How Not to Make Decisions with Computer Models

Roman Frigg

www.romanfrigg.org

CPNSS Centre for Philosophy of Natural and Social Science

LSE

Arts & Humanities Research Council

MSU Managing Severe Uncertainty


philosophy
welcome to philosophy
purity
made simple
hope in a jar
philosophy: where there is hope there can
faith. where there is faith miracles can
miracle
worker
philosophy: miracles begin with a fresh perspective
miraculous anti-aging retinoid pads
disques rétinoïdes anti-âge miraculeux

In Collaboration With

Hailiang Du

CATS, LSE

Dept Stats, LSE

Seamus Bradley

Philosophy, LMU

Lenny Smith

CATS, LSE

Dept Stats, LSE

With Illustrations By

Fiorella Lavado

Independent Artist

www.fiorellalavado.com

“Laplace’s Demon and the Adventures of His Apprentices”, *Philosophy of Science*, January 2014

www.romanfrigg.org

Preview

The Question

At the most general level:

Models often aren’t exact replicas of their target systems.

How bad is it if out model is not an exact replica of the target system?

The Question

At the most general level:

Models often aren’t exact replicas of their target systems.

How bad is it if out model is not an exact replica of the target system?

→ It’s pretty bad.

The Question

A bit more specifically:

A dynamical model has **structural model error** (SME) if its time evolution is relevantly different from that of the target system, possibly due to simplifications and idealisations.

Question: what are the consequencs of SME for a model's predictive capacity?

Take-Home Message - Part 1

If **chaotic** models have even the slightest SME, their capacity to make meaningful forecasts is seriously compromised.

This has dramatic consequences for our ability to make the kind of forecasts about the future that policy makers would like to have.

Attention: *not* the same old story.

So far chaos has been studied in connection with uncertainty about ***initial conditions***.

We ask what happens if we are uncertain about the correct ***model structure****.*

These are completely different problems!

Butterfly effect:

Error in initial conditions

Butterfly effect:

Error in initial conditions

Hawkmoth Effect:

Error in the model structure (equations)

(Erica Thompson)

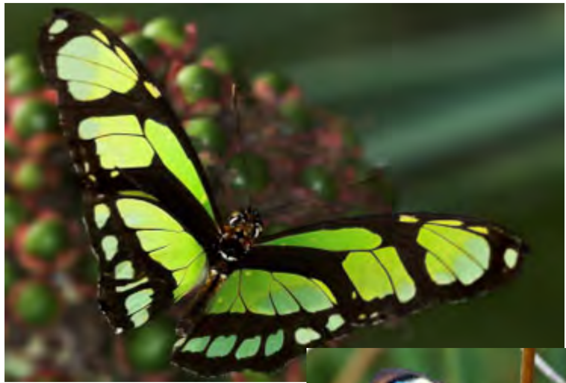
Take-Home Message – Part 2

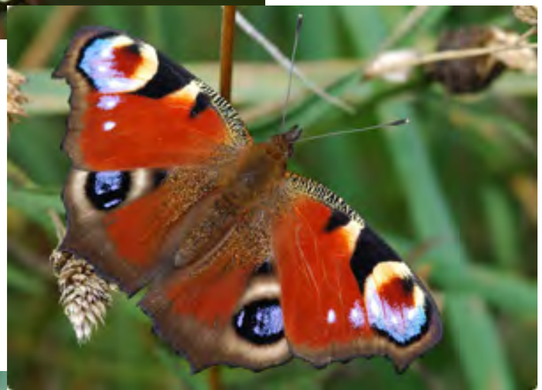
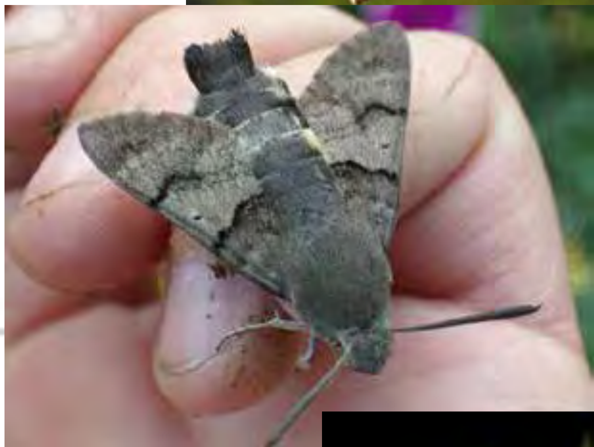
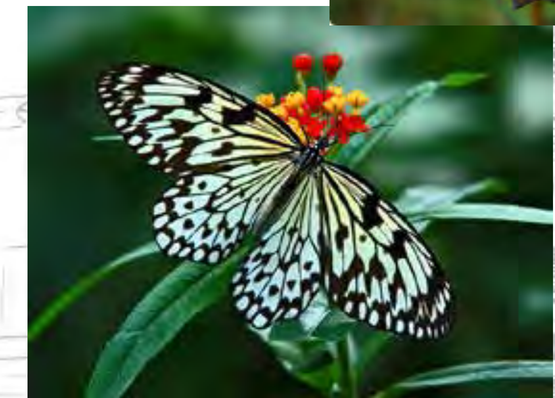
We can mitigate against the butterfly effect by making probabilistic predictions rather than point forecasts.

This route is foreclosed in the case of the hawkmoth effect: nothing can mitigate against that effect!

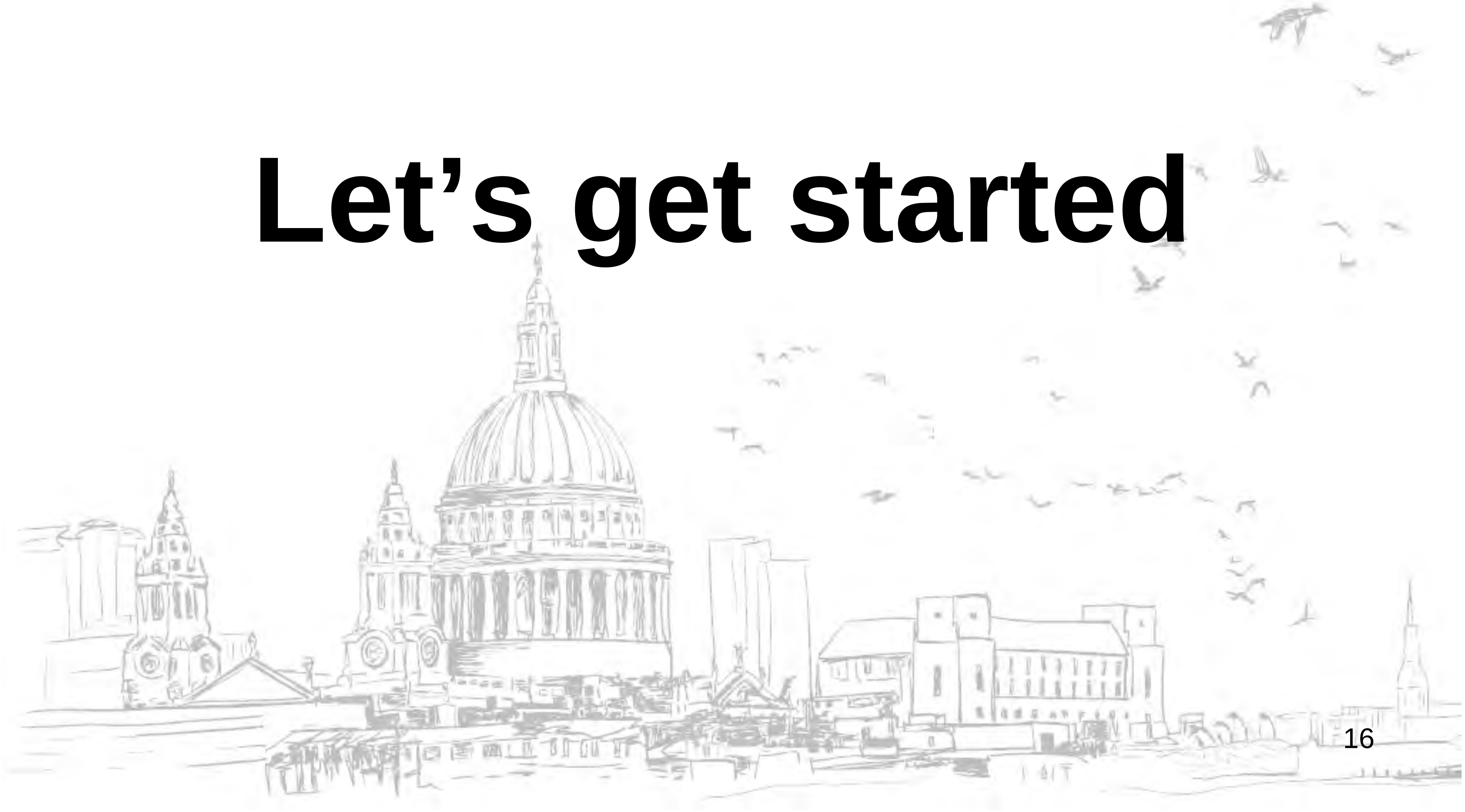
So structural model error and not uncertainty in the initial conditions is what truly limits predictive power.

Or: butterflies are pretty; hawkmoths are ugly.



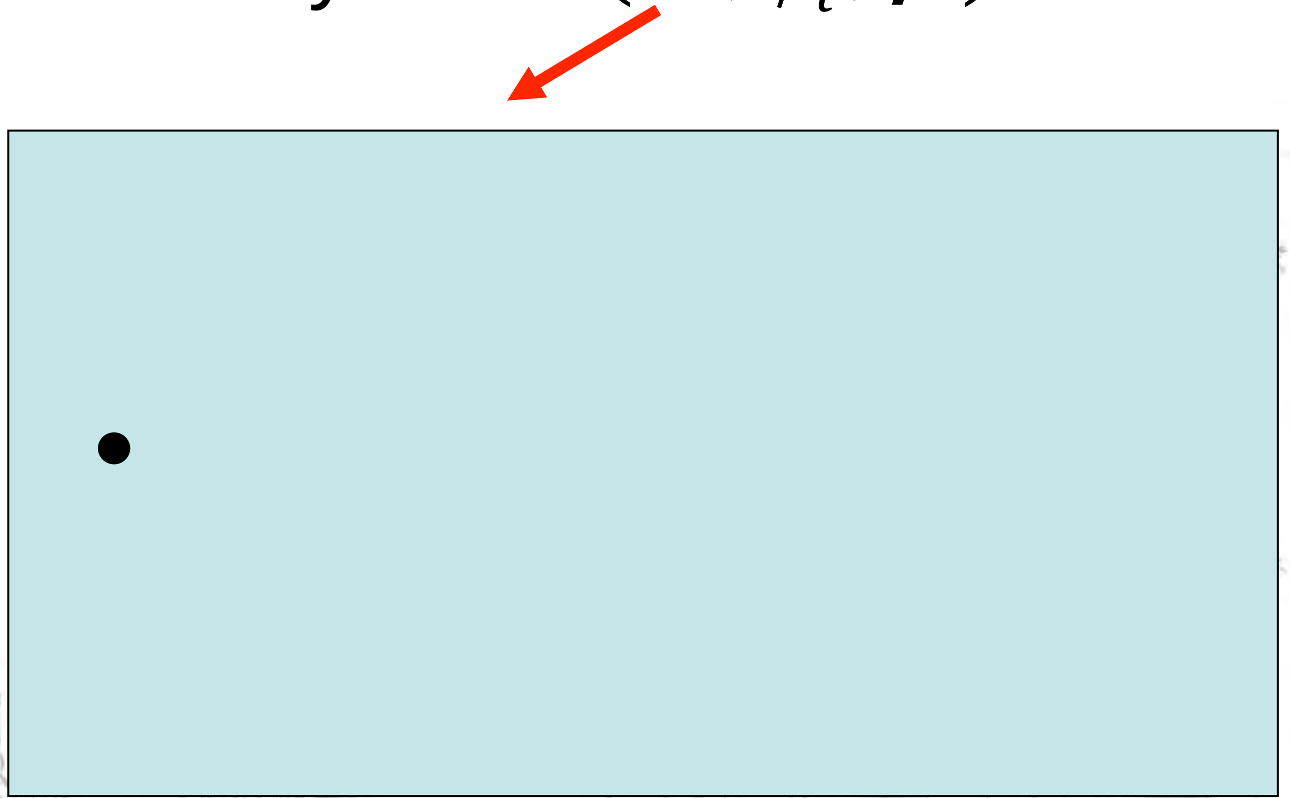





Let’s get started



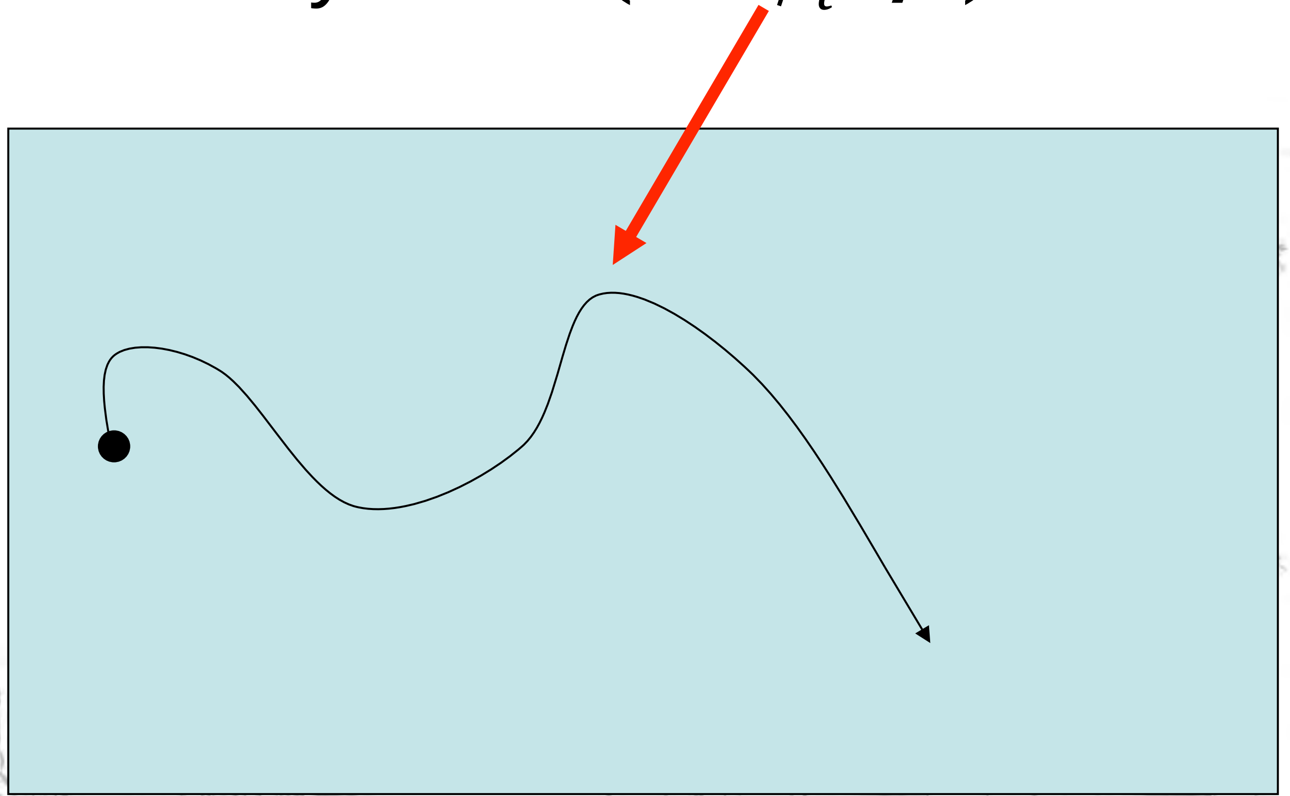
A Primer on Models

Dynamical system (X, ϕ_t, μ)



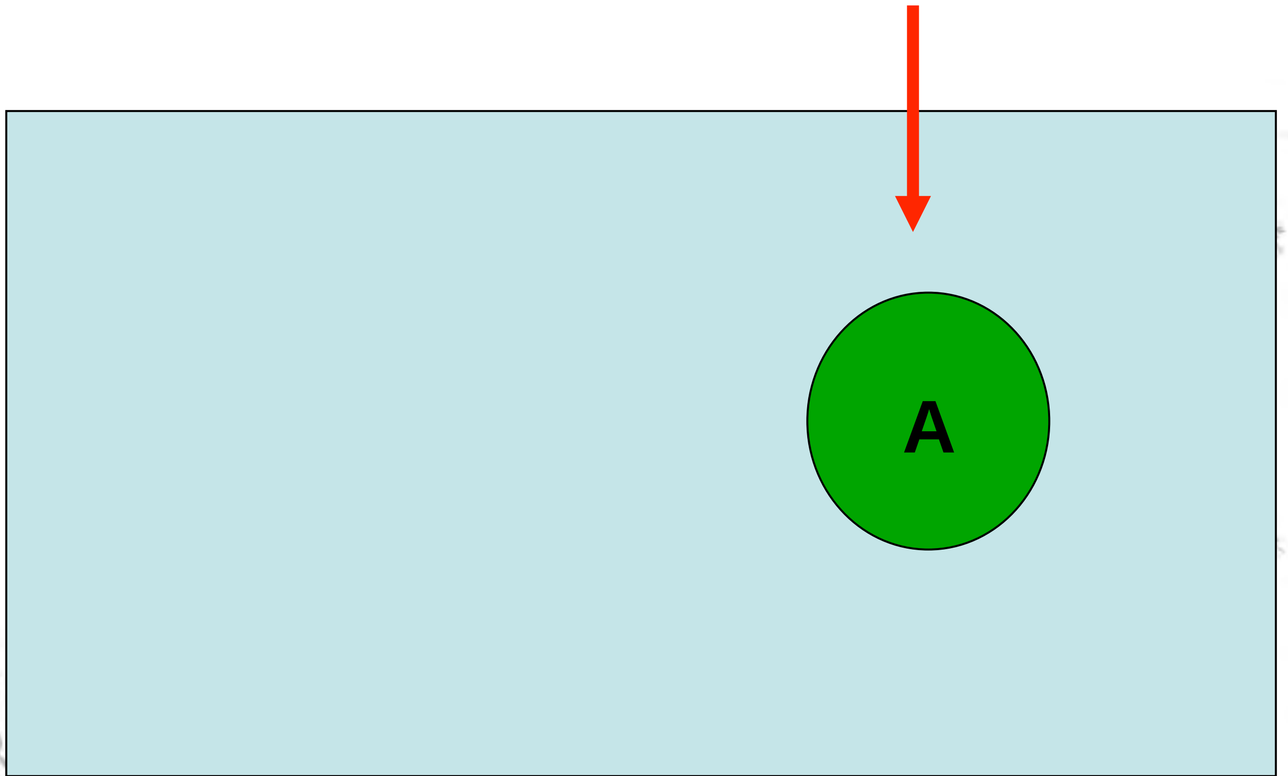
A Primer on Models

Dynamical system (X, ϕ_t, μ)



A Primer on Models

Dynamical system (X, ϕ_t, μ)



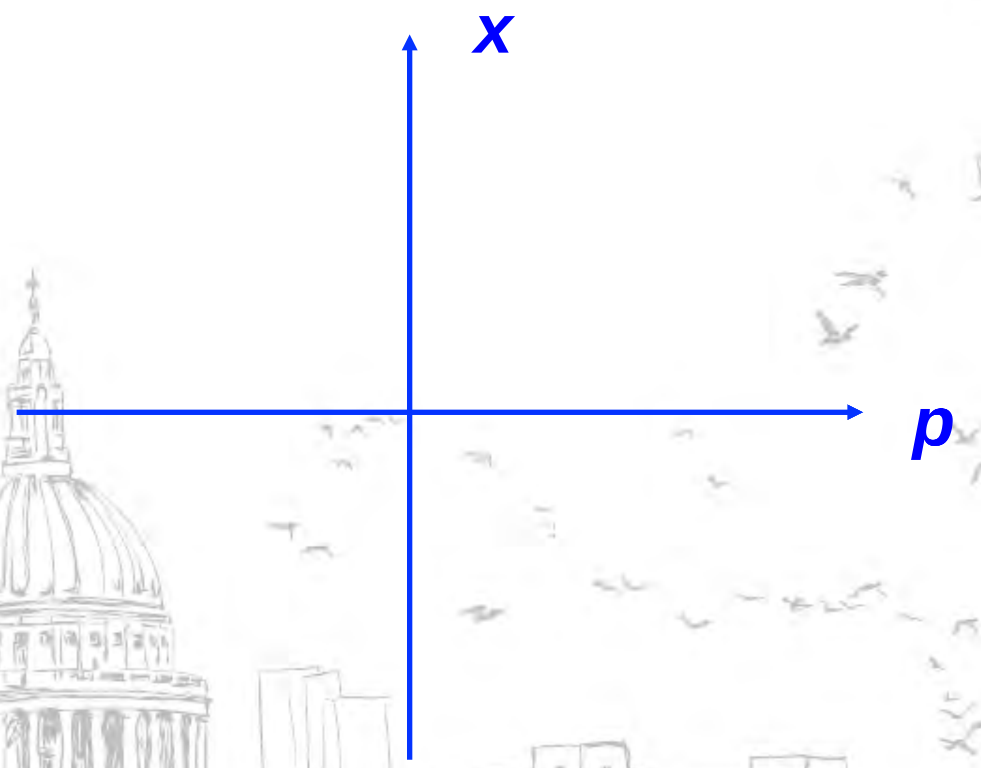
Simple example: stone falling from tower

Position ***x***

momentum ***p***

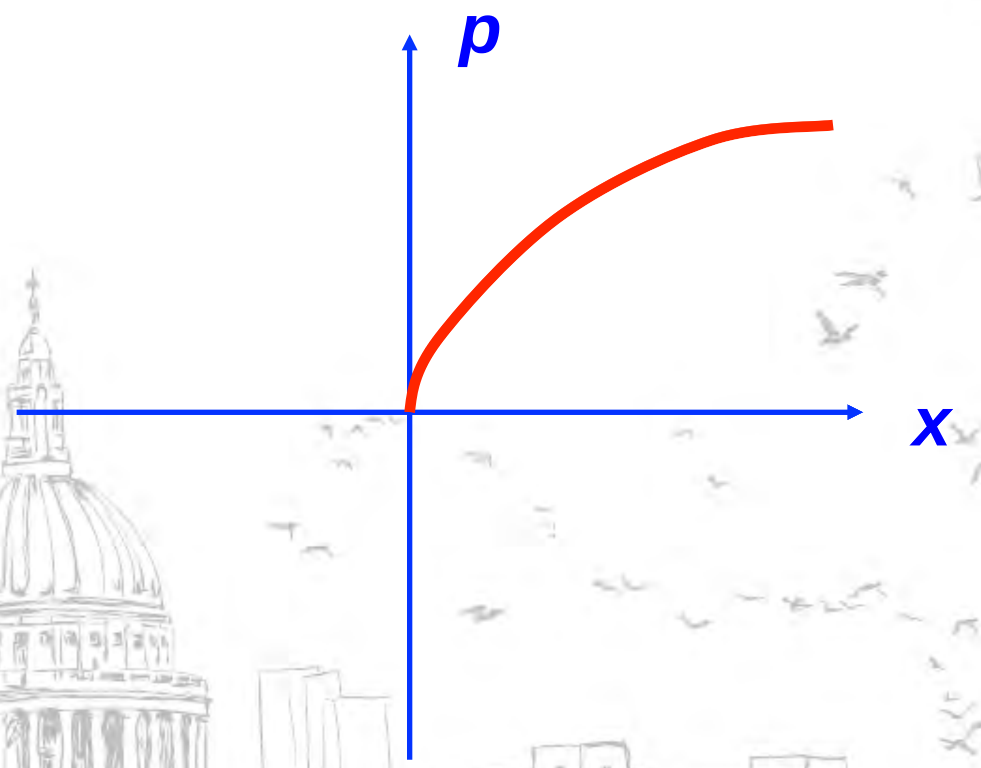
Simple example: stone falling from tower

$$(\underline{X}, \phi_t, \mu)$$



Simple example: stone falling from tower

$$(\underline{X}, \underline{\phi_t}, \mu)$$

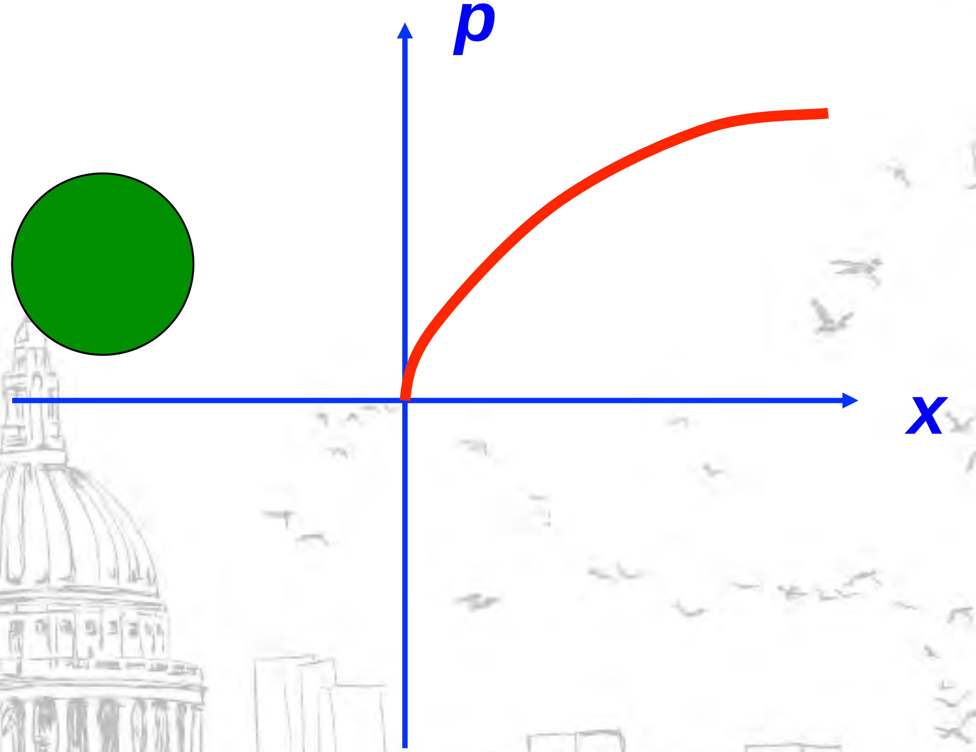


Simple example: stone falling from tower

$$(X, \phi_t, \mu)$$

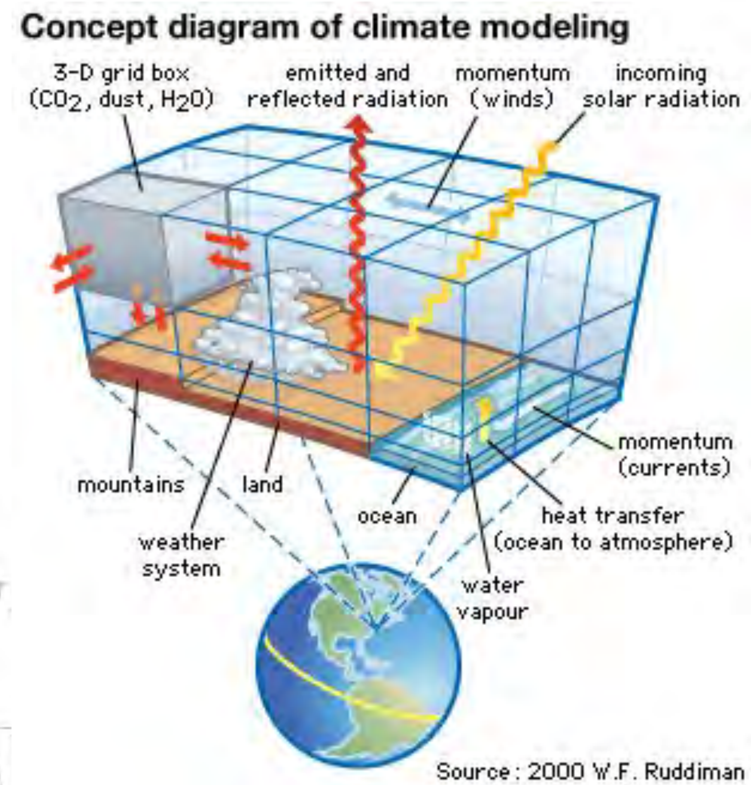
Lebesgue

Measure



Difficult example: global climate model

Difficult example: global climate model



Difficult example: global climate model

Literally 10,000s of climate variables for the entire world

$$(\underline{X}, \phi_t, \mu)$$

Difficult example: global climate model

Literally 10,000s of climate variables for the entire world

$$(X, \phi_t, \mu)$$

The evolution of these variables over time

Difficult example: global climate model

$$(X, \phi_t, \mu)$$

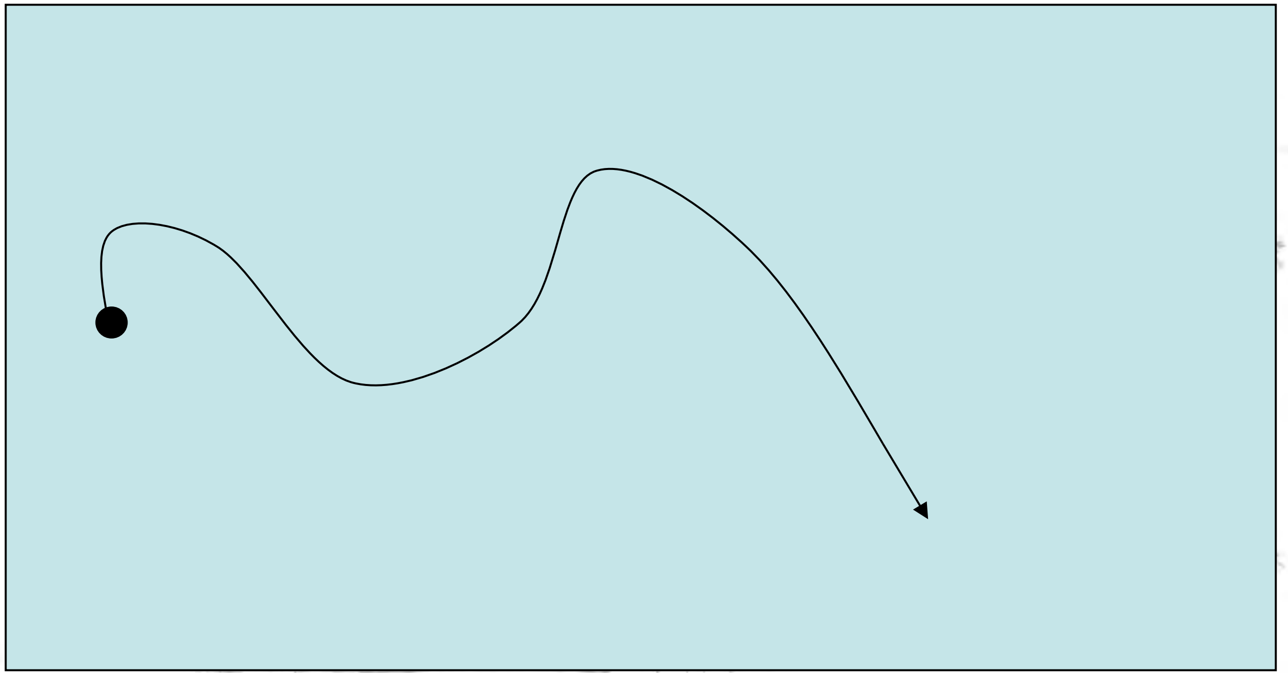
Literally 10,000s of climate variables for the entire world

The evolution of these variables over time

The so-called invariant measure of the dynamics

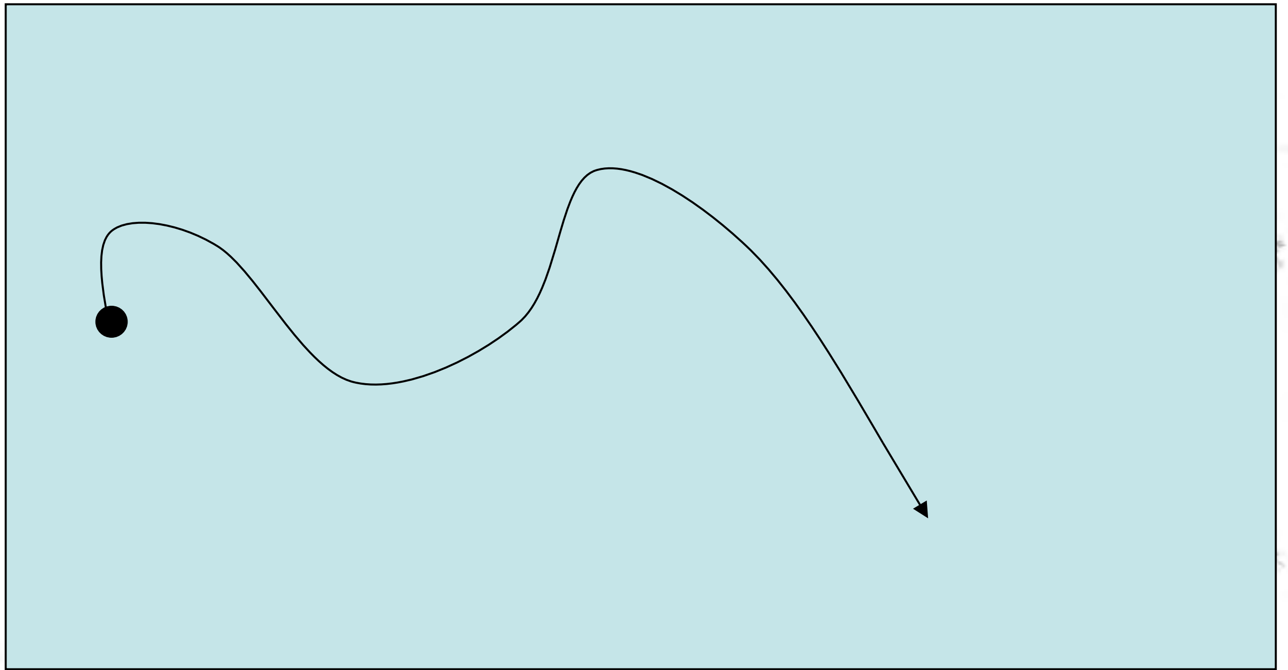
Locating the Issues

Dynamical system (X, ϕ_t, μ)



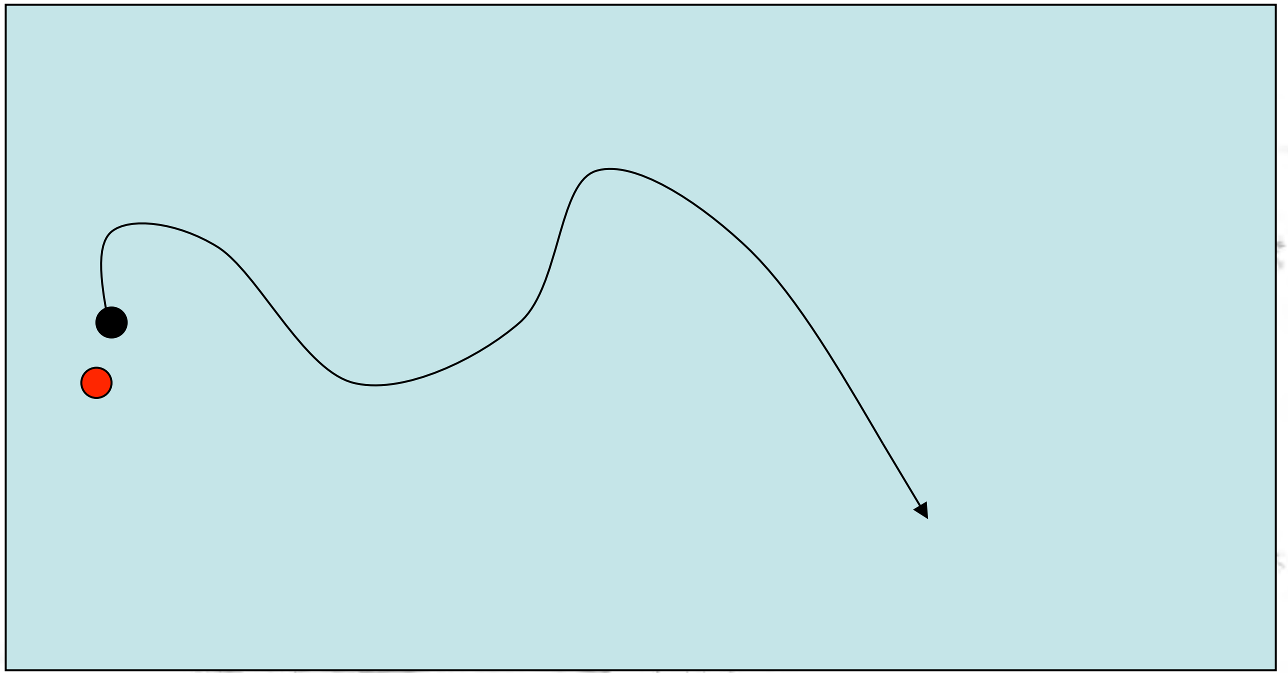
Locating the Issues

Dynamical system (X, ϕ_t, μ)



Locating the Issues

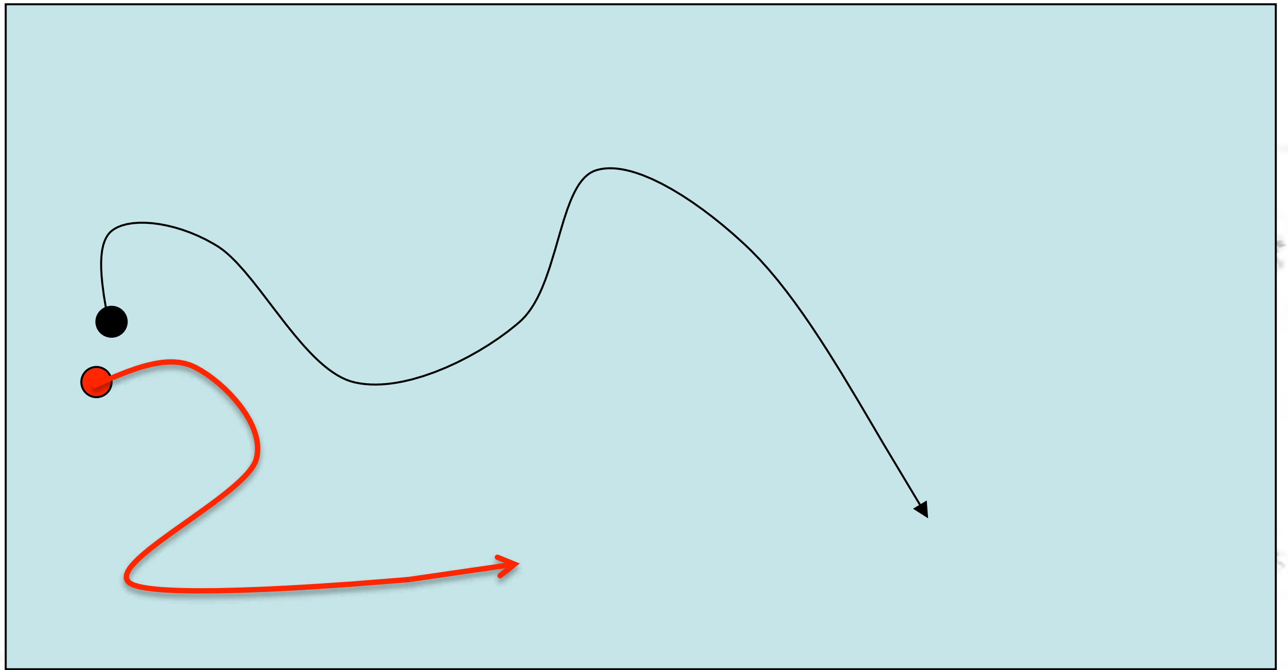
Dynamical system (X, ϕ_t, μ)



Initial Condition Error (ICE)

Locating the Issues

Dynamical system (X, ϕ_t, μ)



Initial Condition Error (ICE)

Locating the Issues

Initial condition error

↓

Butterfly Effect

Locating the Issues

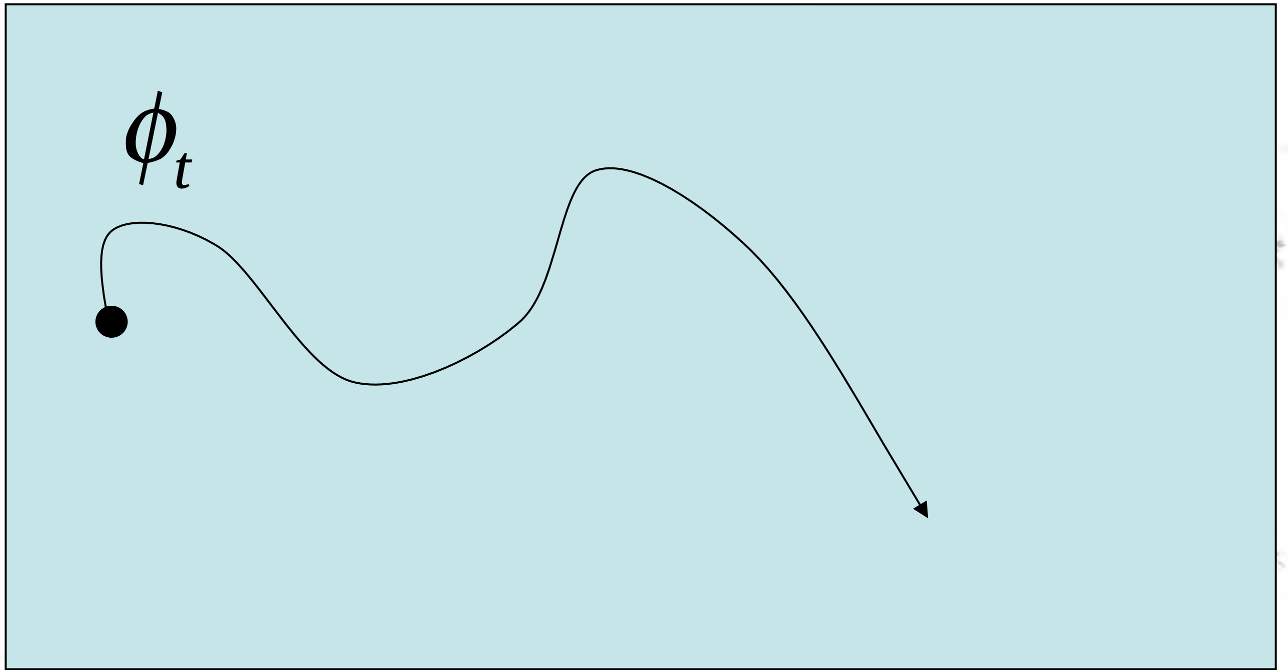
SME: the time evolution of the model differ from the time evolution of the system under study:

$$\phi_t^S = \phi_t^M + \delta_t$$

True dynamics — Model — "Difference"

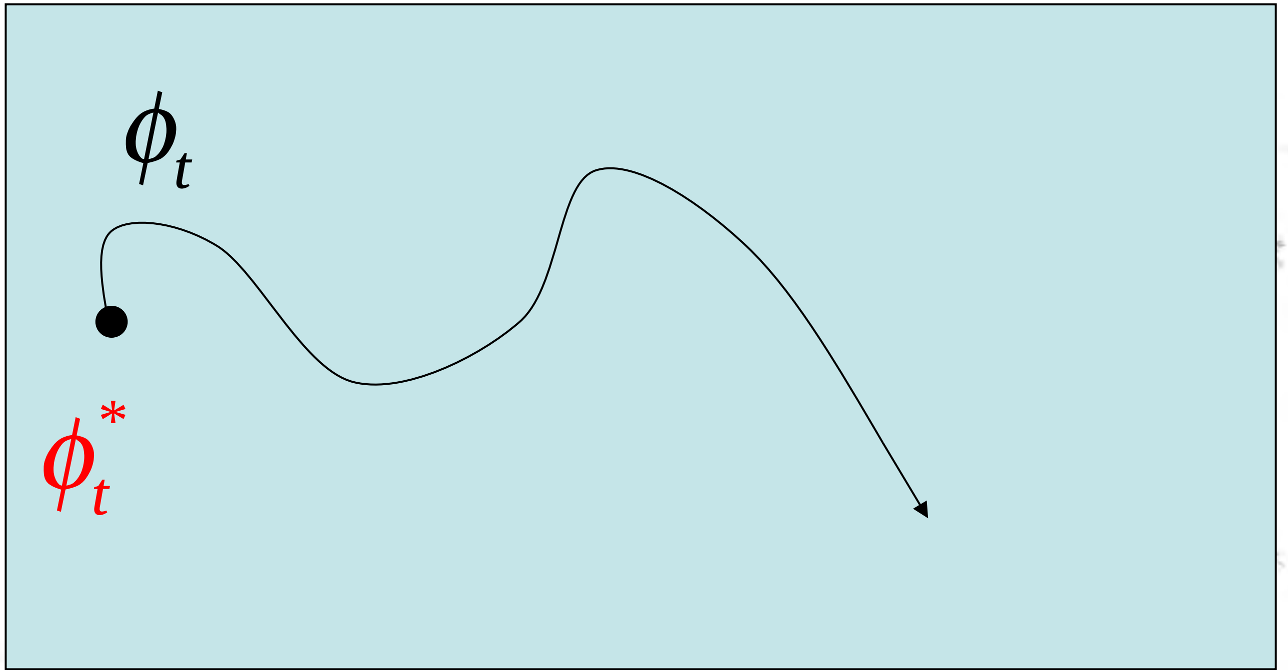
Locating the Issues

Dynamical system (X, ϕ_t, μ)



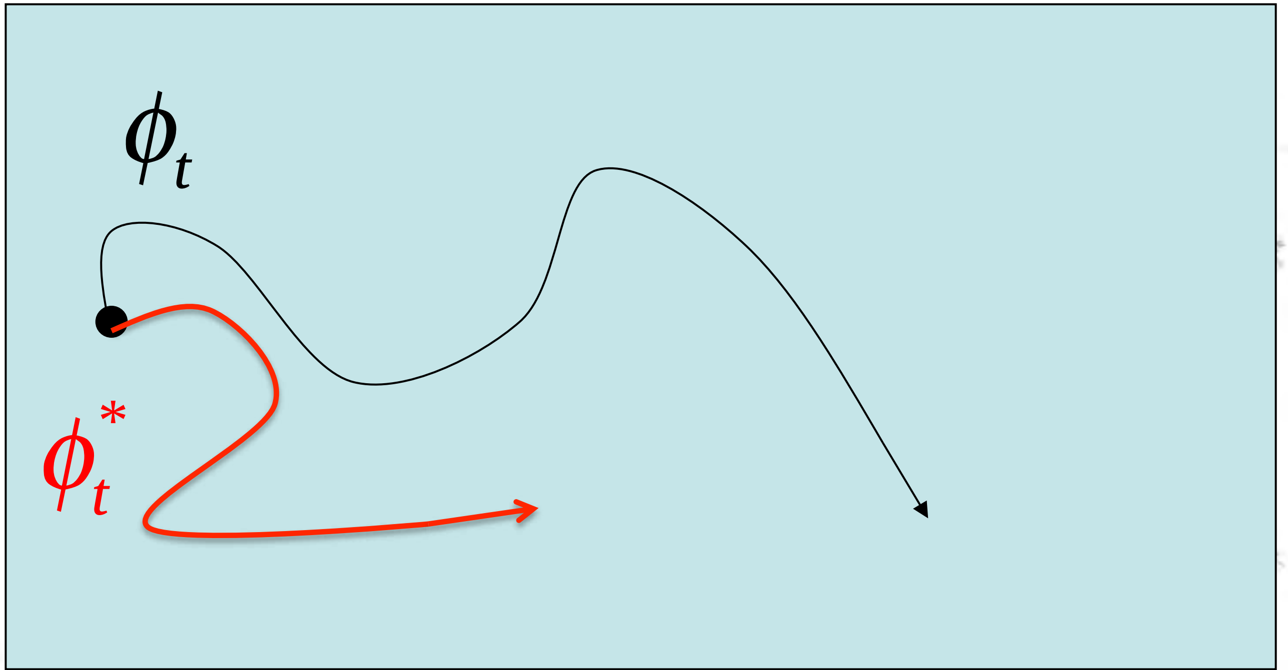
Locating the Issues

Dynamical system (X, ϕ_t, μ)



Locating the Issues

Dynamical system (X, ϕ_t, μ)



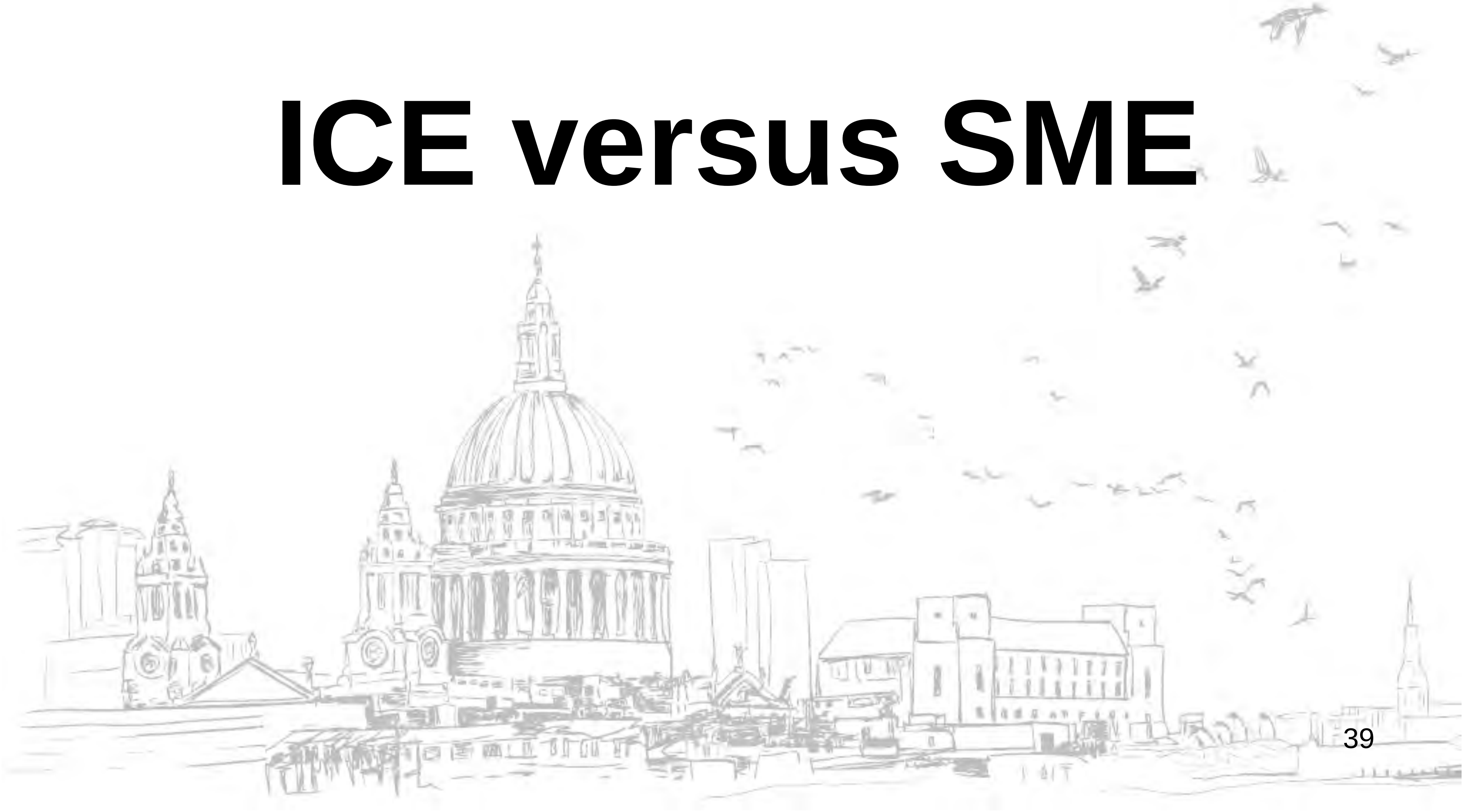
Locating the Issues

Structural Model Error

↓

Hawkmoth Effect

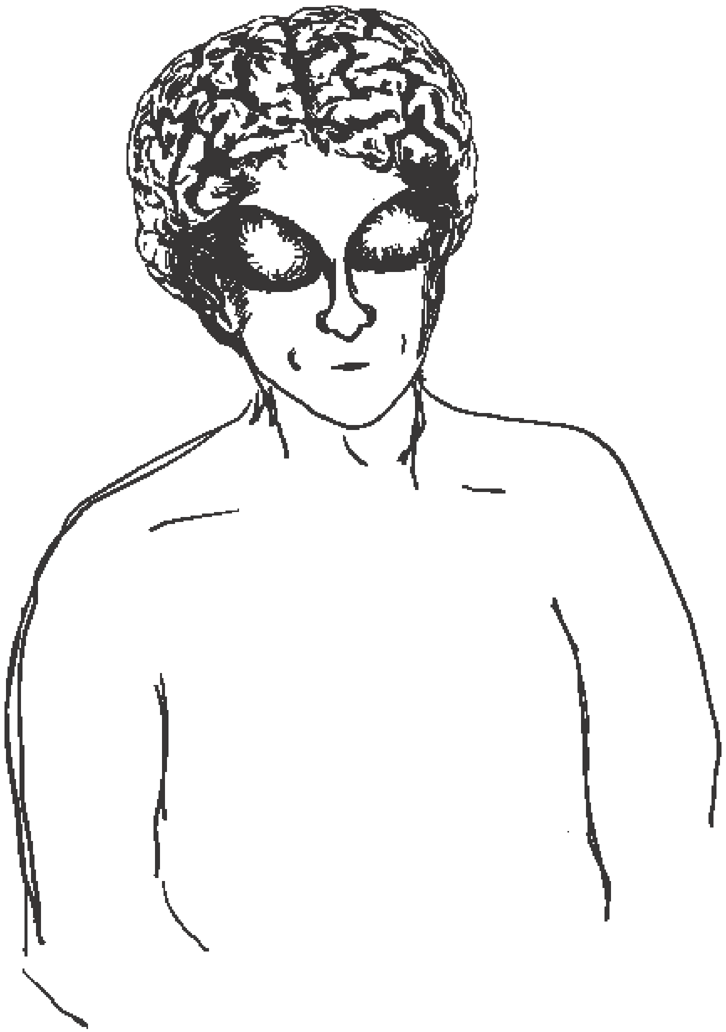
ICE versus SME



Meet Laplace's Demon

1. Unlimited computational power
2. Unlimited dynamical knowledge
3. Unlimited observational power

(Laplace 1814)



The Demon knows everything.

In Laplace's own words: 'nothing would be uncertain and the future, as the past, would be present to [his] eyes'.

So the Demon's model of the world's climate would be trustworthy because it provides the full truth.

But what happens if we are less capable than the Demon?

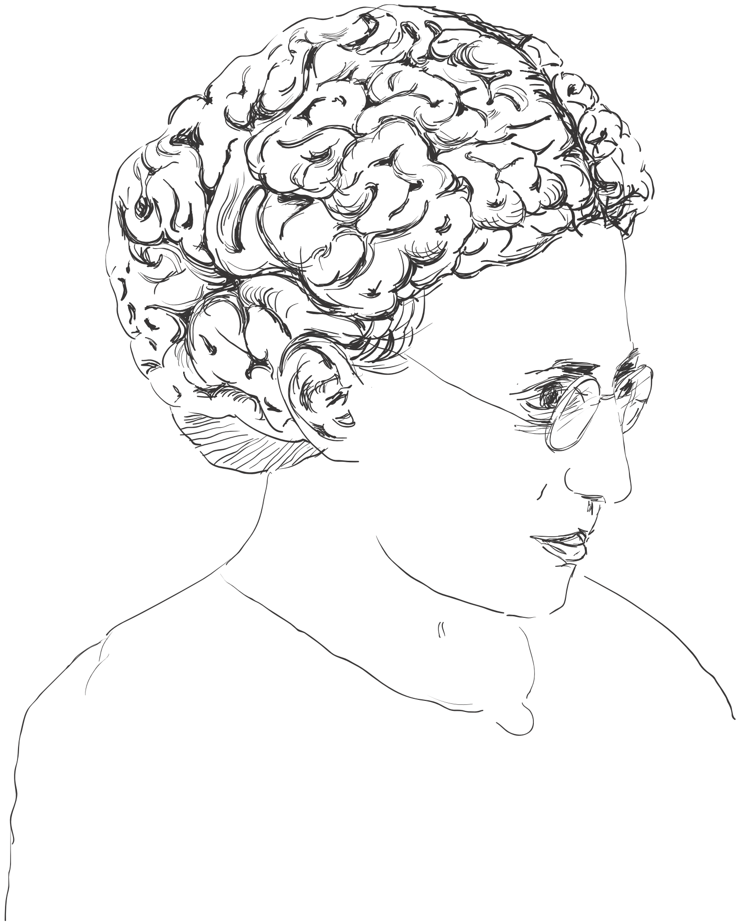
Meet the Senior Apprentice

1. Unlimited computational power
2. Unlimited dynamical knowledge
3. No unlimited observational power

Meet the Senior Apprentice

1. Unlimited computational power
2. Unlimited dynamical knowledge
3. No unlimited observational power

In other words, the Senior Apprentice only has noisy observations.

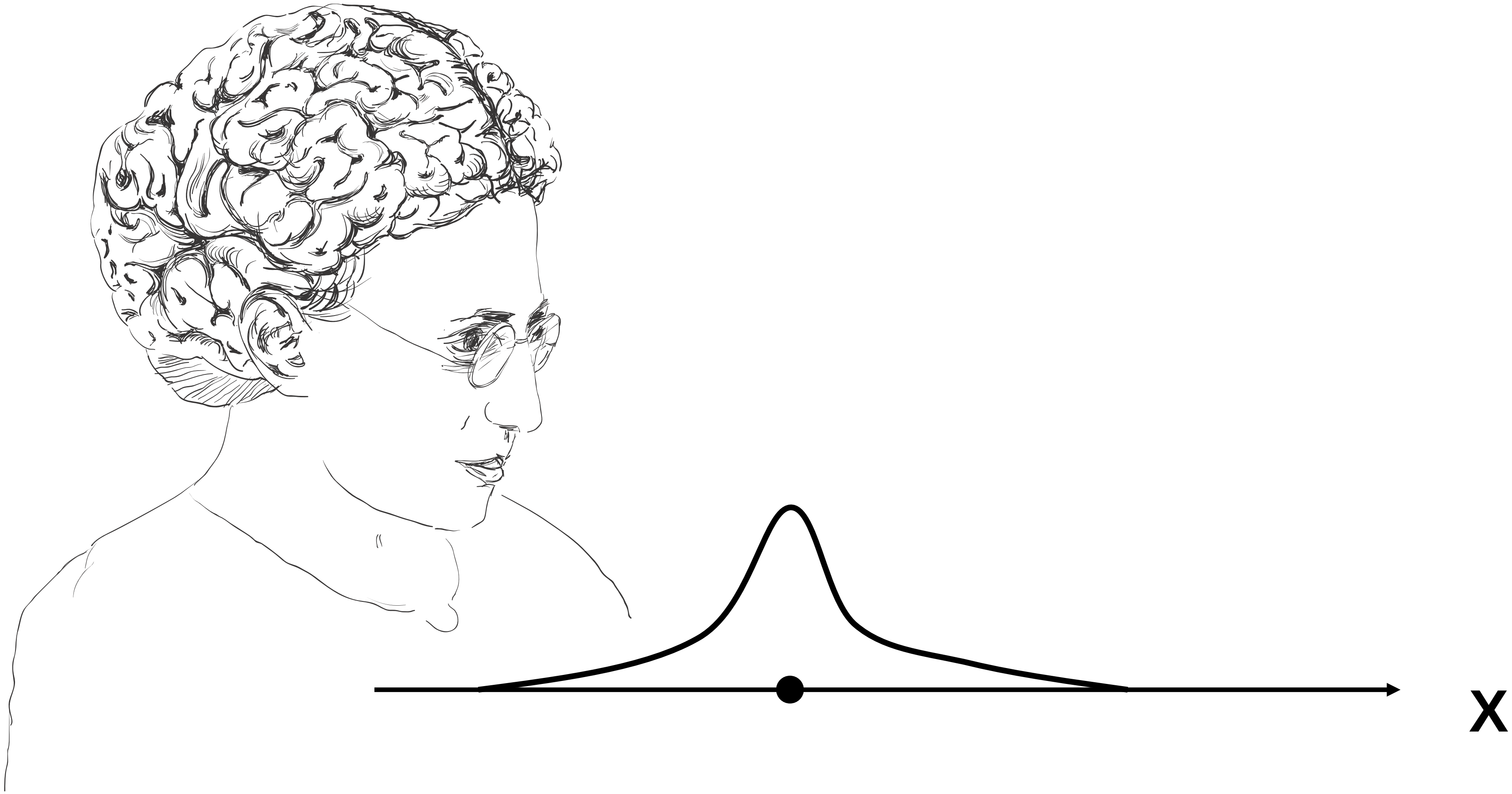


How could the limitation of not having unlimited observational power be overcome?

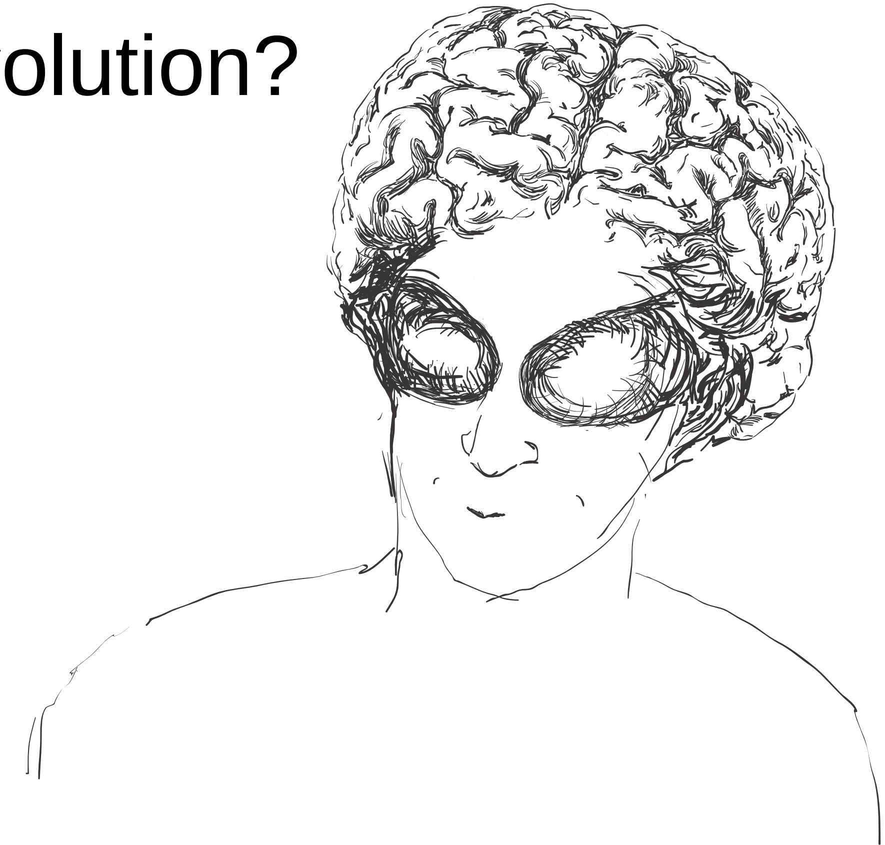
Reply:

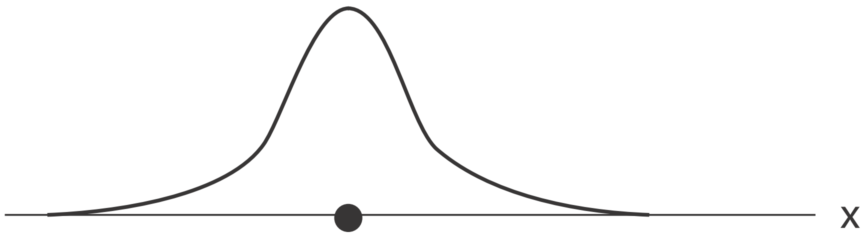
Initial Condition Ensemble

That is, she puts a probability distribution over an approximate initial condition.

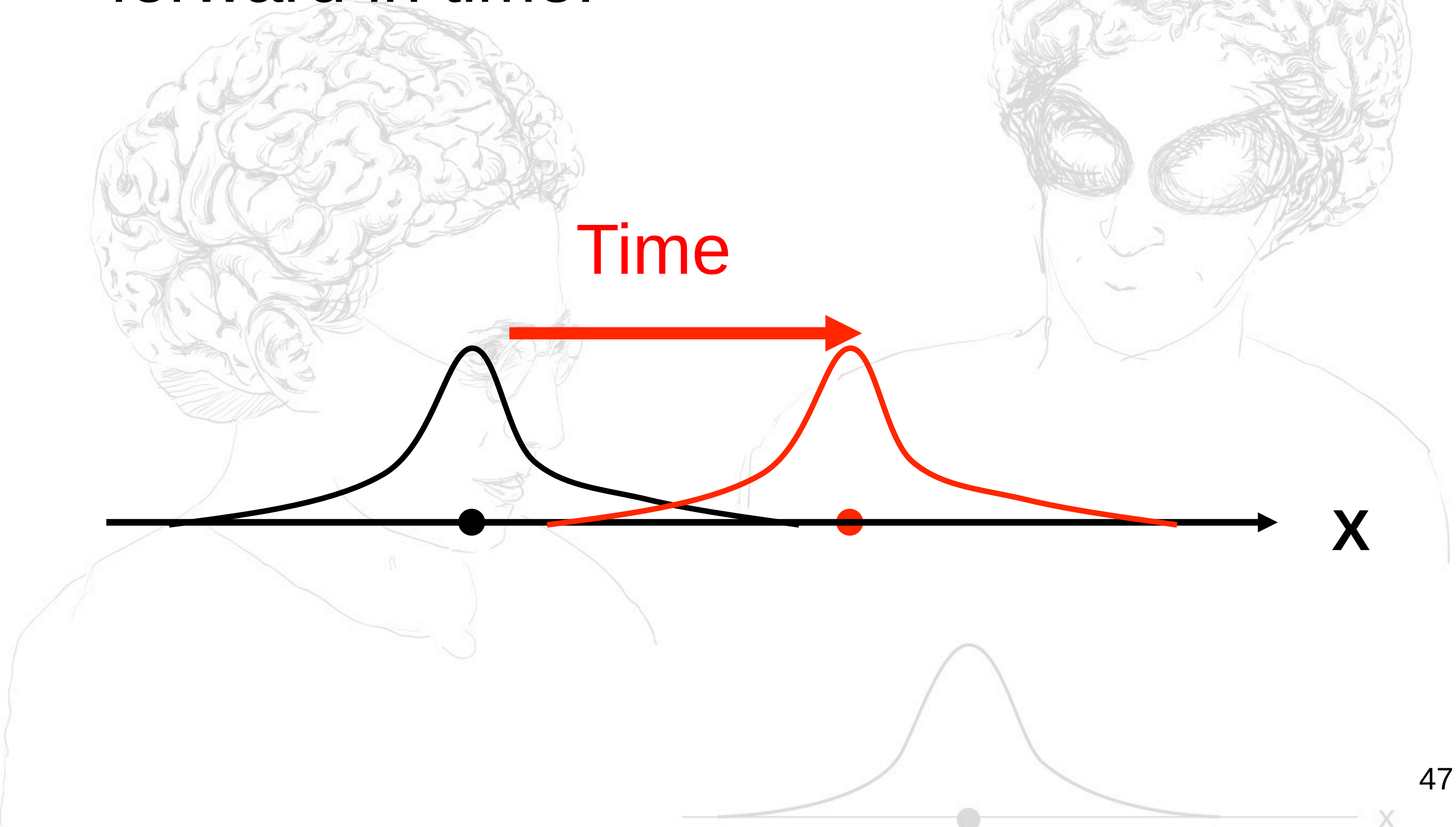


Prediction? Time Evolution?

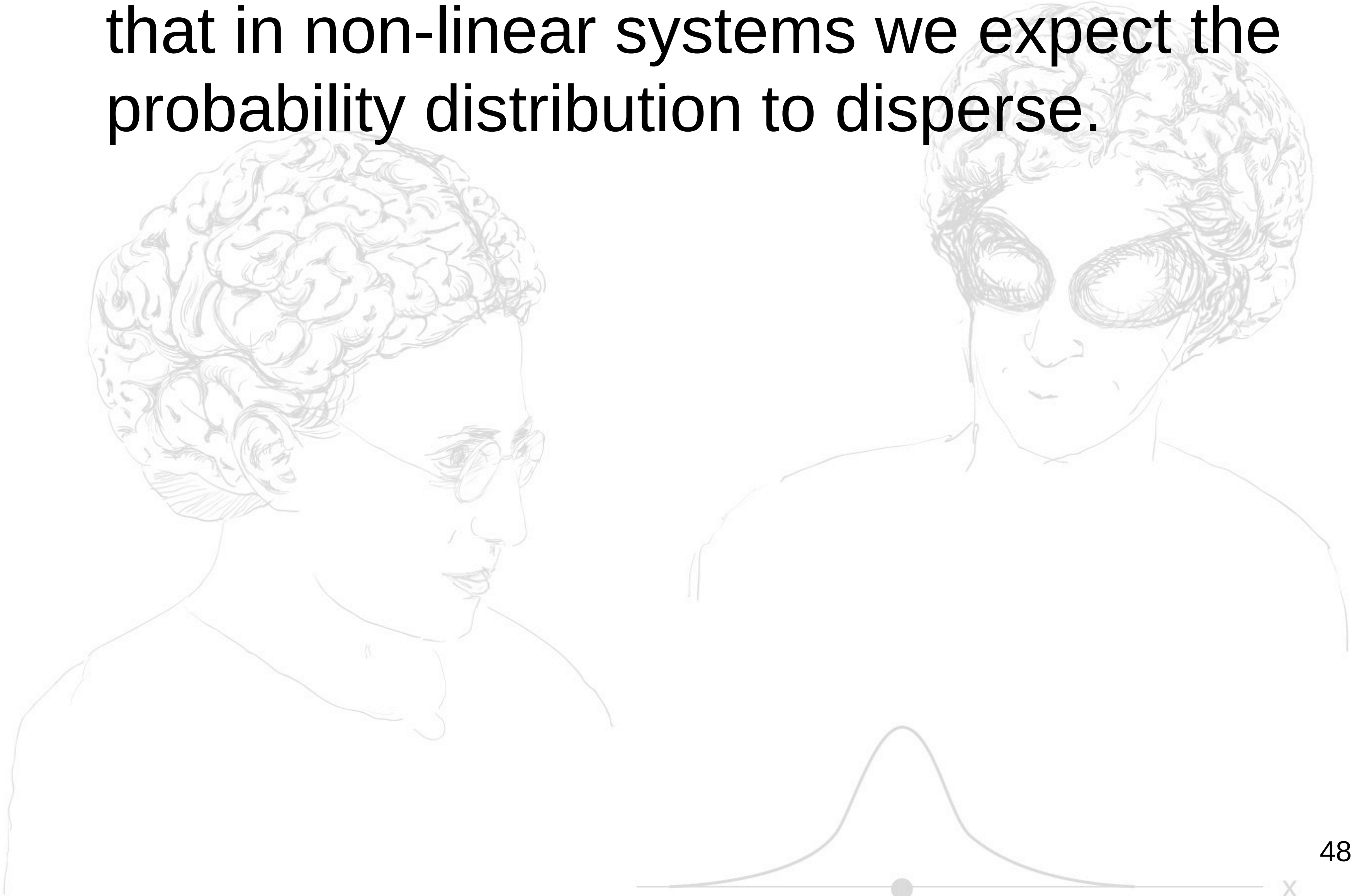




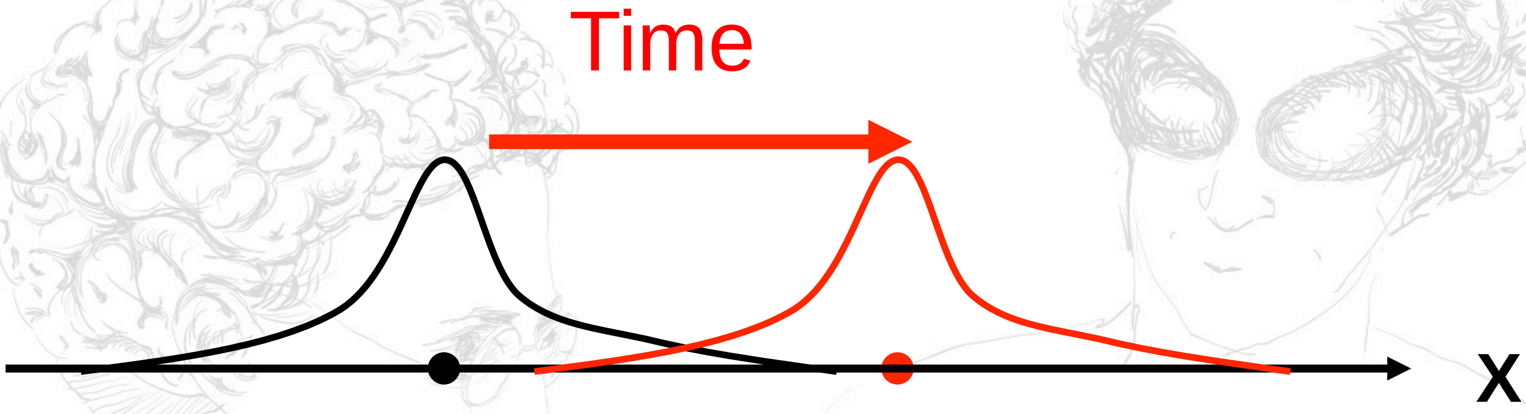
Generate probabilistic predictions by moving the initial probability distribution forward in time:



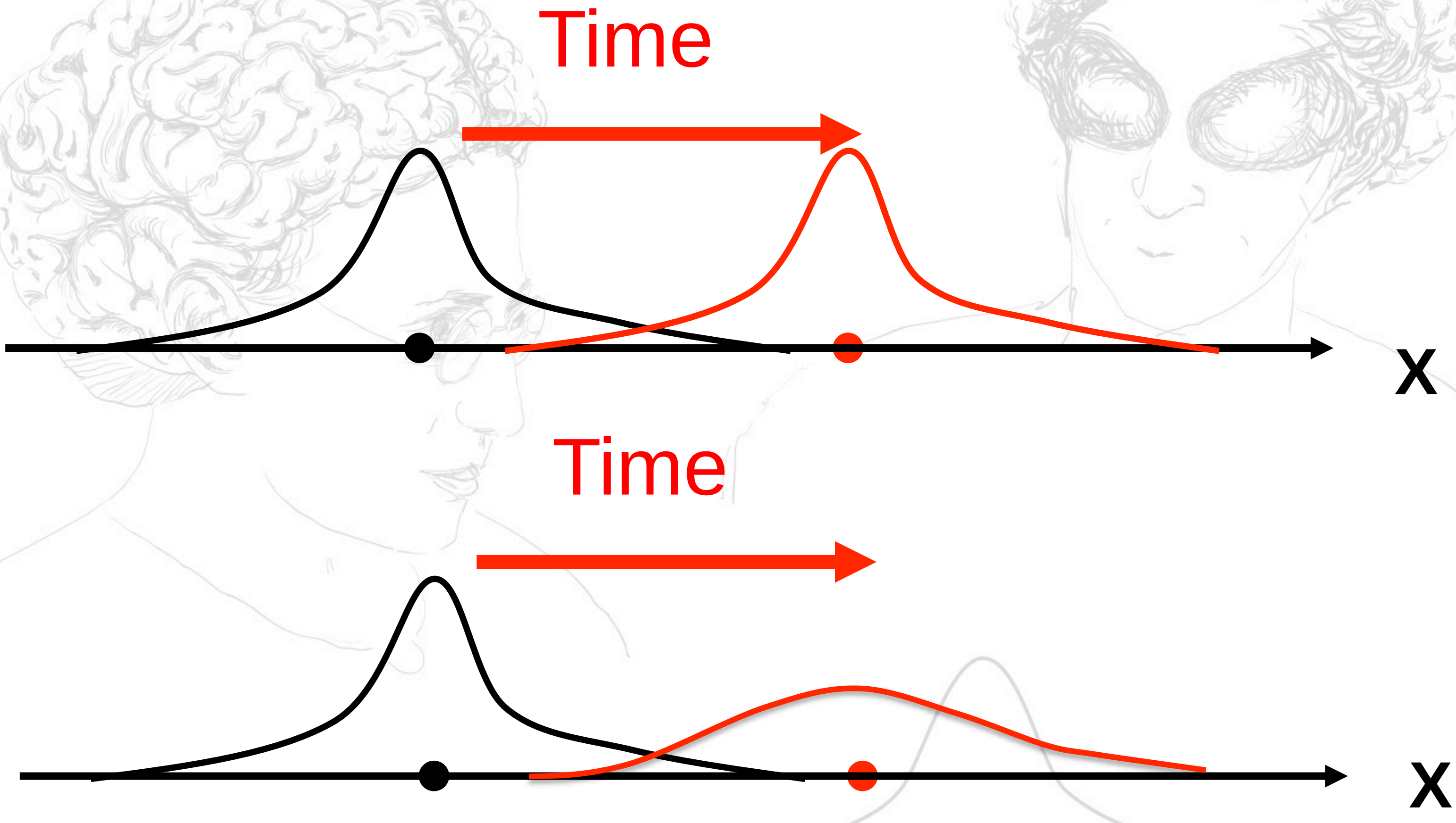
Implications for prediction? They figure that in non-linear systems we expect the probability distribution to disperse.



Implications for prediction? They figure that in non-linear systems we expect the probability distribution to disperse.



Implications for prediction? They figure that in non-linear systems we expect the probability distribution to disperse.



Why dispersion?

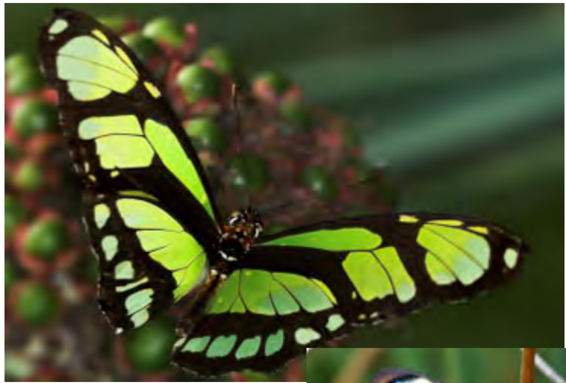
Why dispersion?

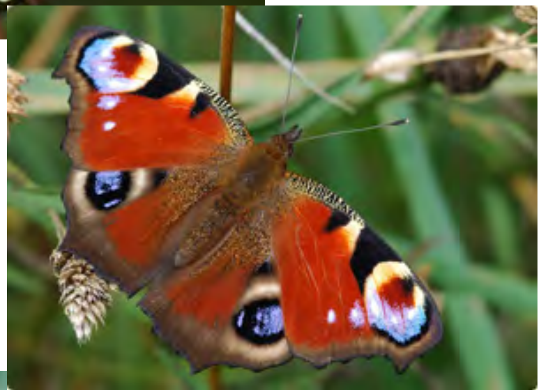
Distributions become ***<u>uninformative</u>*** as time passes, but they ***<u>do not become misleading</u>***.

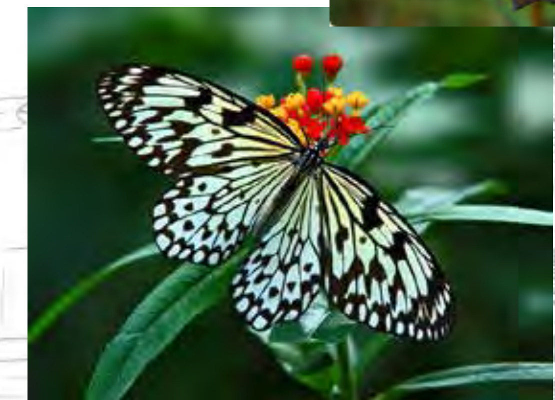
The Senior Apprentice realises that this is the limitation that she has to accept.

It is the price to pay for not having unlimited observational power.

Or: butterflies are pretty; hawkmoths are ugly.



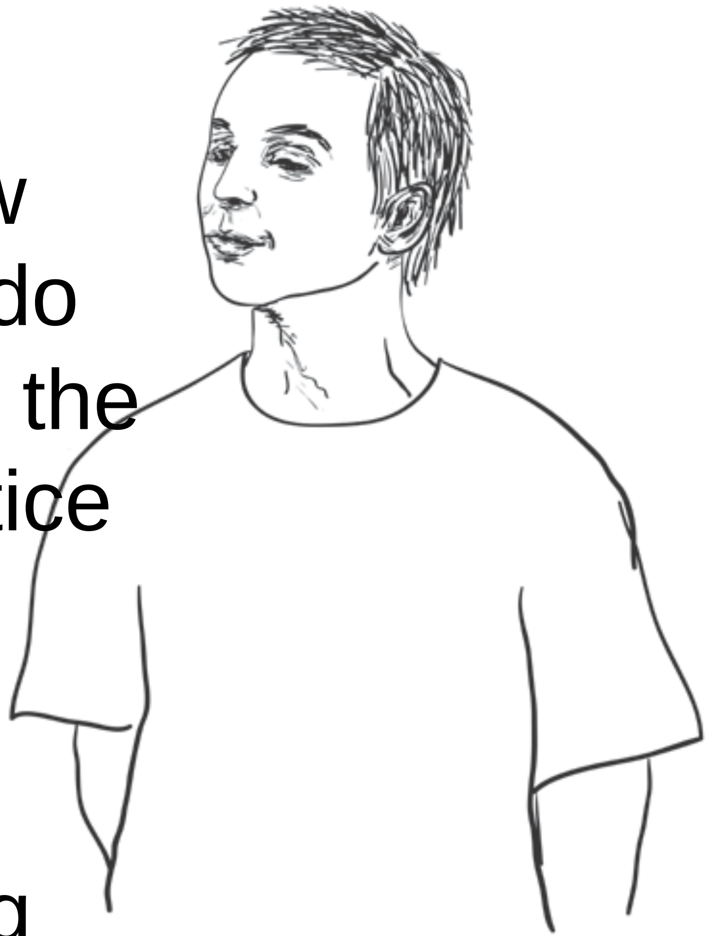




Meet the Freshman Apprentice

1. Unlimited computational power
2. No unlimited dynamical knowledge
3. No unlimited observational power

The Freshman Apprentice now claims he can do everything that the Senior Apprentice can do, his additional limitation notwithstanding

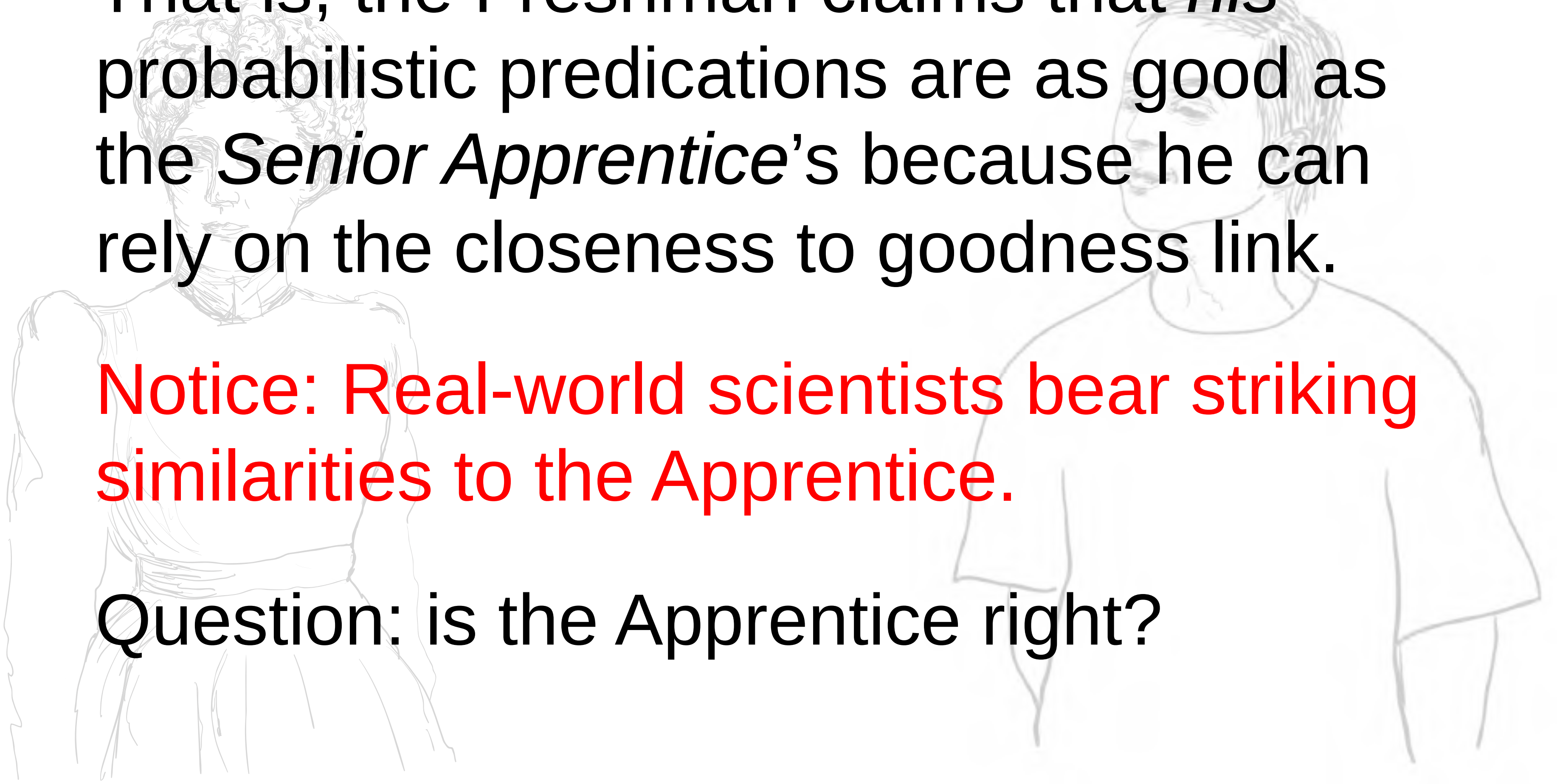


Recall: The Freshman can't formulate the exact dynamics of a system.

Reaction: Distortions and idealisations of all kind are acceptable as long as the resulting model is close enough to the truth.

This is the ***closeness-to-goodness link***.

→ This is a crucial part!

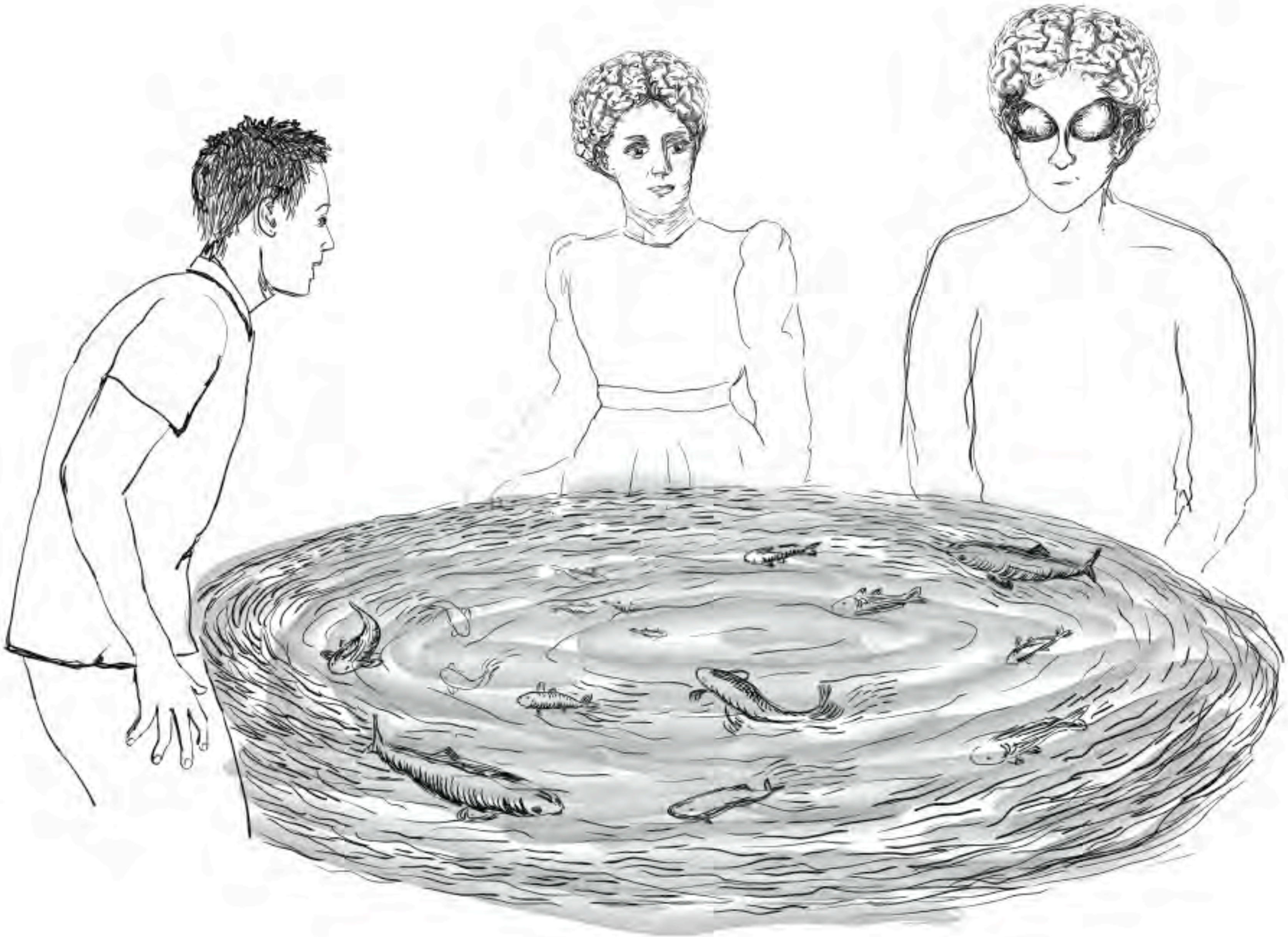


That is, the Freshman claims that *his* probabilistic predications are as good as the *Senior Apprentice*'s because he can rely on the closeness to goodness link.

Notice: Real-world scientists bear striking similarities to the Apprentice.

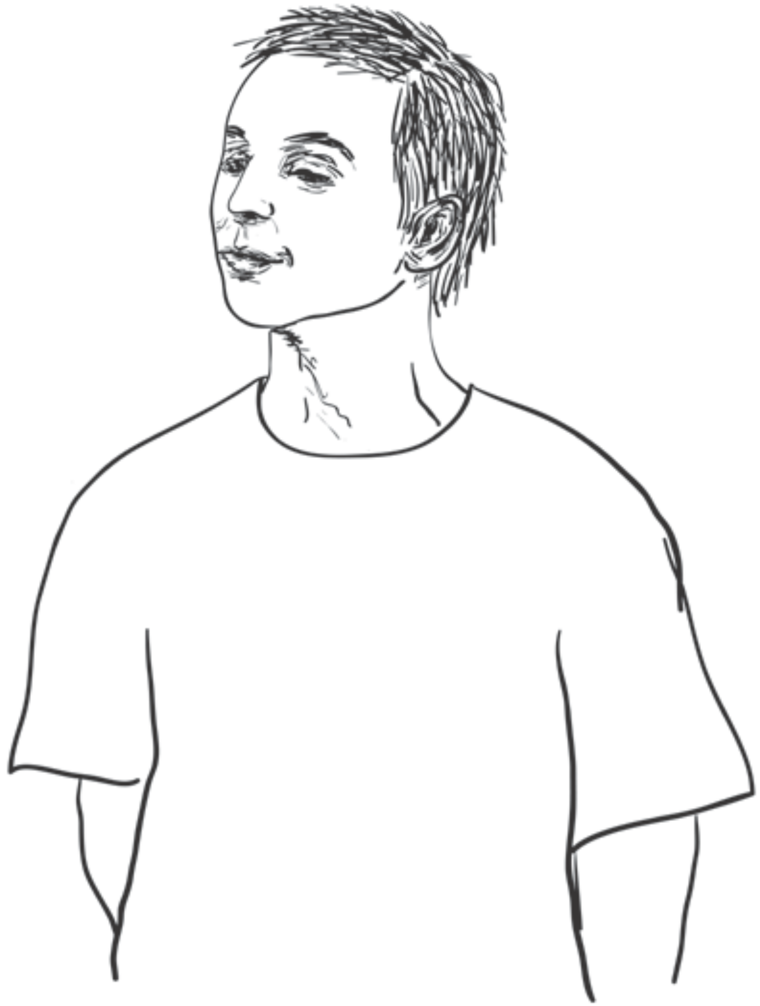
Question: is the Apprentice right?


No way!



Population density:

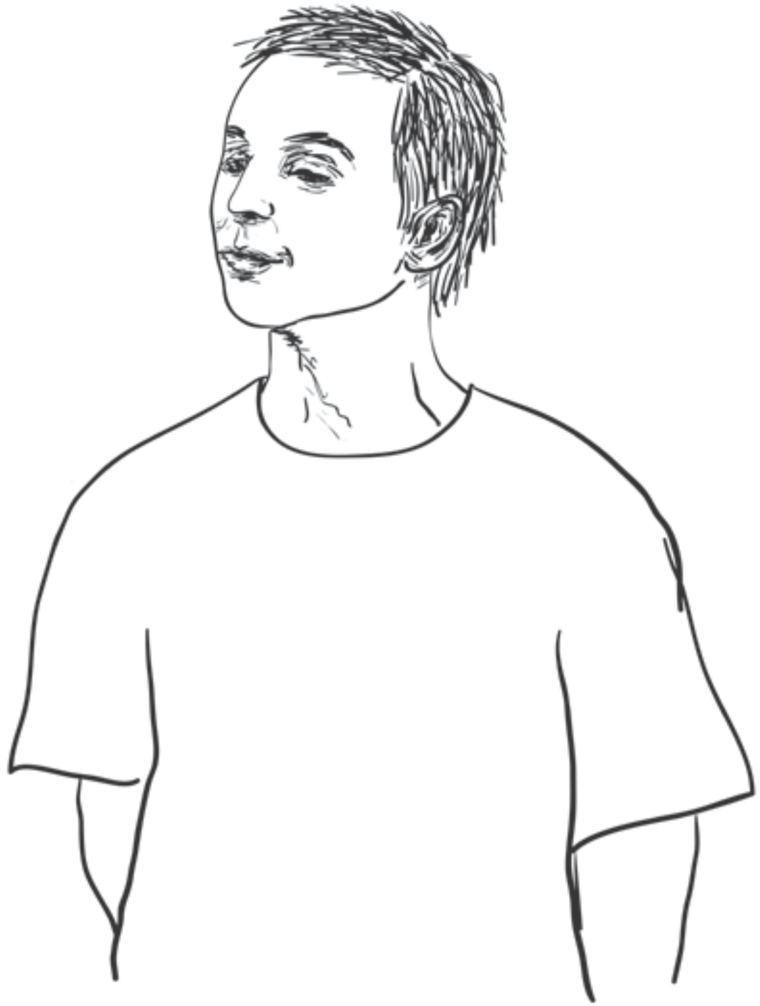
$$\rho = \frac{\#\,\text{fish} \,/\, m^3}{\#\,\text{max}\,\text{fish} \,/\, m^3}$$



Population density:

$$\rho = \frac{\#\,\text{fish} / m^3}{\#\,\text{max fish} / m^3}$$

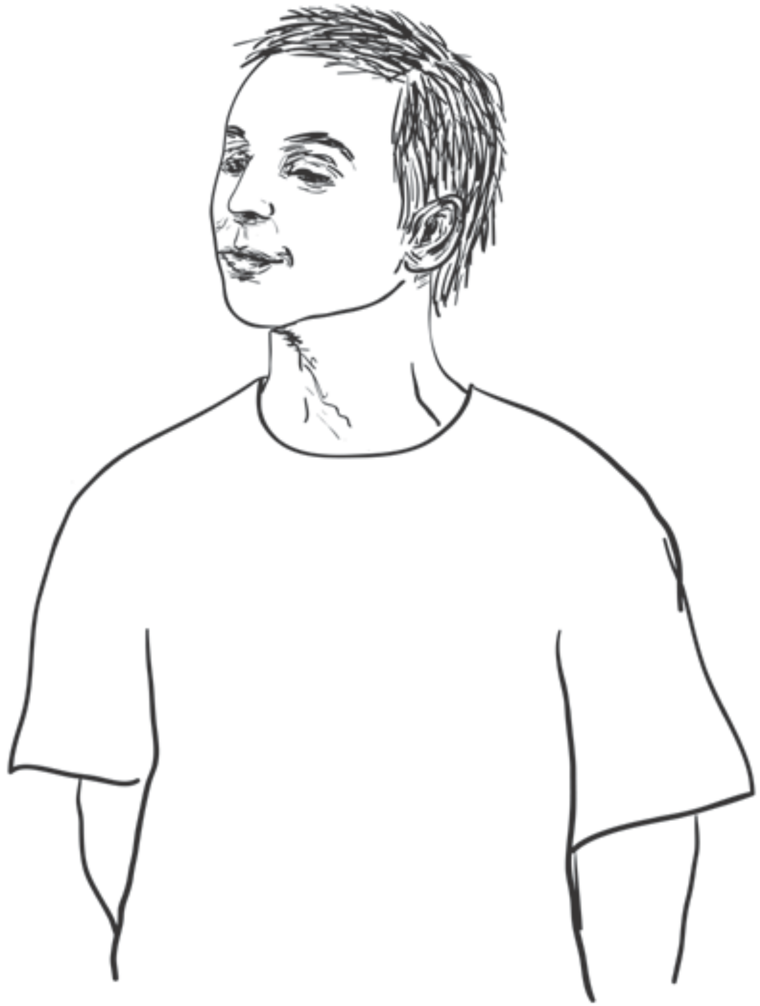
Hence: $\rho \in [0,1]$



Population density:

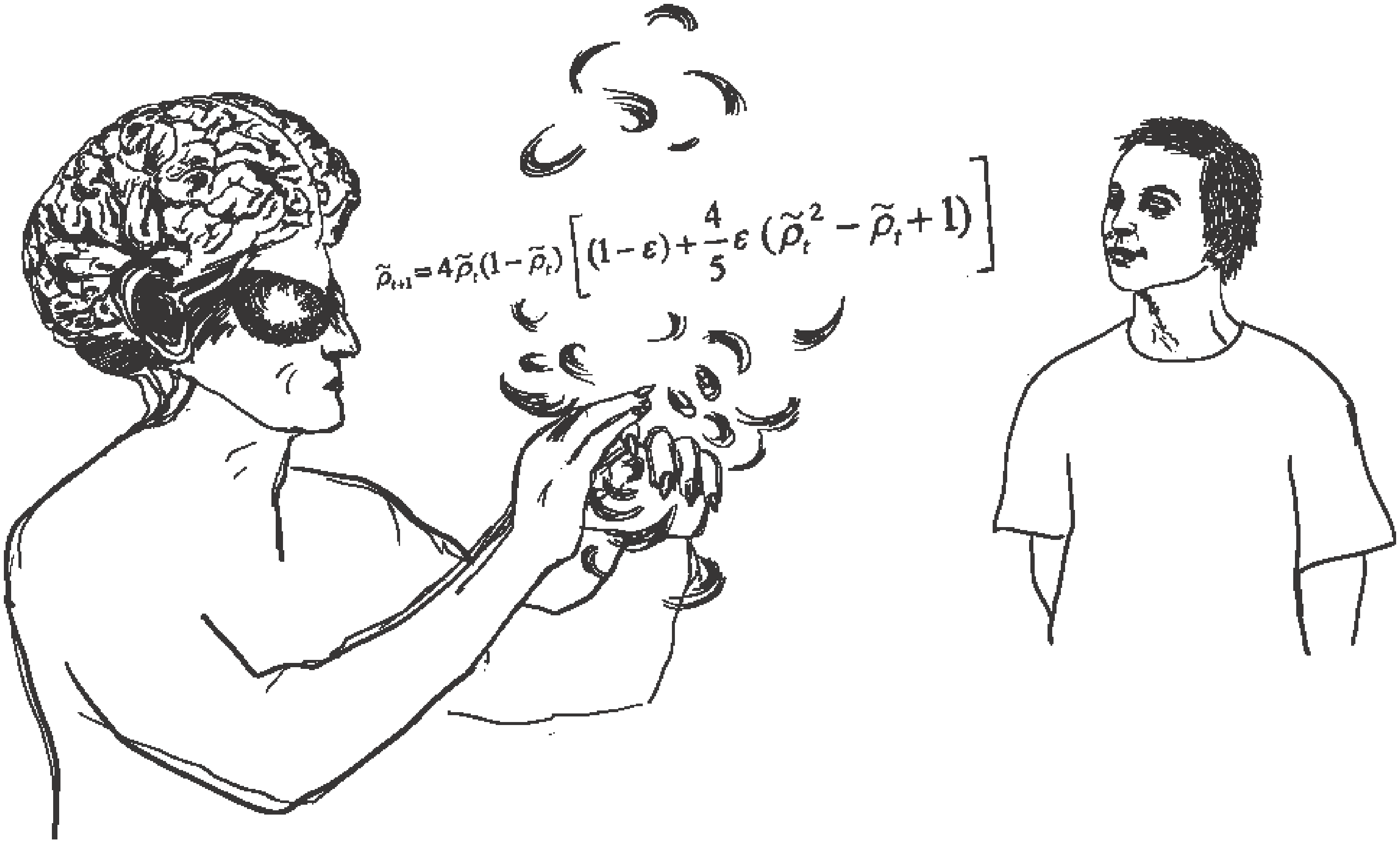
$$\rho = \frac{\#\,\text{fish} / m^3}{\#\,\text{max fish} / m^3}$$

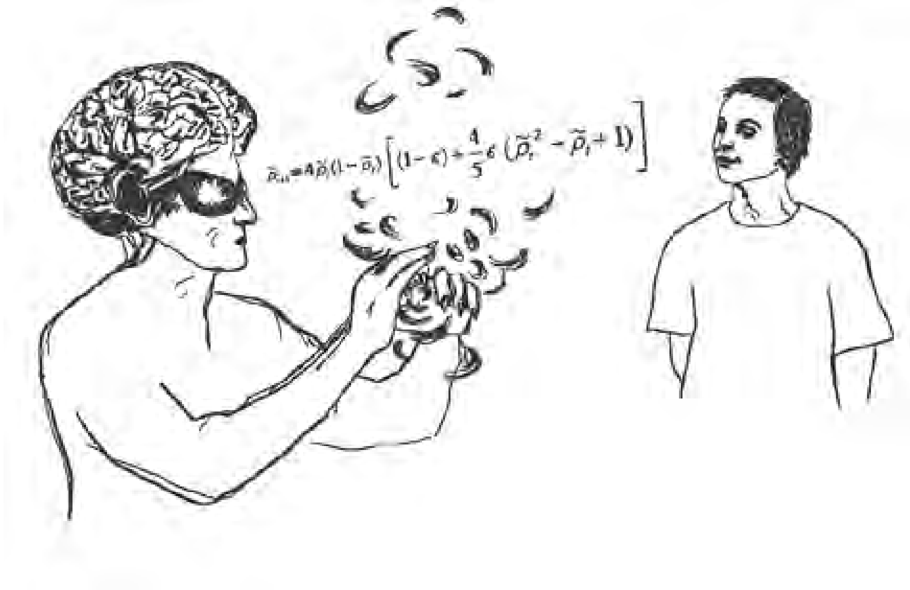
Hence: $\rho \in [0,1]$



Model:

$$\rho_{t+1} = 4\rho_t(1-\rho_t)$$

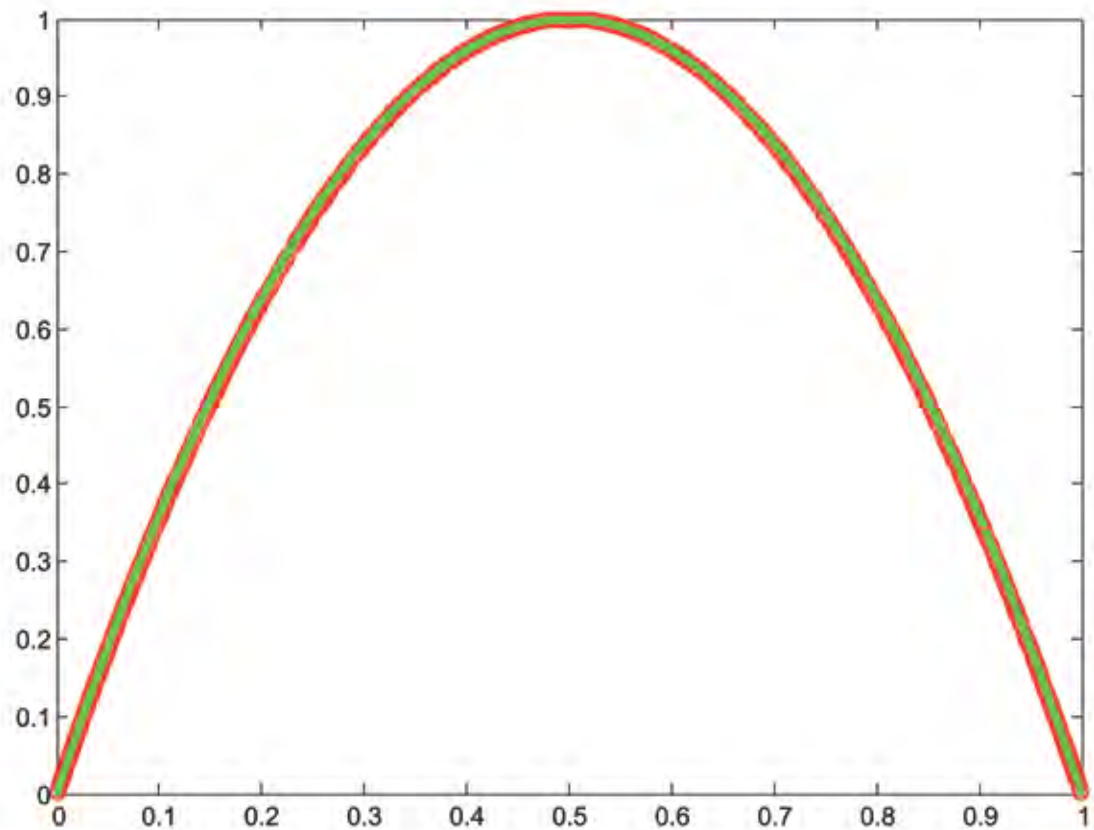

$\tilde{\rho}_{t+1} = 4\tilde{\rho}_t(1-\tilde{\rho}_t)\left[(1-\varepsilon)+\frac{4}{5}\varepsilon\,(\tilde{\rho}_t^{\,2}-\tilde{\rho}_t+1)\right]$



$$\tilde{\rho}_{t+1} = 4\tilde{\rho}_t(1-\varepsilon)(1-\tilde{\rho}_t)+\varepsilon\frac{16}{5}\tilde{\rho}_t(1-2\tilde{\rho}_t{}^2+\tilde{\rho}_t{}^3)$$

where $\varepsilon = 0.1$

The Apprentice remains defiant:



Green – Apprentice and Red - Demon

Mathematically:

$\rho_{t+1} = 4\rho_t(1-\rho_t)$ + small perturbation

$$\tilde{\rho}_{t+1} = 4\tilde{\rho}_t(1-\varepsilon)(1-\tilde{\rho}_t) + \varepsilon\frac{16}{5}\tilde{\rho}_t(1-2\tilde{\rho}_t^{\,2}+\tilde{\rho}_t^{\,3})$$

Mathematically:

$$\rho_{t+1} = 4\rho_t(1-\rho_t) \text{ + small perturbation}$$

$$\tilde{\rho}_{t+1} = 4\tilde{\rho}_t(1-\varepsilon)(1-\tilde{\rho}_t) + \varepsilon\frac{16}{5}\tilde{\rho}_t(1-2\tilde{\rho}_t{}^2 + \tilde{\rho}_t{}^3)$$

One step error: 0.001

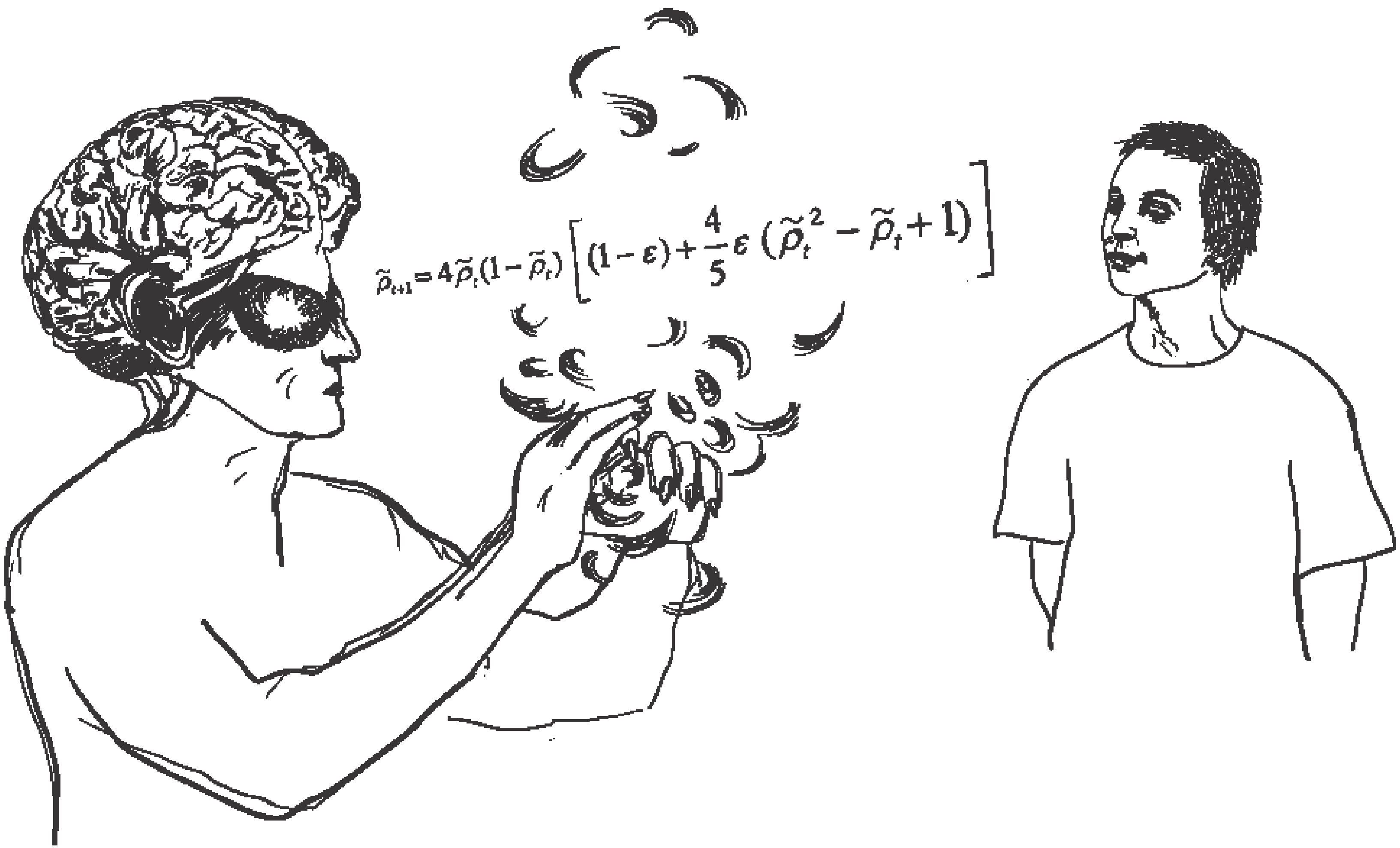
Mathematically:

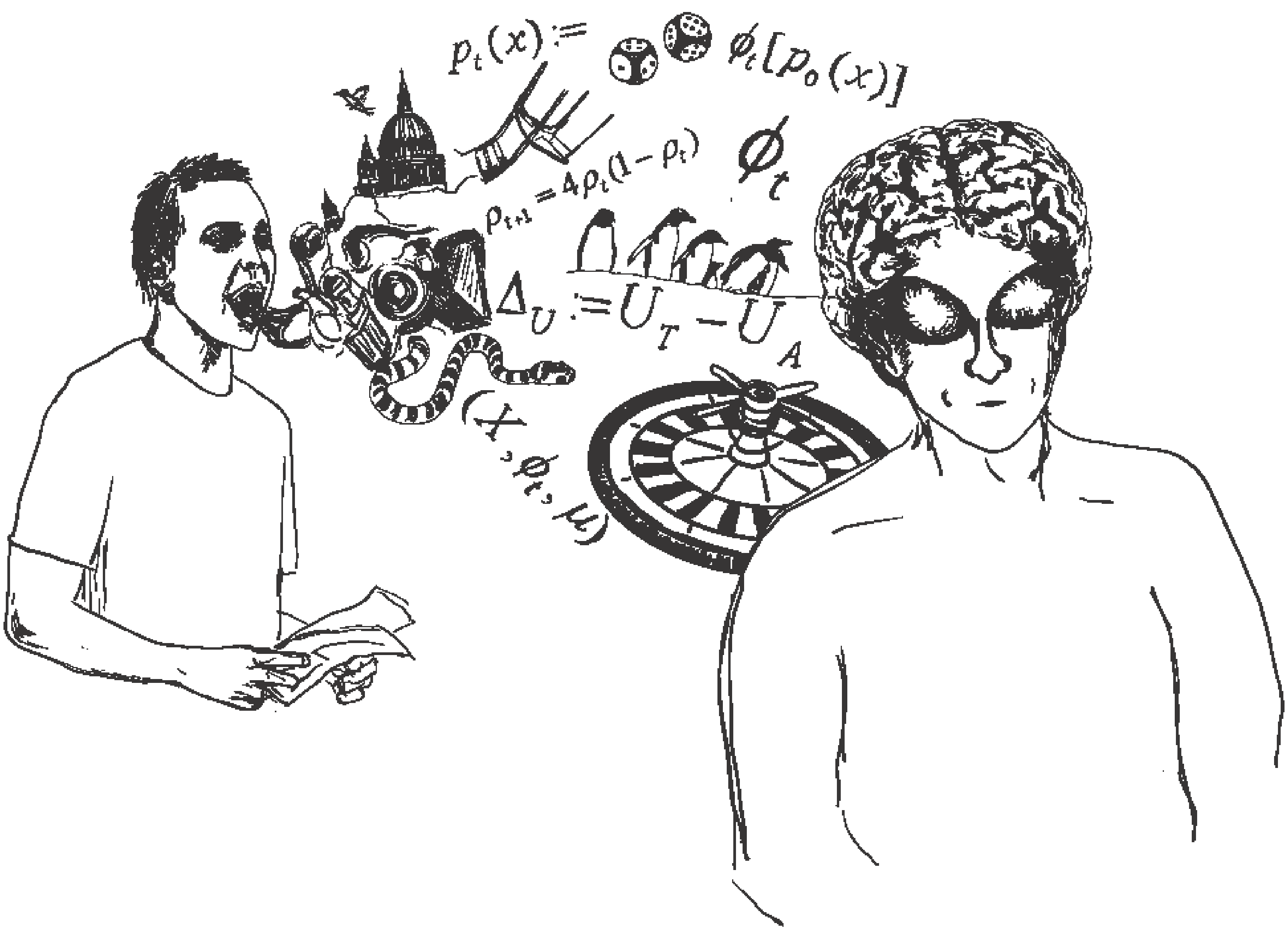
$$\rho_{t+1} = 4\rho_t(1-\rho_t) \text{ + small perturbation}$$

$$\tilde{\rho}_{t+1} = 4\tilde{\rho}_t(1-\varepsilon)(1-\tilde{\rho}_t) + \varepsilon\frac{16}{5}\tilde{\rho}_t(1-2\tilde{\rho}_t^{\ 2} + \tilde{\rho}_t^{\ 3})$$

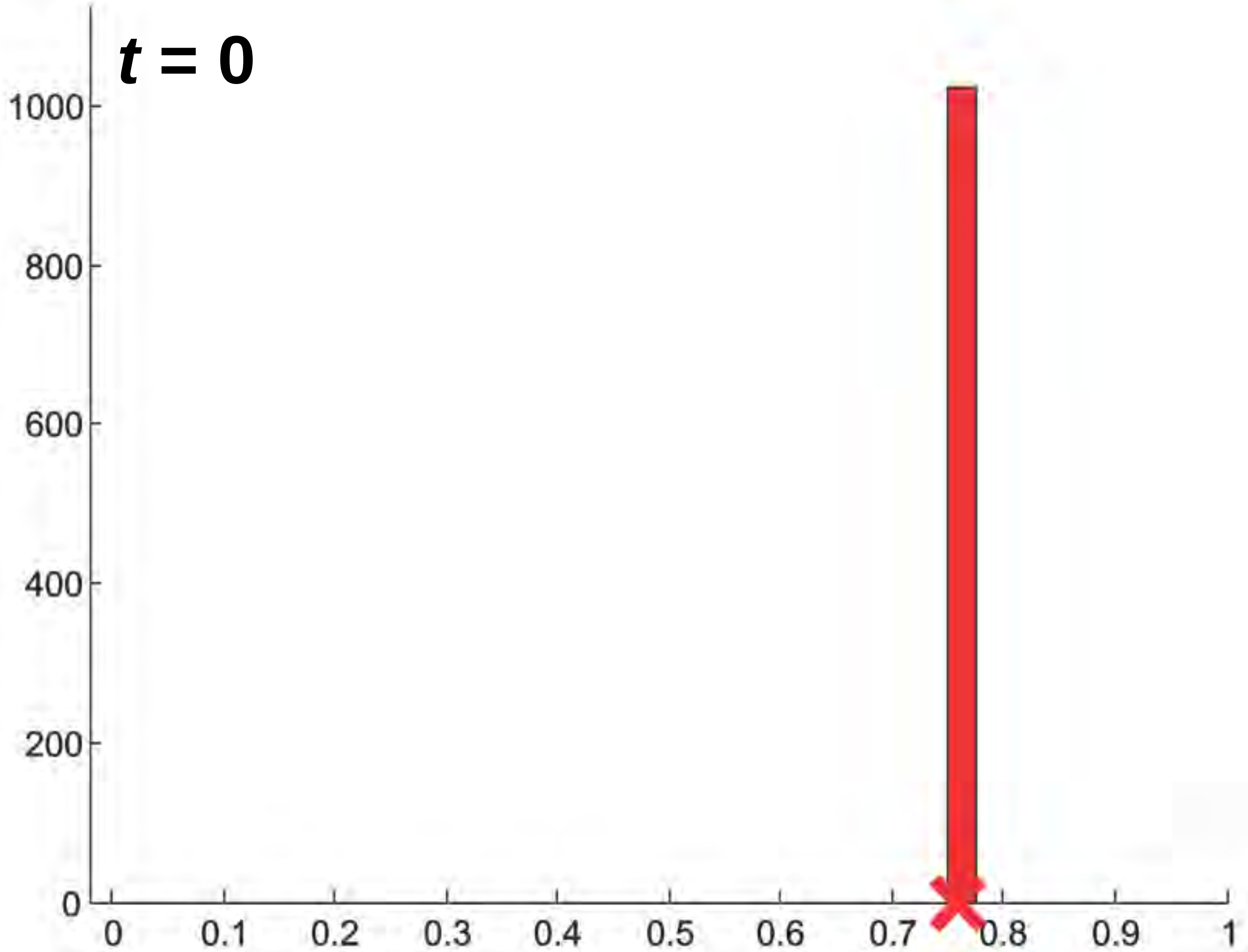
One step error: 0.001

Closeness-to-goodness link: this is close enough and predictions are reliable.

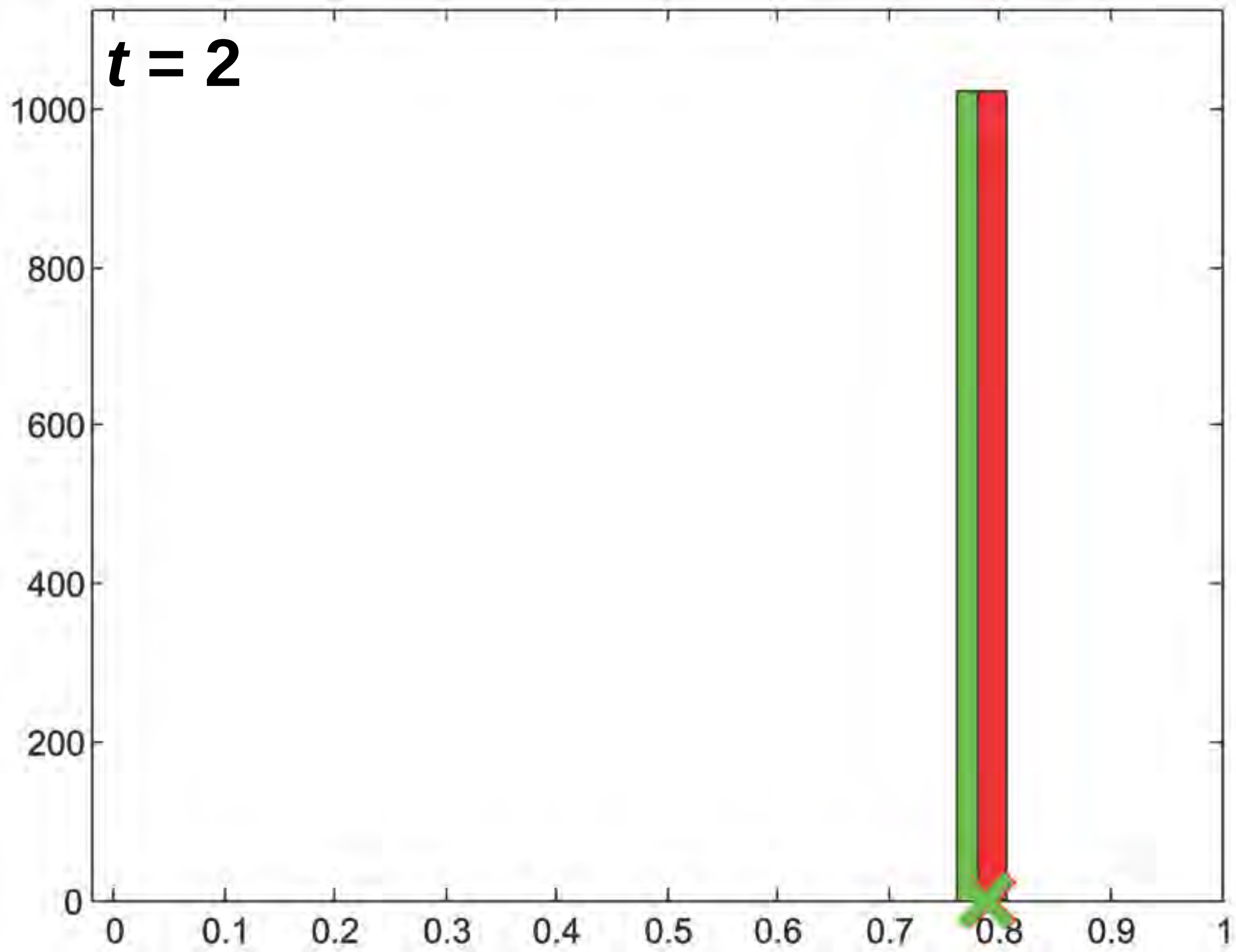

$\tilde{\rho}_{t+1} = 4\tilde{\rho}_t(1-\tilde{\rho}_t)\left[(1-\varepsilon)+\frac{4}{5}\varepsilon\left(\tilde{\rho}_t^{\,2}-\tilde{\rho}_t+1\right)\right]$

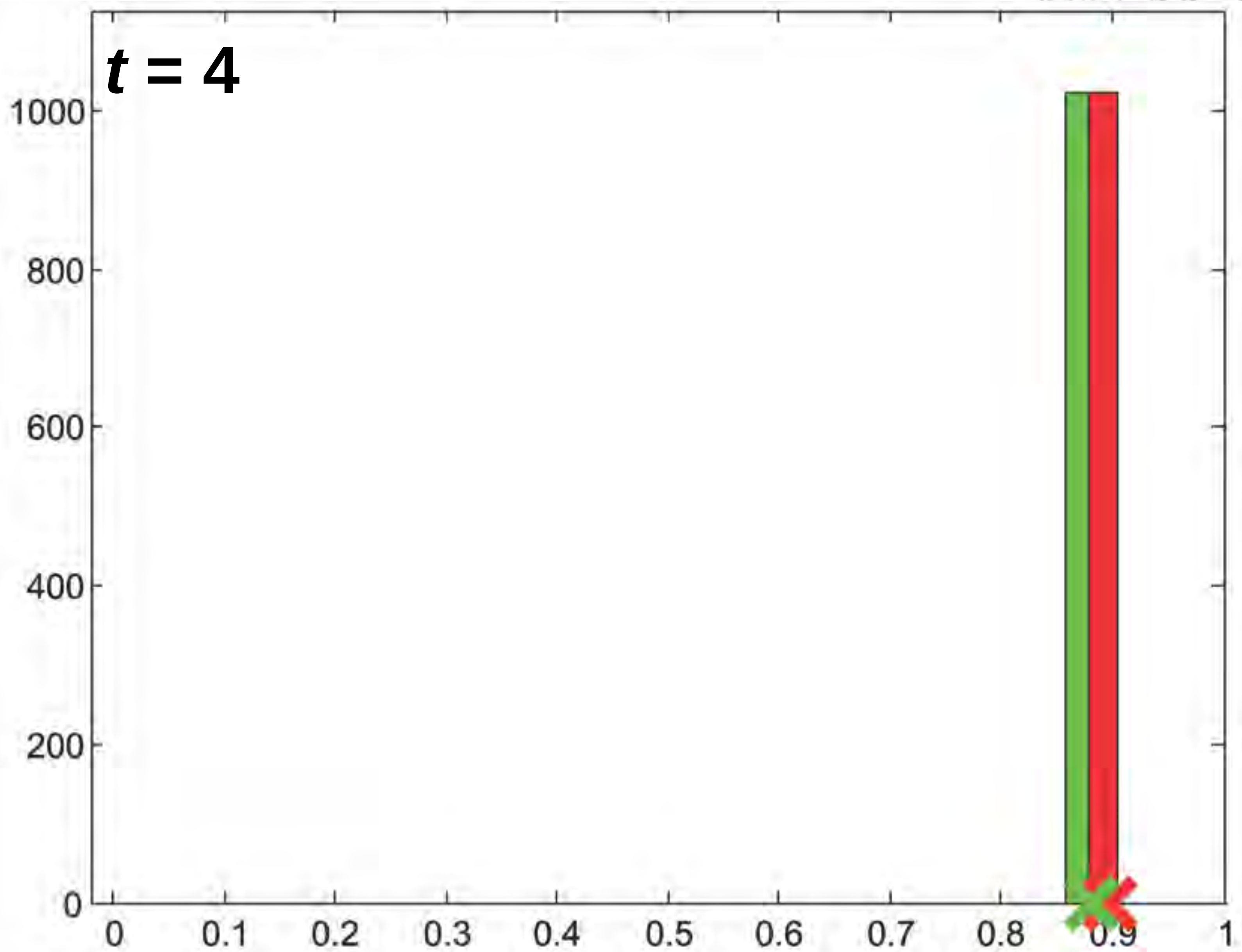

$p_t(x) := \phi_t[p_0(x)]$
$\rho_{t+1} = 4\rho_t(1-\rho_t)$
ϕ_t
$\Delta_U := U_T - U_A$
(X, ϕ_t, μ)

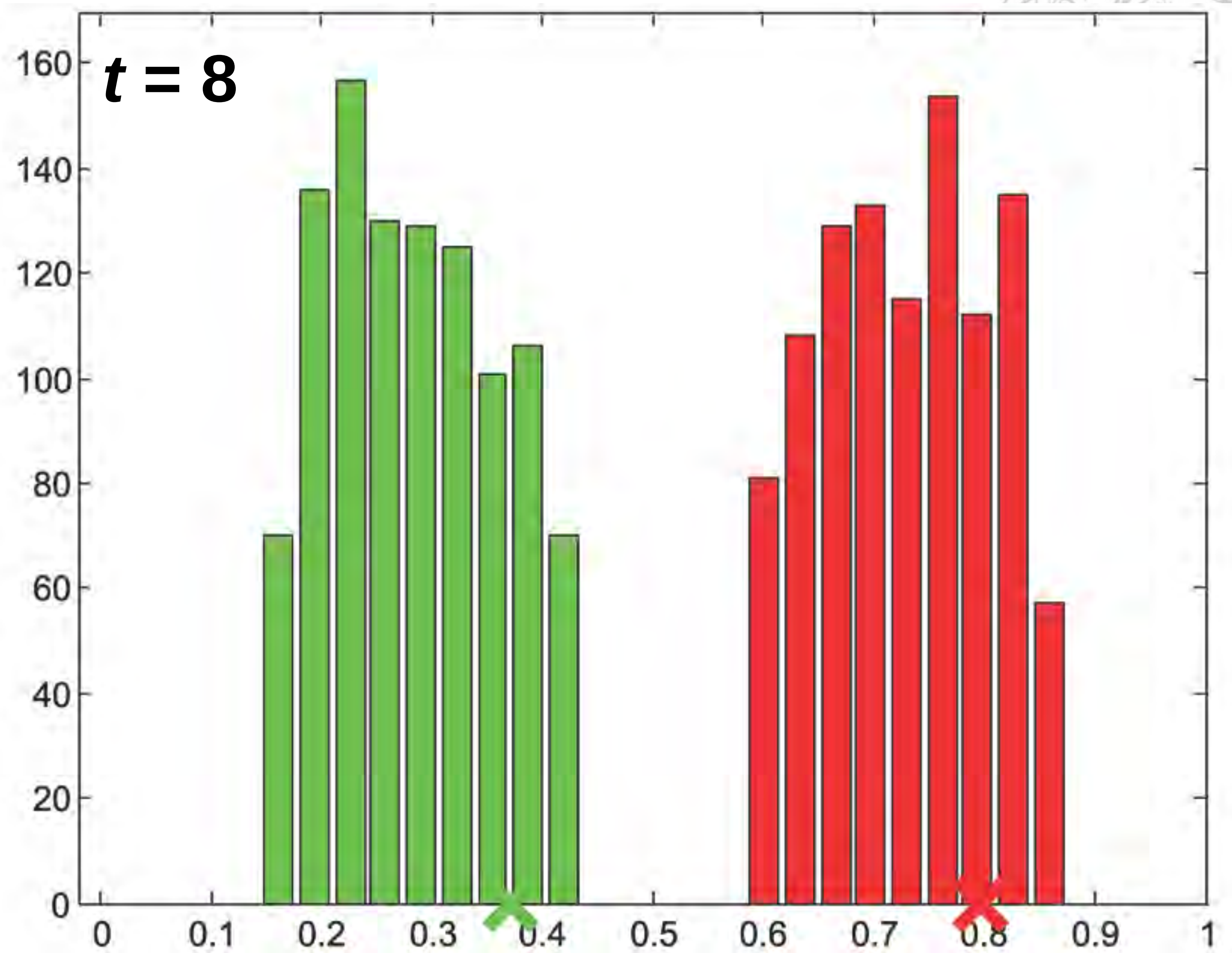
They all do the Calculation ….

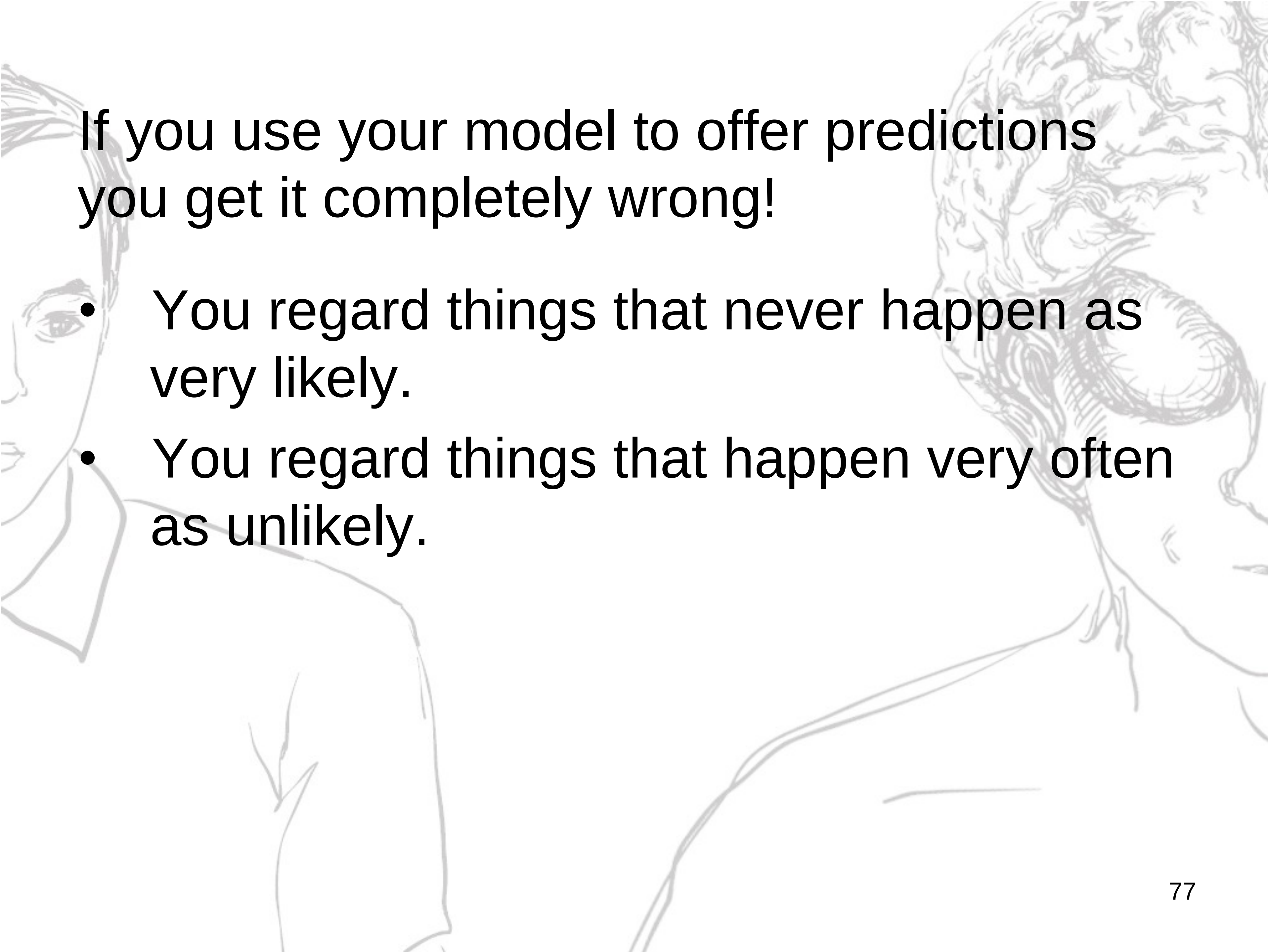


t = 0
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


t = 2
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


t = 4
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1


t = 8
160
140
120
100
80
60
40
20
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1



If you use your model to offer predictions you get it completely wrong!

- You regard things that never happen as very likely.
- You regard things that happen very often as unlikely.

And if you use your model to offer bets (or insurance policies) on certain events, you are losing money!

Probability p on event E: $p(E)$

Odds on E: $o(E) = 1/p$ → pay-out if E occurs

And if you use your model to offer bets (or insurance policies) on certain events, you are losing money!

Probability p on event E: $p(E)$

Odds on E: $o(E) = 1/p$ → pay-out if E occurs

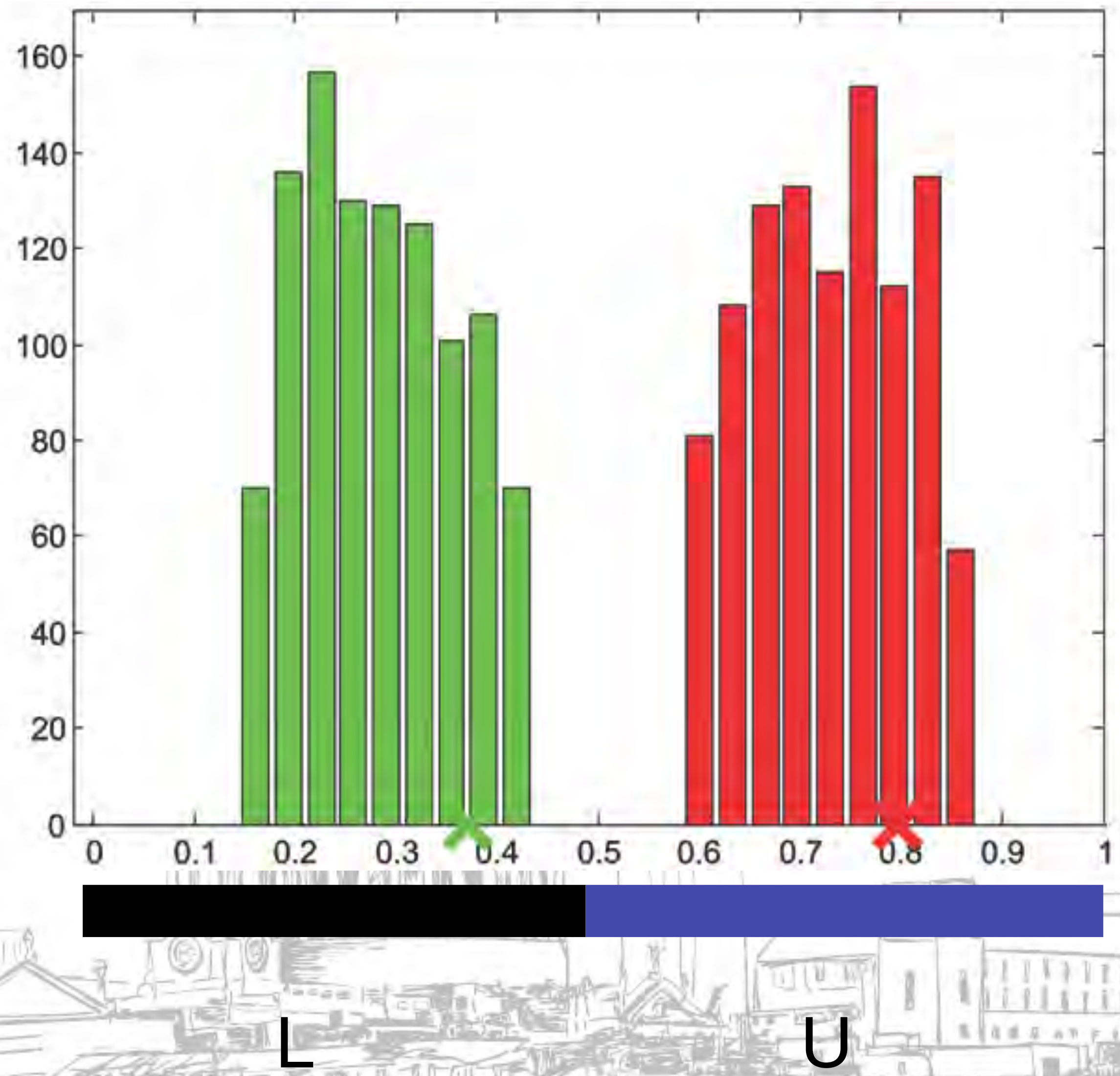
Example: coin

p or heads is ½.

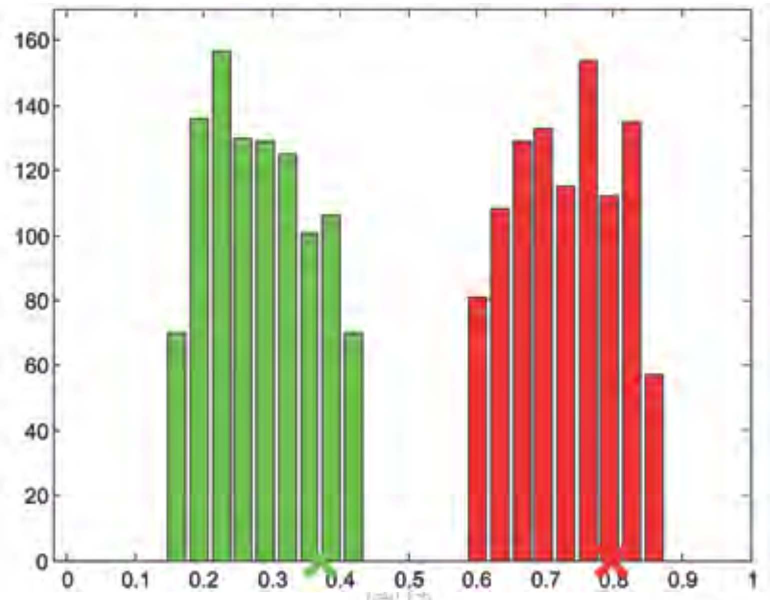
Odds on heads is 2.

If you bet £1 on heads and head occurs you get £2 back.

“Lower Half” against “Upper Half



“Lower Half” against “Upper Half

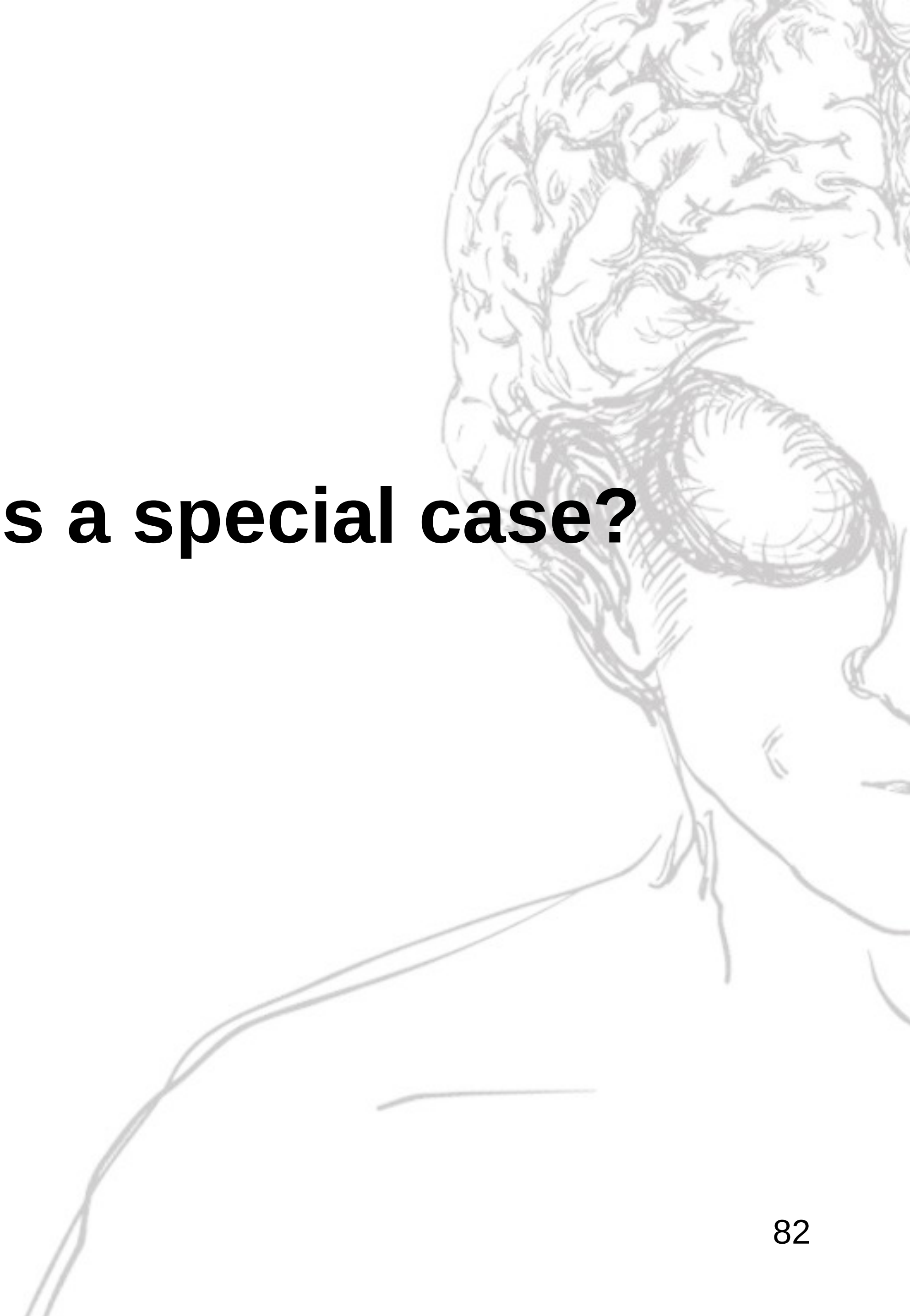


Model: $p(U) = 0$ and $o(U) \rightarrow \infty$

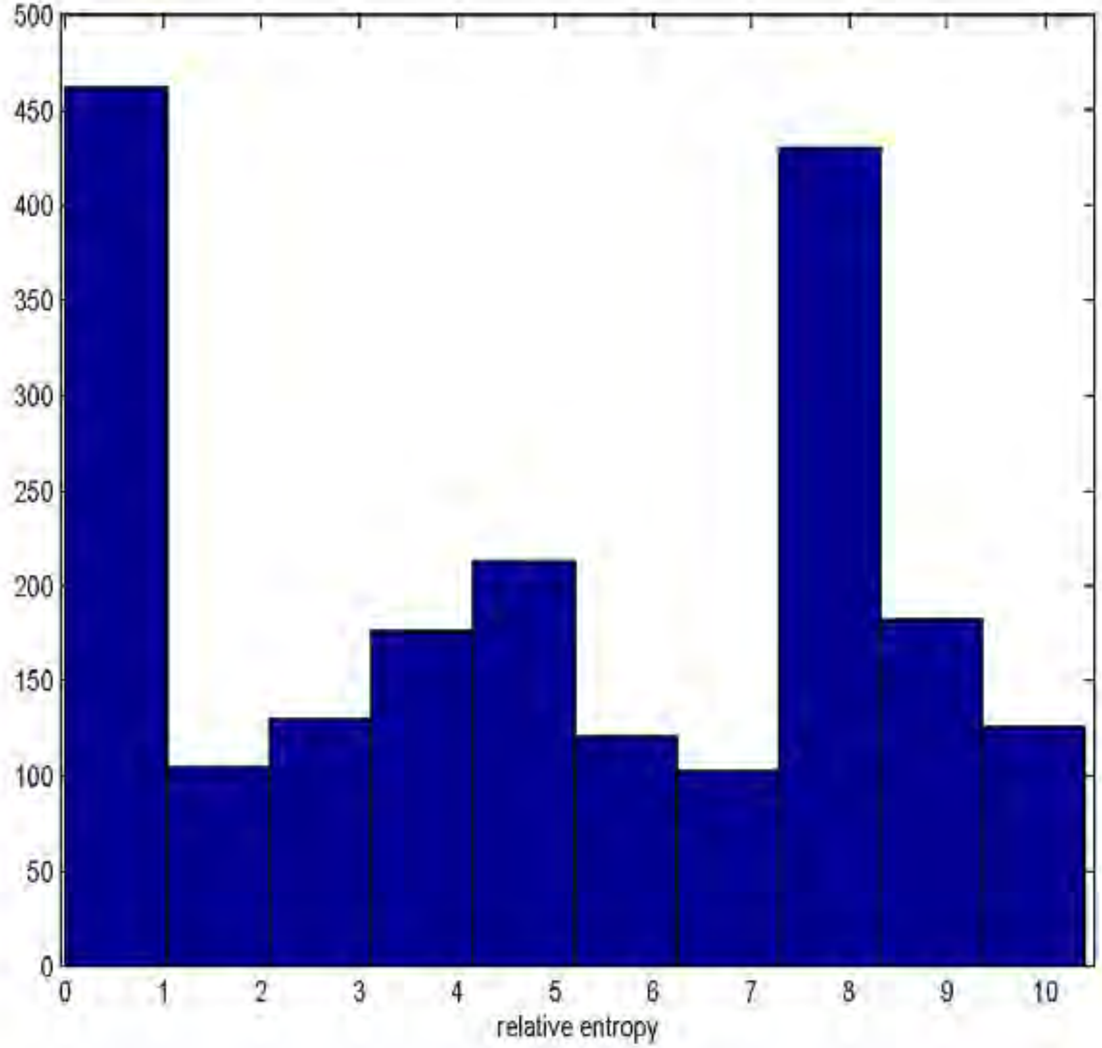
System: $p(U) = 1$

So U happens with probability 1 and you have to pay out infinite gains!

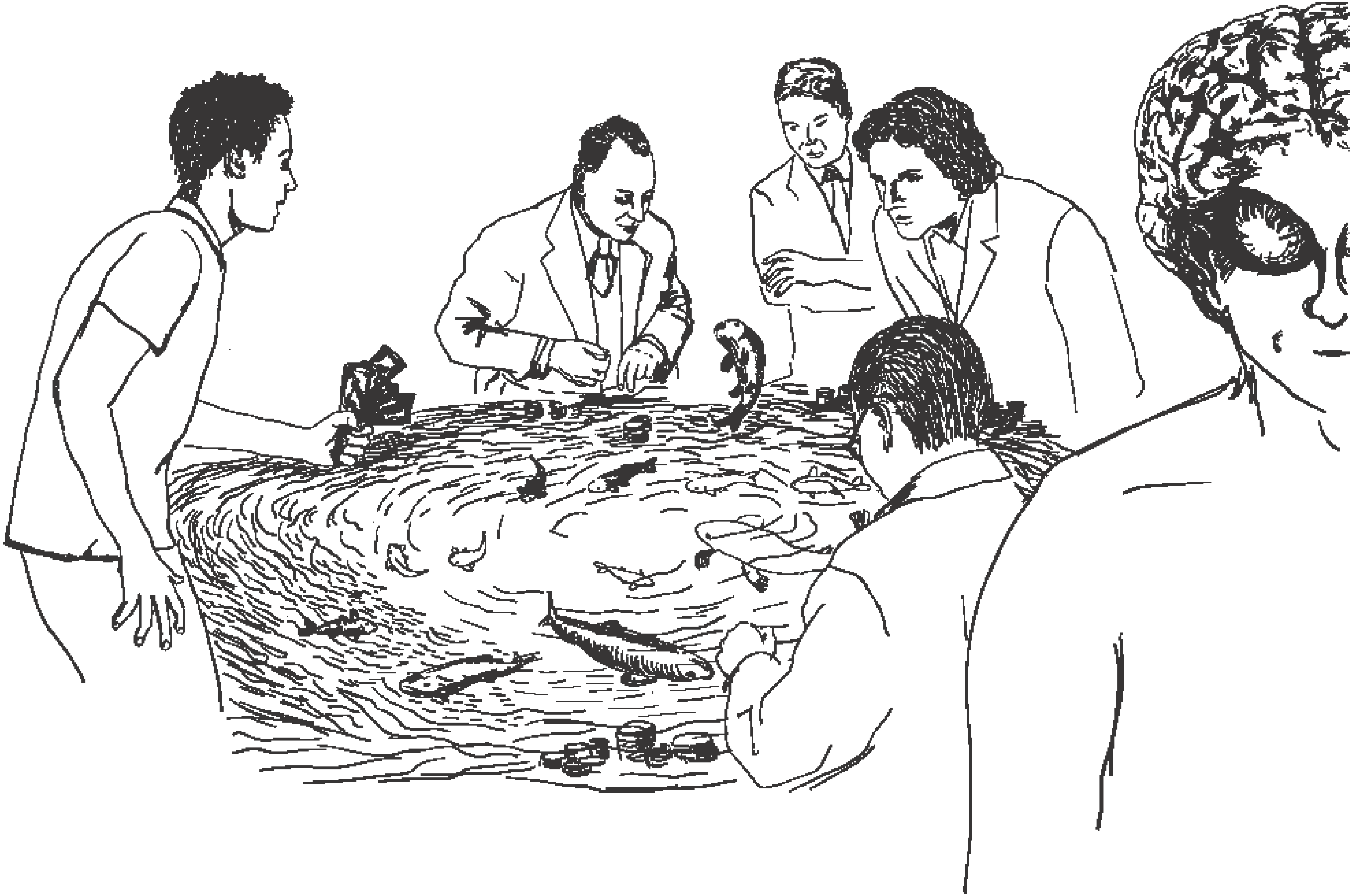
Question: is this a special case?



Relative Entropy of 2048 initial distributions (t=8)



The Pond Casino



Nine punters with £1000 each.

In every round they bet 10% of their wealth on events with probability in the interval:

1st Punter: [1/2, 1]

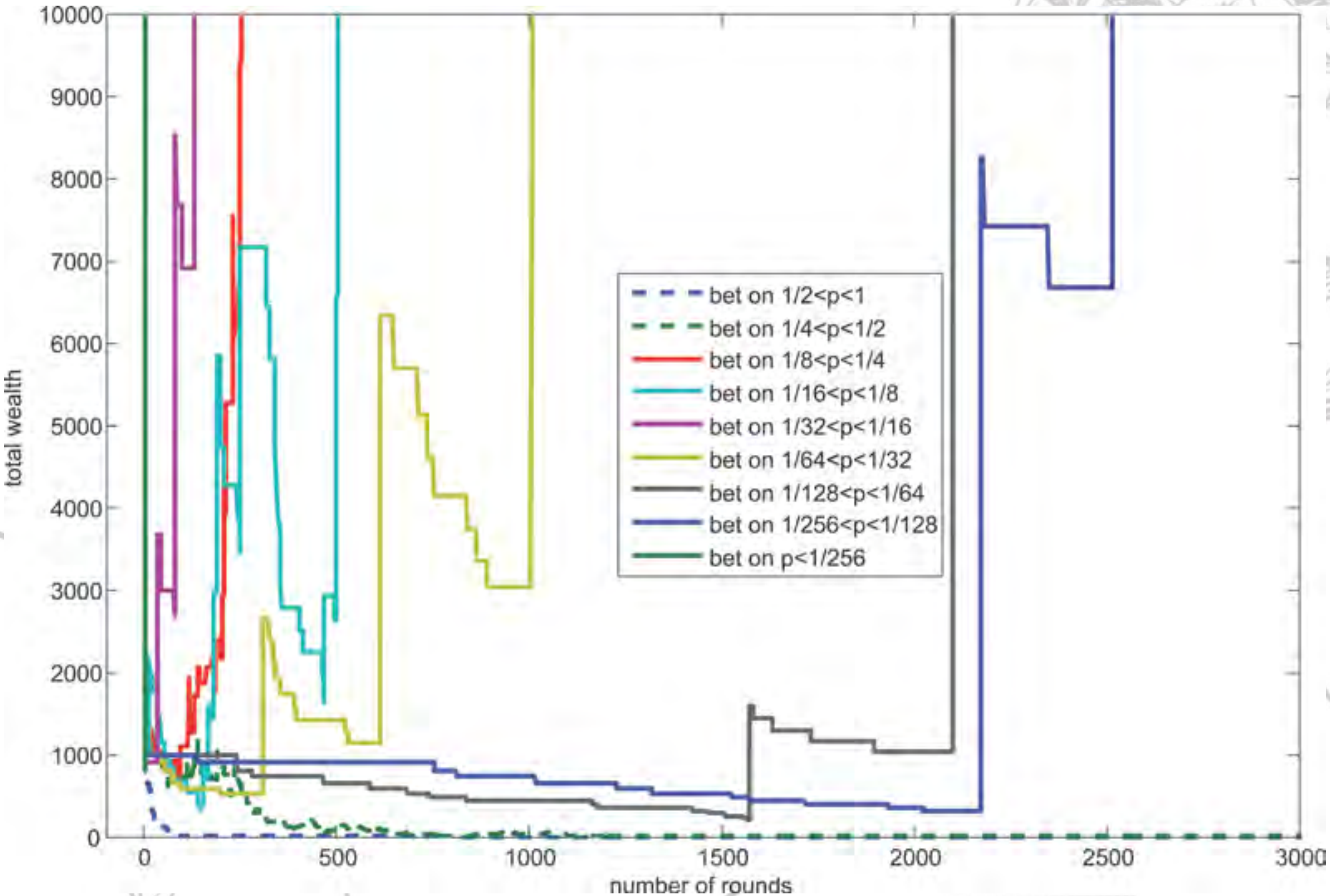
2nd Punter: [1/4), 1/2)

…

9th Punter: [0, 1/256)

How are they doing?

Punters' wealth



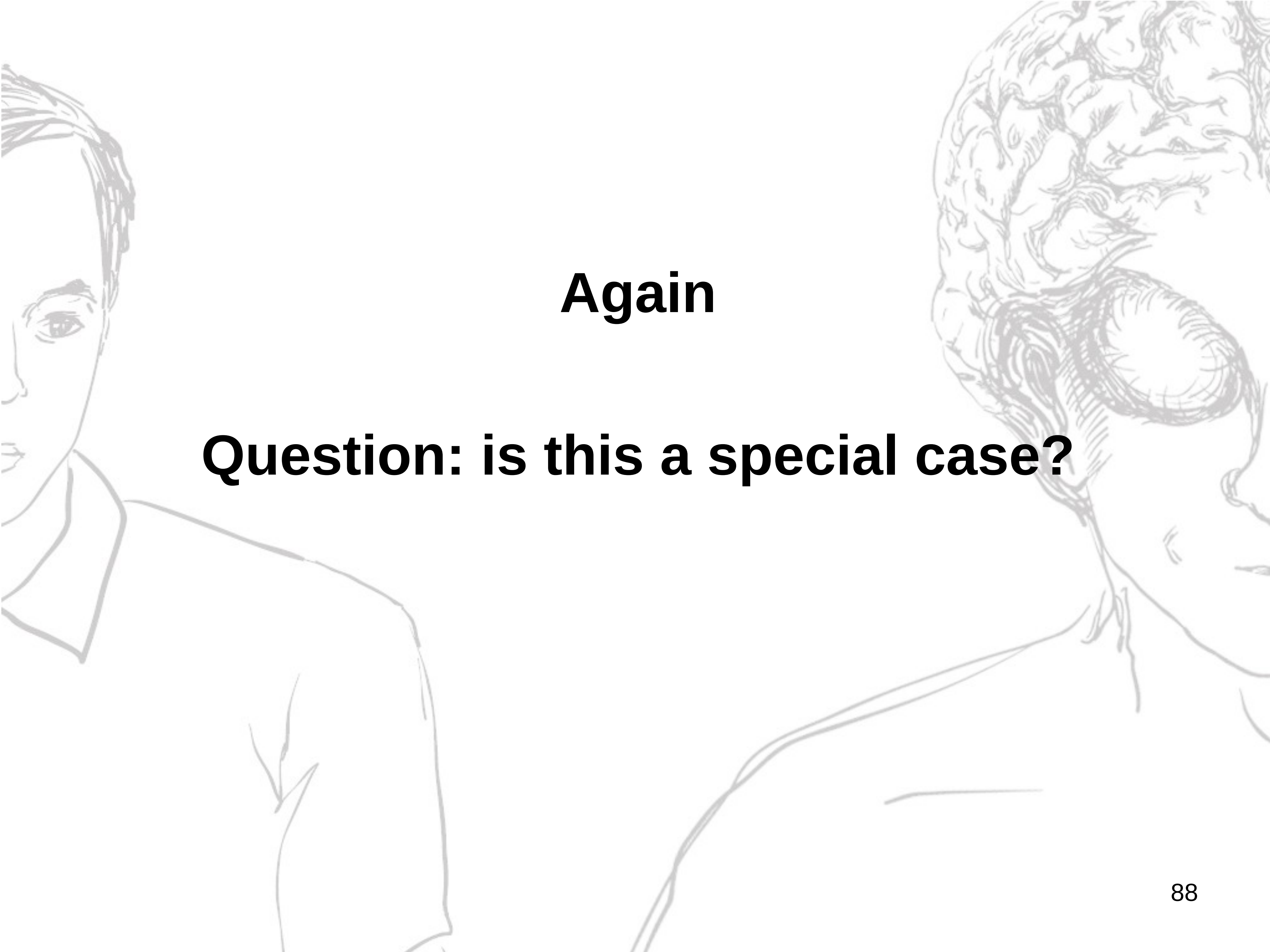
Time (Number of rounds played)

Result:

- 7 out of the 9 punters make enormous gains!
- The casino runs up huge losses.

→ Insurance companies …

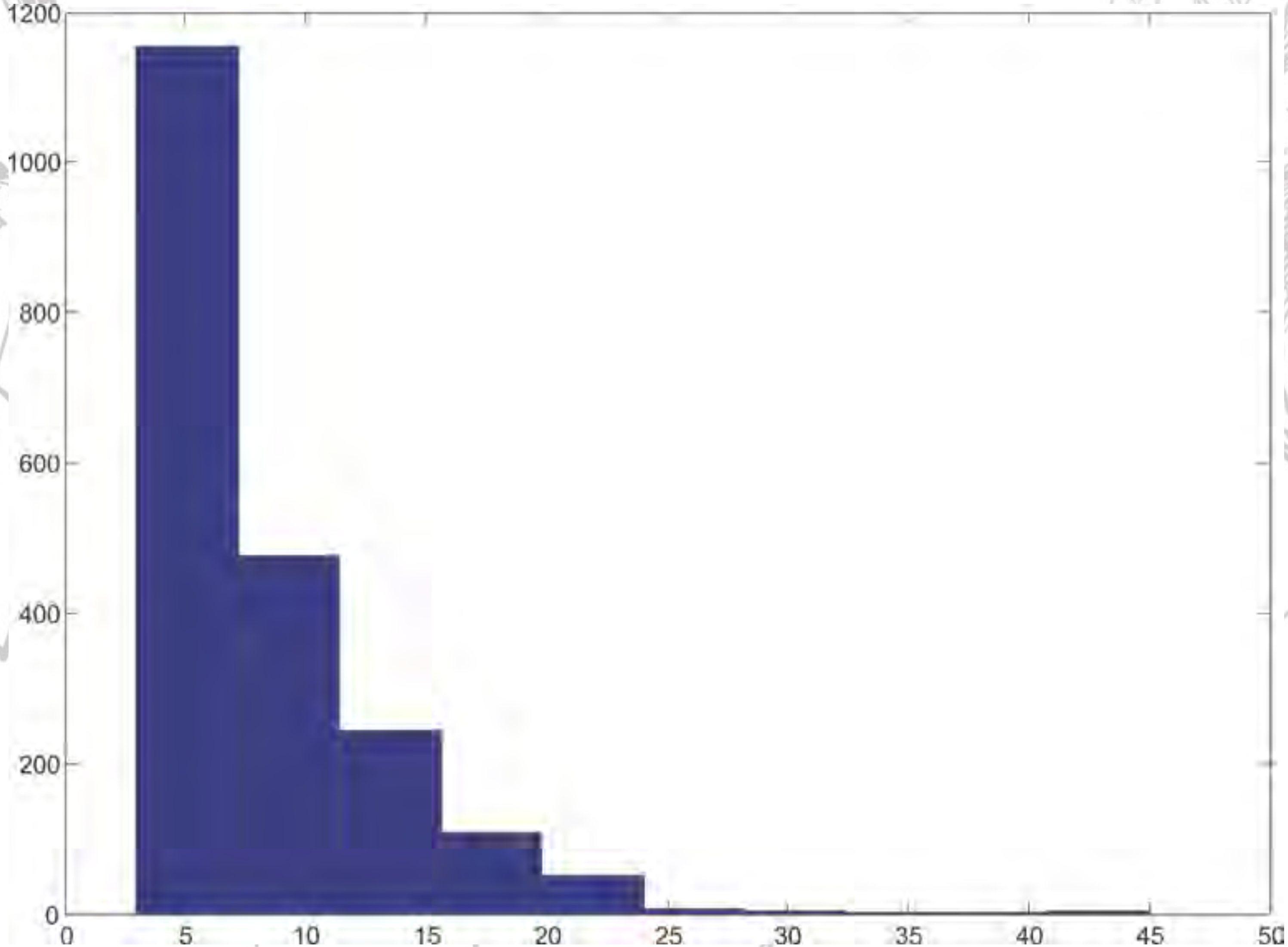
But: is this just a bad “bad luck event”?



Again

Question: is this a special case?

Time to bust for 2048 casinos:



Conclusion:

Even though the model is very close to the truth, it provides ruinous predictions!

Hence: If chaotic models have even the slightest model error, their capacity to make meaningful (and policy relevant!) probabilistic forecasts is lost.

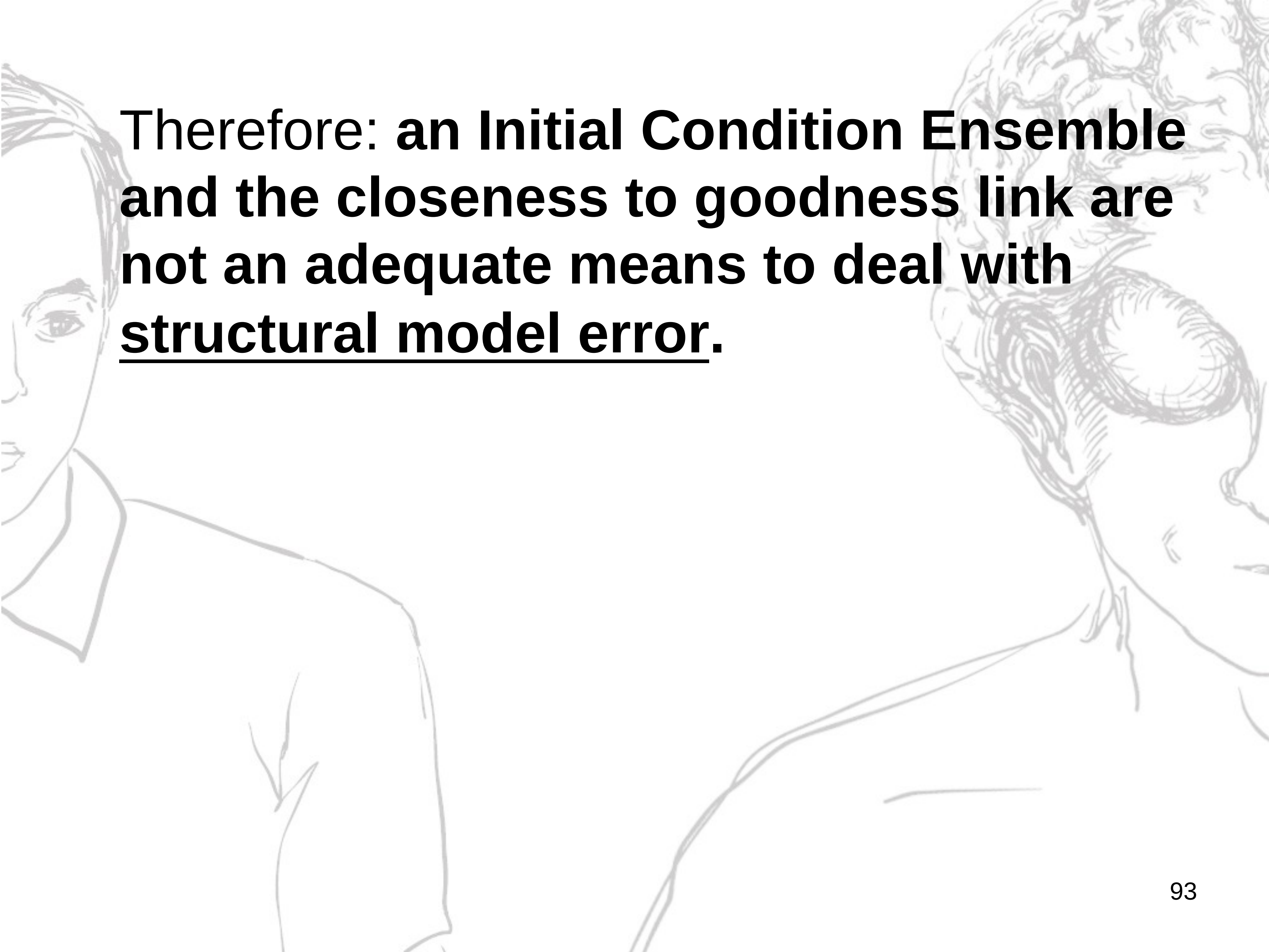
Conclusion:

Even though the model is very close to the truth, it provides ruinous predictions!

Hence: If chaotic models have even the slightest model error, their capacity to make meaningful (and policy relevant!) probabilistic forecasts is lost.

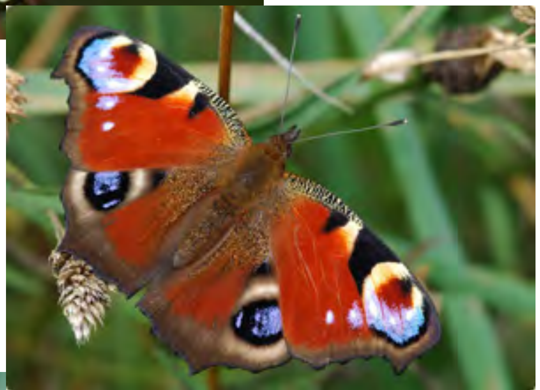
The closeness-to-goodness link is wrong!

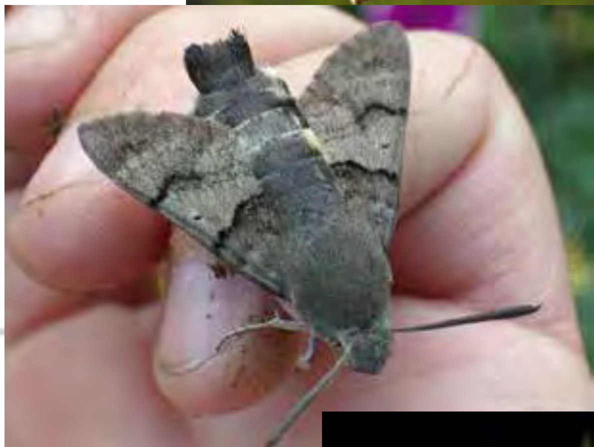
The failure of the closeness-to-goodness link gives raise to the ***hawkmoth effect***: the smallest deviation in model structure leads to completely different results, both for deterministic *and* probabilistic forecasts.

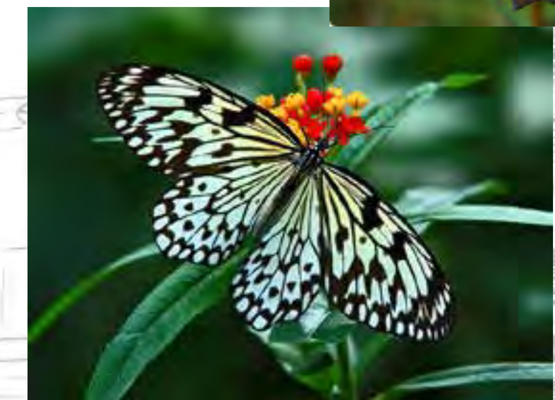


Therefore: **an Initial Condition Ensemble and the closeness to goodness link are not an adequate means to deal with <u>structural model error</u>.**

Or: butterflies are pretty; hawkmoths are ugly.







Irrelevance?

These considerations are relevant in some applied contexts.

For instance: UKCP09.

For details see:

Frigg, Stainforth and Smith:
'The Myopia of Imperfect Climate Models: The Case of UKCP09', forthcoming in *Philosophy of Science* (December 2013)

Reinventing the wheel?

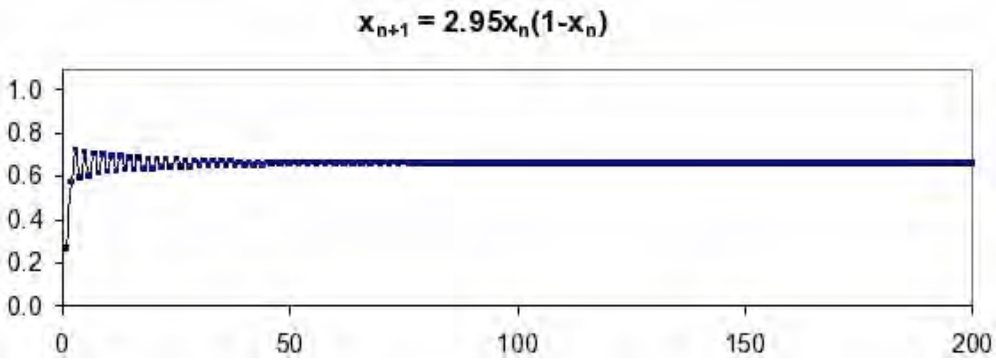
Feigenbaum’s classical discussion:

$$\rho_{t+1} = \alpha \rho_t (1 - \rho_t)$$

Parameter:

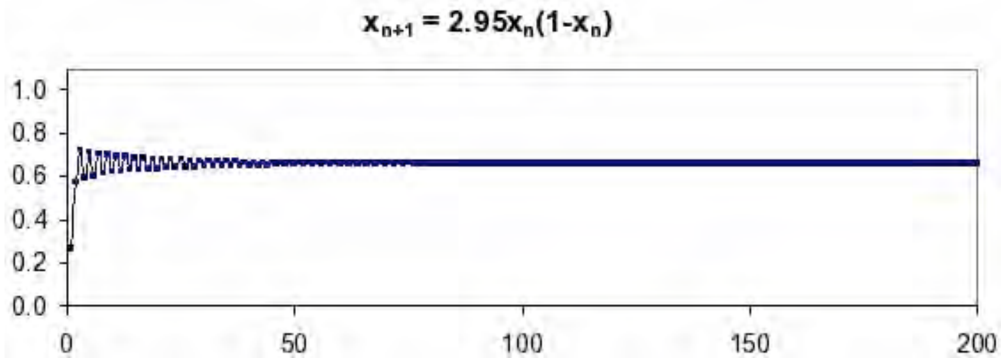
$$\alpha \in [0, 4]$$

Time series for different parameter values:

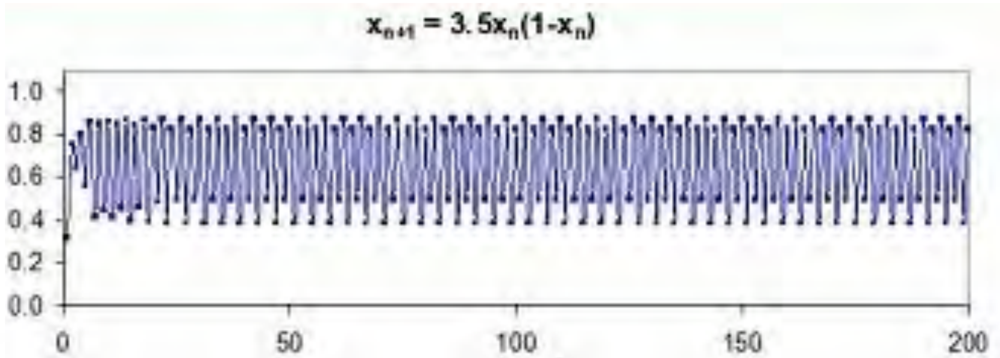


$\alpha = 2.95$

Time series for different parameter values:

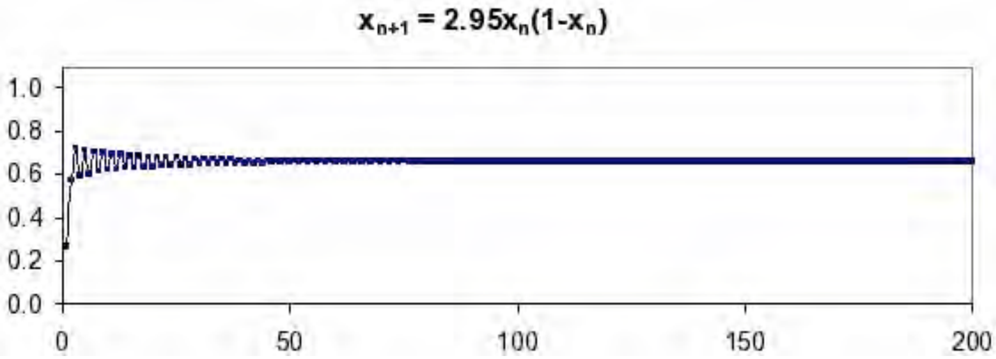


$\alpha = 2.95$

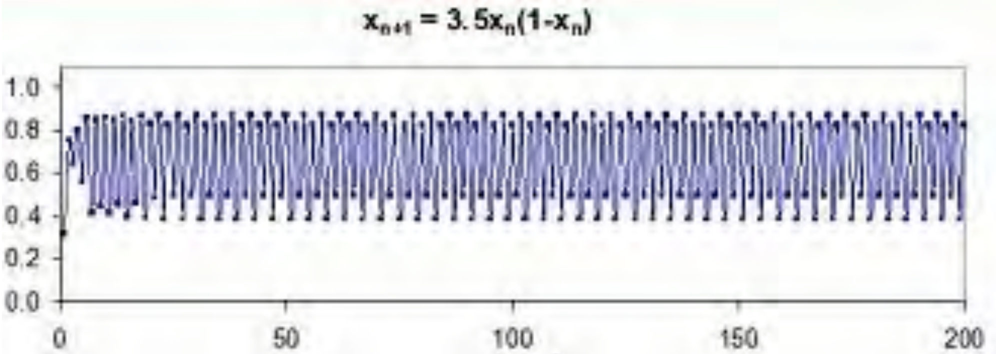


$\alpha = 3.5$

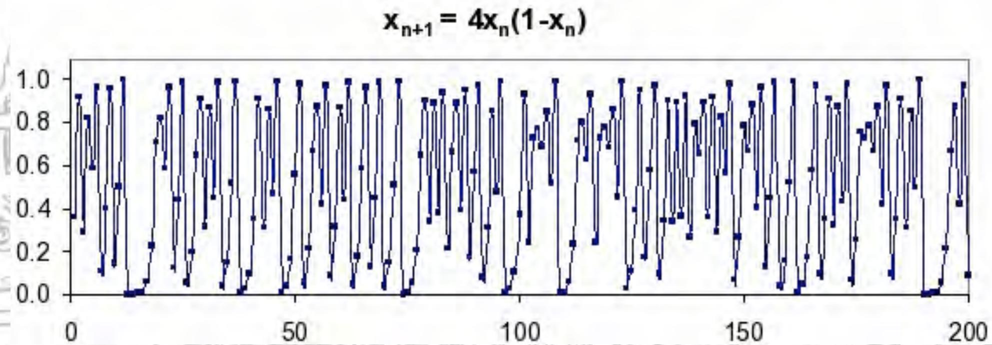
Time series for different parameter values:



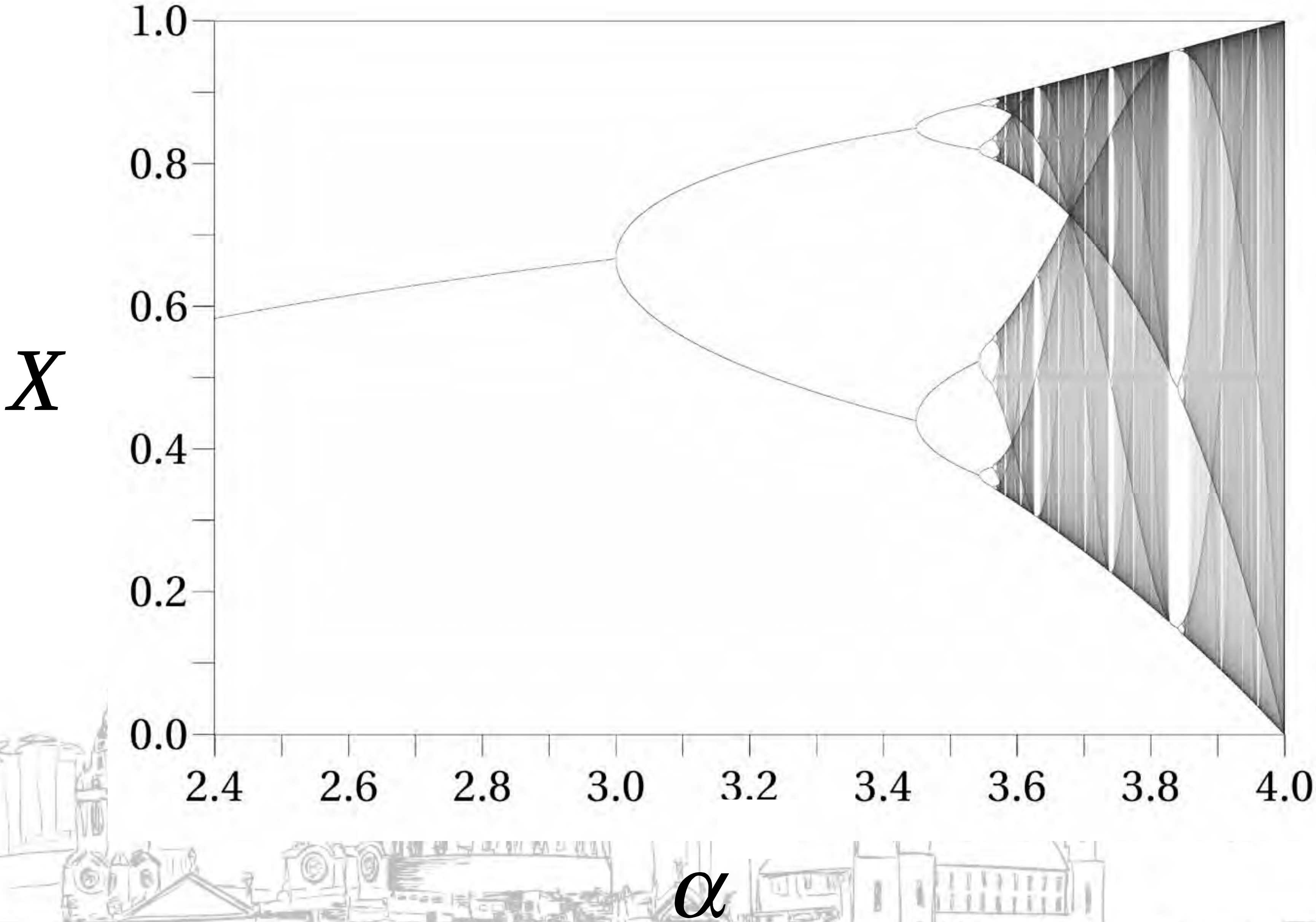
$\alpha = 2.95$

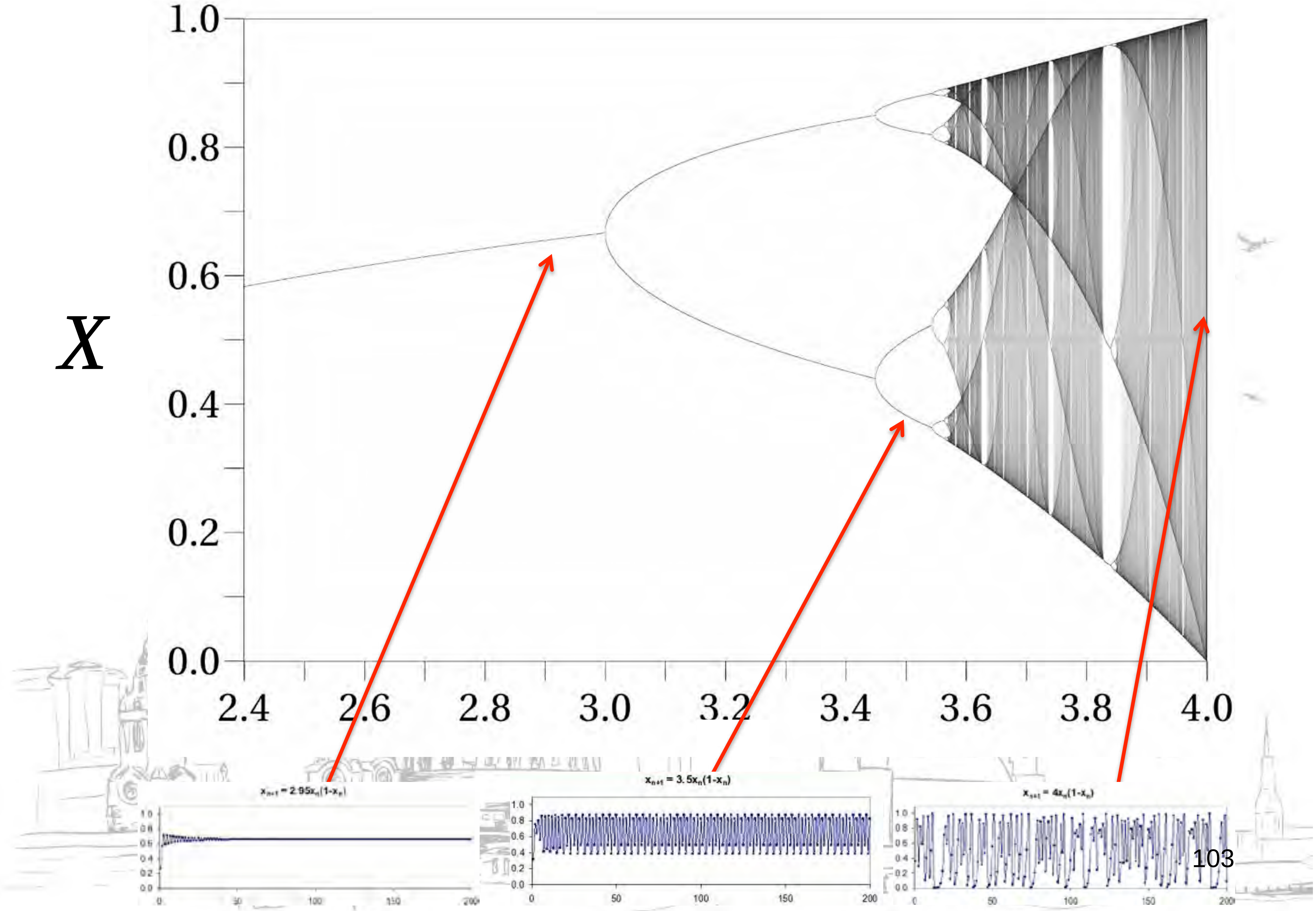


$\alpha = 3.5$



$\alpha = 4$


1.0
0.8
0.6
0.4
0.2
0.0
X
2.4
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0
α



X
1.0
0.8
0.6
0.4
0.2
0.0
2.4
2.6
2.8
3.0
3.2
3.4
3.6
3.8
4.0
$x_{n+1} = 2.95x_n(1-x_n)$
$x_{n+1} = 3.5x_n(1-x_n)$
$x_{n+1} = 4x_n(1-x_n)$

This is a study of **parameter variation**.

It provides information about what happens if we are uncertain about parameter values.

But: it provides **<u>no</u>** information about what happens when we are **uncertain about the model structure**.

What if the true equation is not exactly

$$\rho_{t+1} = \alpha \rho_t (1 - \rho_t) \,?$$

Overselling an example?

Recall our conclusion: the closeness to goodness link is not an adequate means to deal with structural model error.

Why is this a ***general*** problem and not just a problem of our example?

There is an elaborate mathematical theory of structural stability:

Andronov and Pontrjagin, Peixoto, Palis, Smale, Mañé, Hayashi.

But:

Stability proofs are forthcoming only for two-dimensional flows!

But that is a very special kind of system!

In general the situation is more involved:

Axiom A: the system is uniformly hyperbolic.

Strong transversality condition: stable and unstable manifolds must intersect transversely at every point.

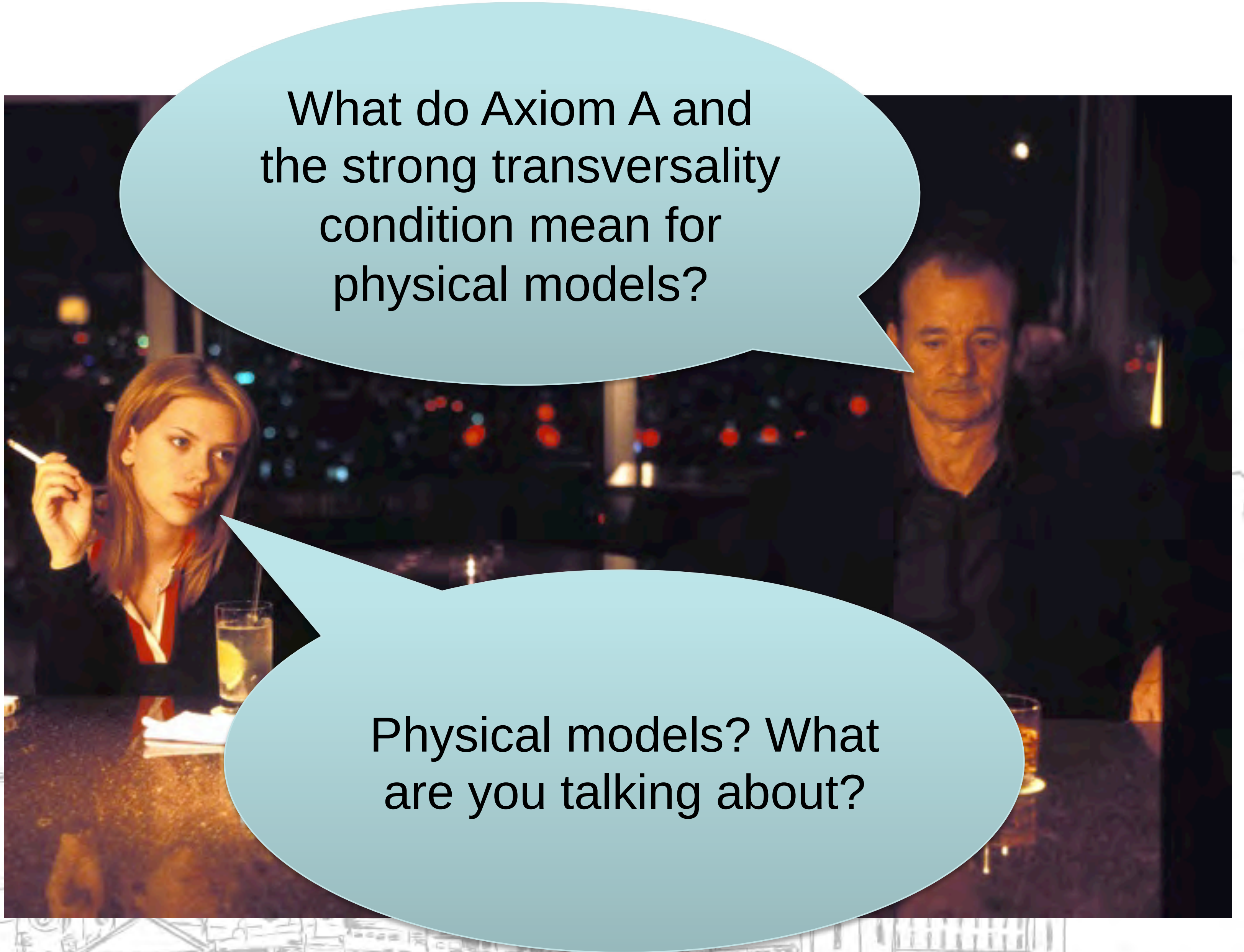
Palis and Smale (1970) conjectured that a system is structurally stable iff it satisfies Axiom A and the strong tranversality condition.

Proofs:

Mañé (1988) for maps

Hayashi (1997) for flows.


What do Axiom A and the strong transversality condition mean for physical models?


What do Axiom A and the strong transversality condition mean for physical models?
Physical models? What are you talking about?

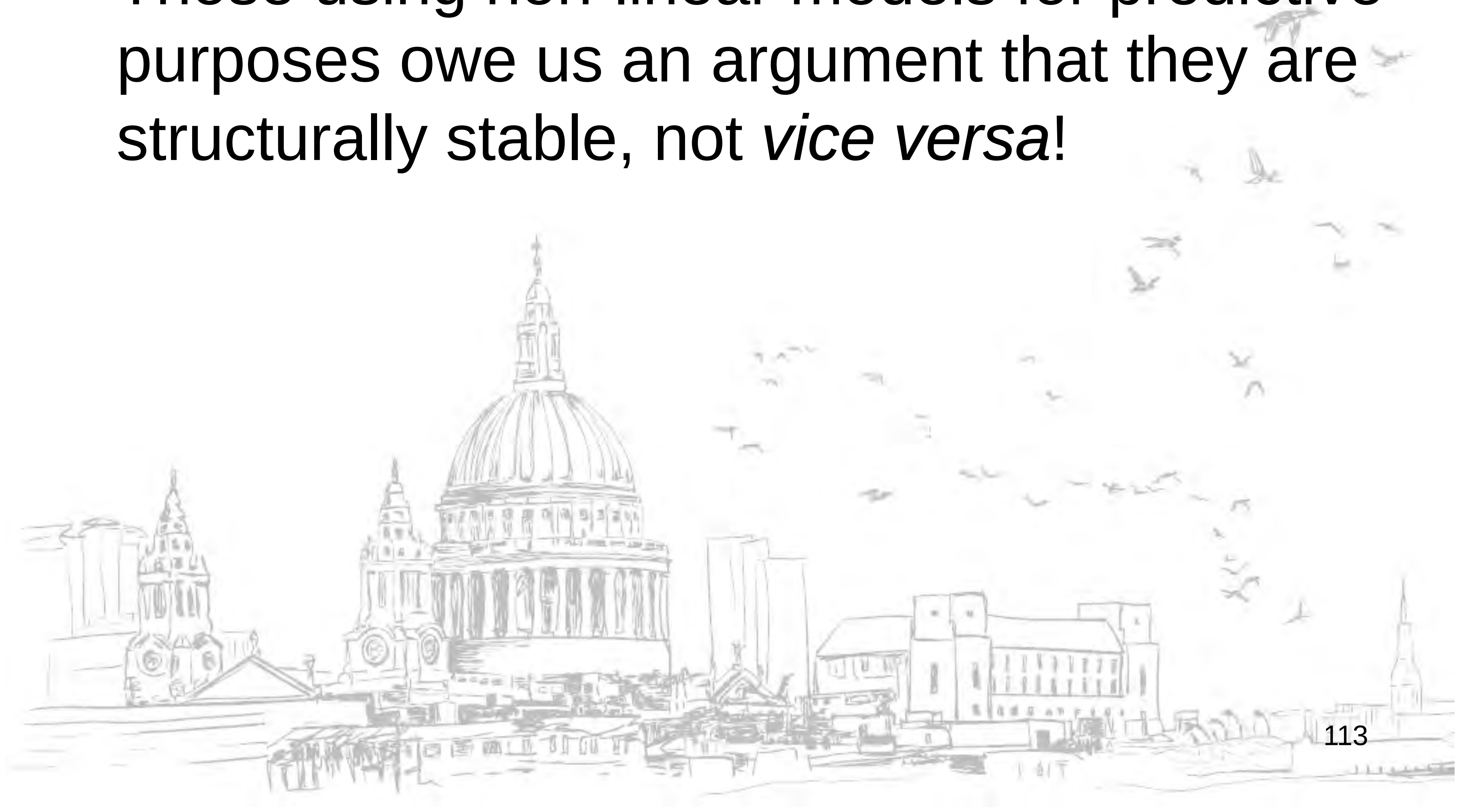
But:

Smale (1966): structural stability is not generic in the class of diffeomorophisms on a manifold: the set of structurally stable systems is open but not dense.

Smith (2002) and Judd and Smith (2004): if the model’s and the system’s dynamics are not identical, then ‘no state of the model has a trajectory consistent with observations of the system’ (2004, 228).

Minimal conclusion: shift of the onus of proof!

Those using non-linear models for predictive purposes owe us an argument that they are structurally stable, not *vice versa*!



Models in action

(ultra-short version)

Points to consider:

- Time scales
- Natural measures
  - but are they structurally stable?
  - do they exist at all (transient systems!)
- Aims of modelling: explanation and understanding versus exact prediction.

Thank you!