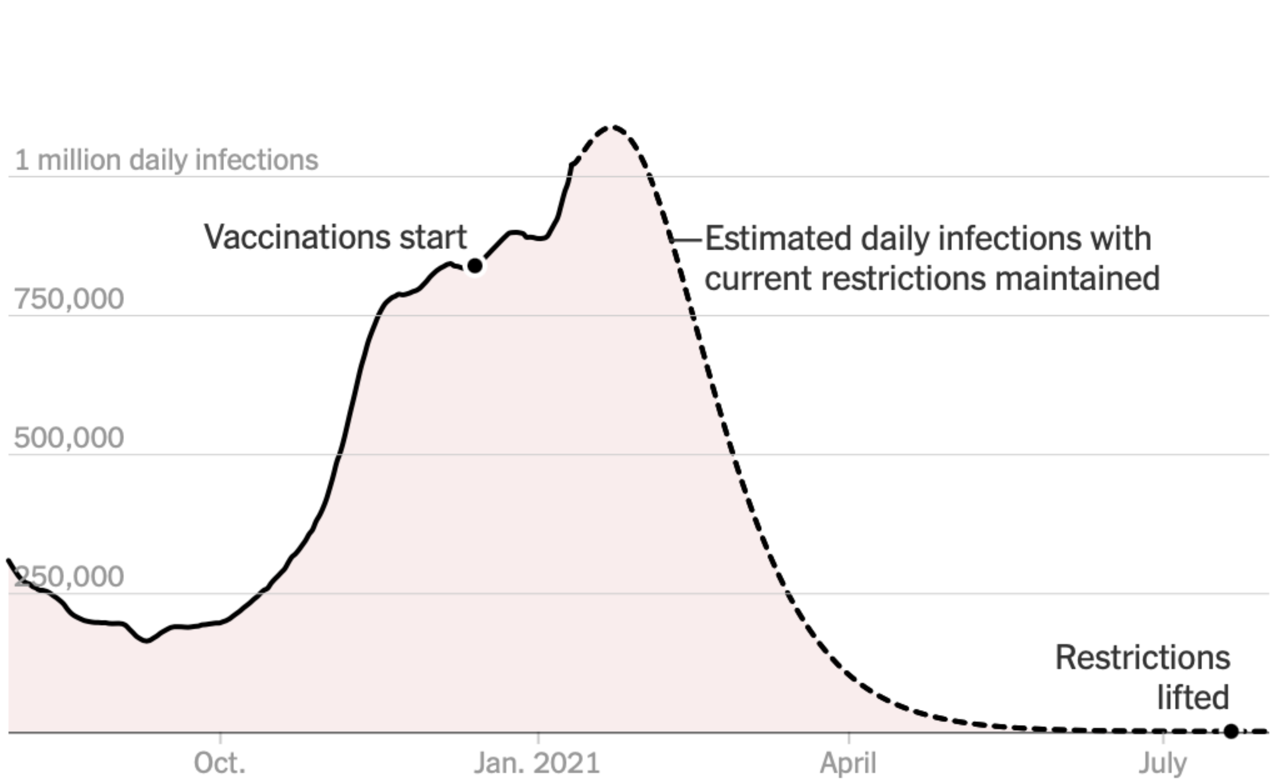
ABM of Epidemiological Behaviour

Protective efficacy high	T1 low and Sensitivity high
Protective efficacy medium	T1 high and Sensitivity med-high
Protective efficacy low	T1 low and Sensitivity high

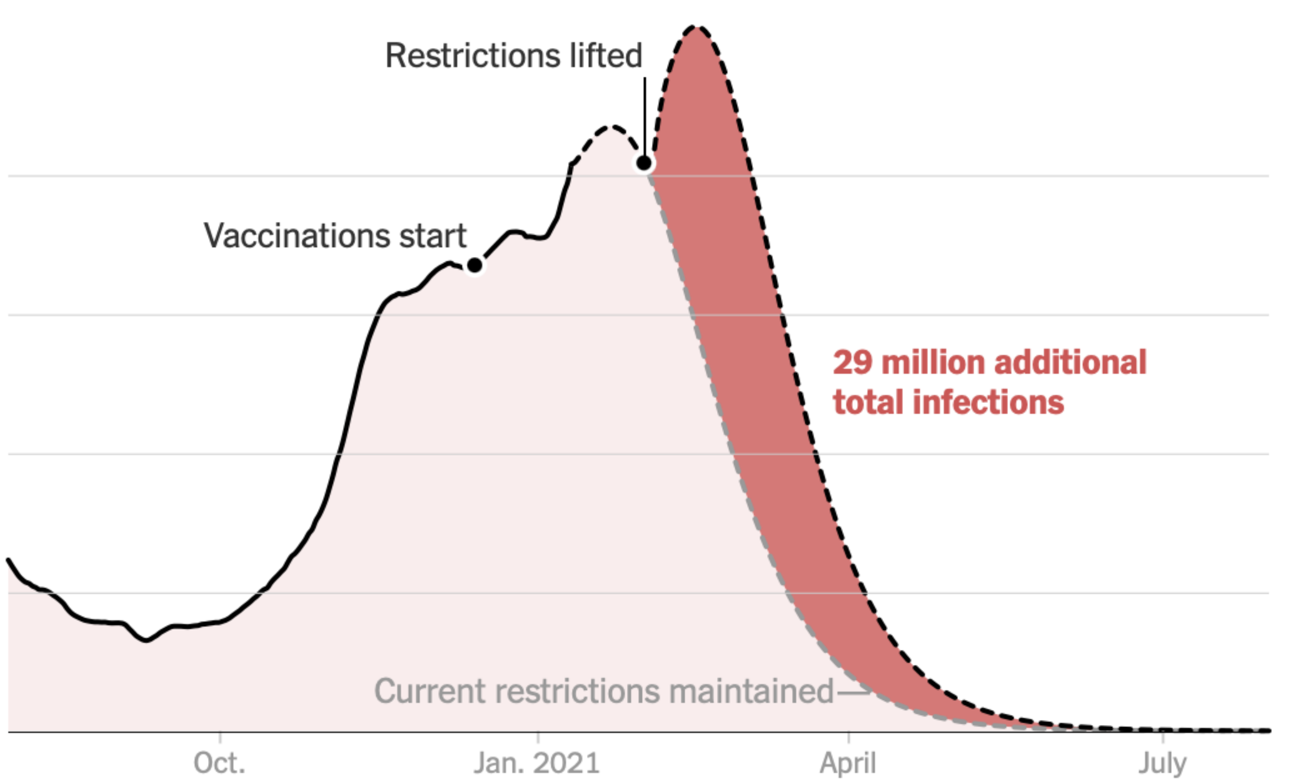
IS THIS REASONABLE?

- Vaccine efficacy 70-90%
- Social distancing, masks, no large gatherings, schools closed, etc. efficacy of x?
- We constantly reason forwards and backwards to understand the efficacy of certain things. That is in the end what lies at the heart of the step by step opening of the UK roadmap.

If current restrictions remain in place until late July

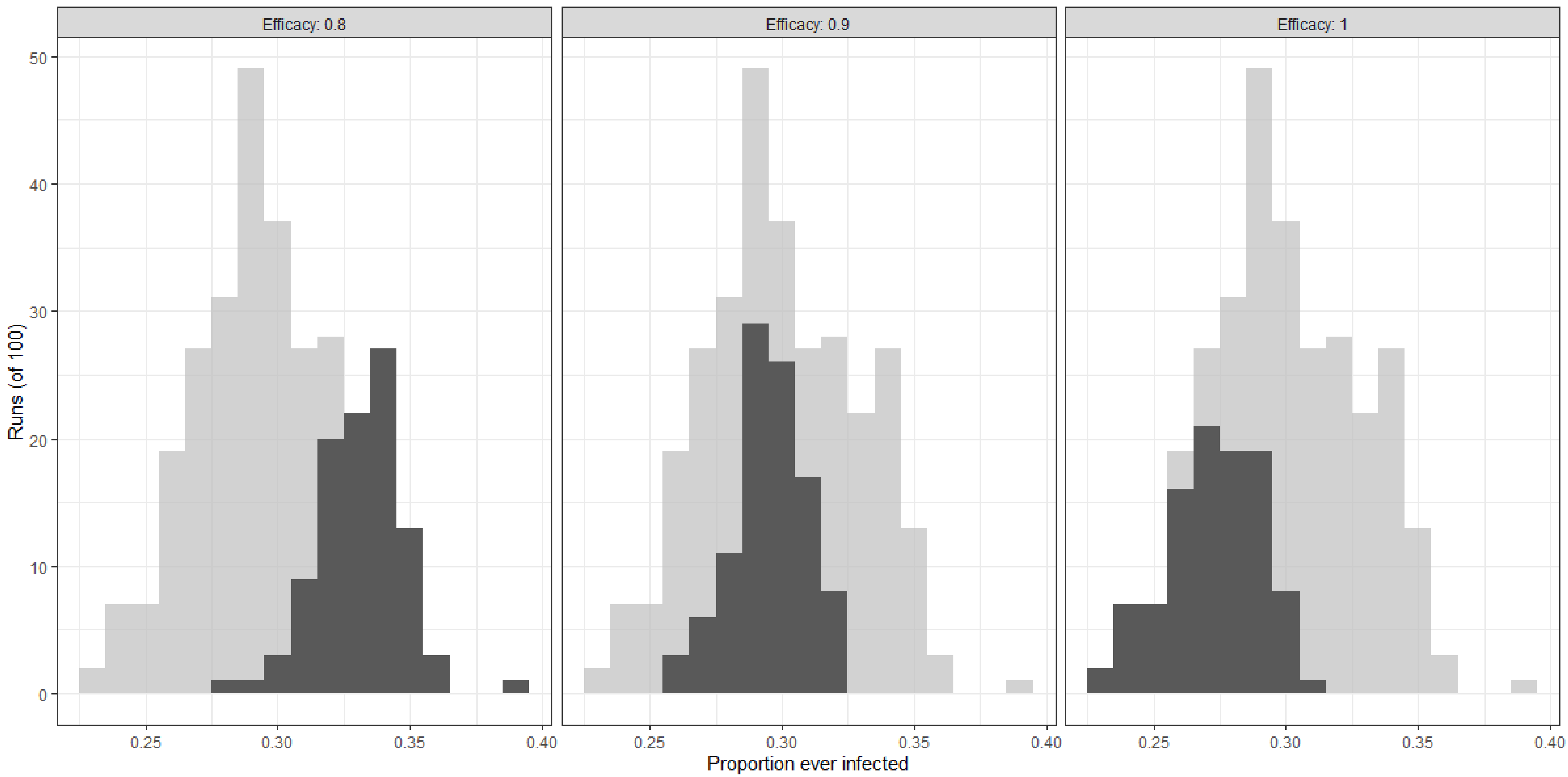


If restrictions are lifted in February



https://www.nytimes.com/interactive/2021/01/24/us/covid-vaccine-rollout.htm

DISTRIBUTION OF RESULTS UNDER THE THREE EFFICACY ASSUMPTIONS



PROBABILITIES OF OBSERVING GIVEN PROPORTIONS OF INFECTED POPULATIONS BY LEVELS OF EFFICACY

	Proportion of population ever infected				
Level of efficacy	<=0.275	>0.275 & <=0.29	>0.29 & <=0.30	>0.30 & <=0.31	>0.31
Ideal 1.0	0.53	0.30	0.13	0.04	0
Improved 0.9	0.09	0.27	0.26	0.24	0.14
Standard 0.8	0	0.02	0.01	0.03	0.94

- We tried to “cut” the space of empirical possibilities (for prop of IP) into areas that would be good at predicting efficacy levels
- We first tried five intervals that we later assembled into three

EXTRACTING THE POSTERIORS FOR BAYESIAN UPDATING WITH THE PRIORS ALL SET AT 0.33 AND A NARROW CENTRAL INTERVAL

Level of efficacy	Posteriors after observation of evidence	Sensitivity	Type I Error	Likelihood Ratio	Posterior-Prior
Ideal 1.0 (prior = 0.33)	Infected population <= 0.29, posterior = 0.68	0.83	0.19	4.37	0.35
Improved 0.9 (prior = 0.33)	I.P. 0.29 < p <= 0.30, posterior = 0.65	0.26	0.07	3.71	0.32
Standard 0.8 (prior = 0.33)	I.P. > 0.30, posterior = 0.69	0.97	0.21	4.62	0.36

- Relatively unsatisfactory results
- The chosen intervals for the tests (prop of IP) weren’t very good at predicting efficacy
- The central interval was a Smoking Gun for “improved” because it was very narrow
- The right interval was a Hoop Test for “standard”

EXTRACTING THE POSTERIORS FOR BAYESIAN UPDATING WITH THE PRIORS ALL SET AT 0.33 AND A LARGER CENTRAL INTERVAL

Level of efficacy	Posteriors after observation of evidence	Sensitivity	Type I Error	Likelihood Ratio	Posterior-Prior
Ideal 1.0 (prior = 0.33)	I.P. <= 0.275, posterior = 0.84	0.53	0.05	10.60	0.51
Improved 0.9 (prior = 0.33)	I.P. 0.275 < p <= 0.31, posterior = 0.58	0.77	0.27	2.85	0.25
Standard 0.8 (prior = 0.33)	I.P. > 0.31, posterior = 0.87	0.94	0.07	13.43	0.54

- The likelihood ratios for the first and third intervals are much better than before
- The first is a smoking gun for "ideal"; the second is a "doubly decisive" for "standard"
- Perhaps these aren't very useful either because you could look at the charts and guess but our aim here was to present a proof of concept – the idea that you can estimate values for Bayesian TBE with computer-based simulation

PART OF A LARGER PROGRAMME

- ABM is time intensive
- Integrating ABM with other empirical methods to investigate small parts of reality
  - e.g. Castellani et al - QCA and ABM (https://www.springer.com/gp/book/9783319097336)
  - Integrating with Participatory Systems Mapping?