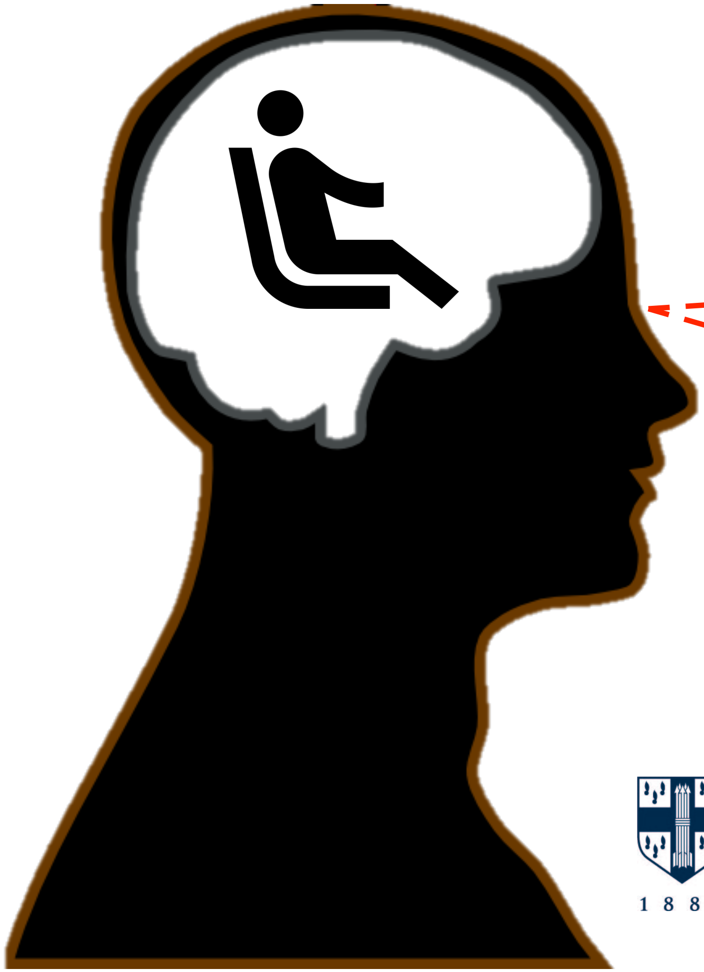
Toward a formal model of free will



Ian Durham

SAINT ANSELM COLLEGE
1889

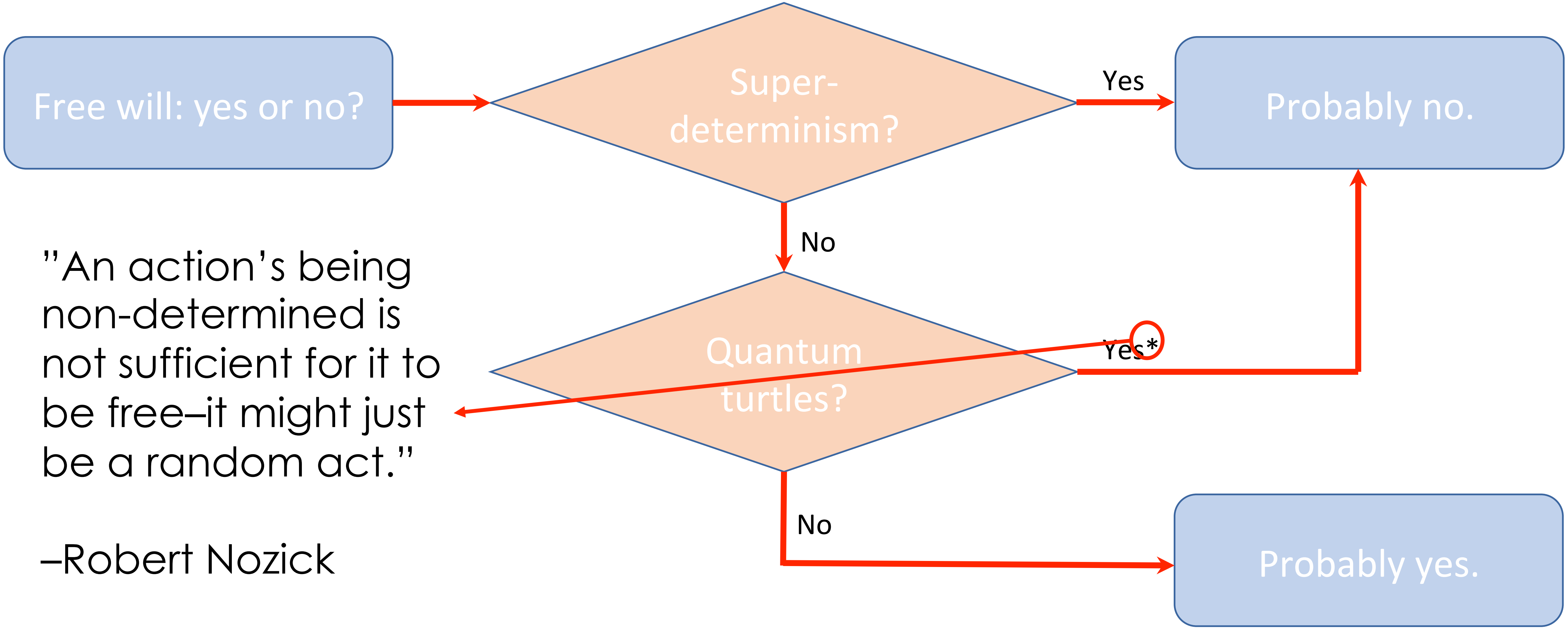
Models of Consciousness

September 9-12, 2019, Oxford University

Speakers include: Roger Penrose, Adrian Kent, Stuart Hameroff, Peter Grindrod, Yakov Kremnitzer, Jonathan Mason, Mauro D'Ariano, William Marshall, Chetan Prakash, Ian Durham, Johannes Kleiner

http://www.models-of-consciousness.org

Simplified free will flow chart



"An action's being non-determined is not sufficient for it to be free–it might just be a random act."

–Robert Nozick

The essence of free will

Carrots?

Peppers?

A continuum of processes

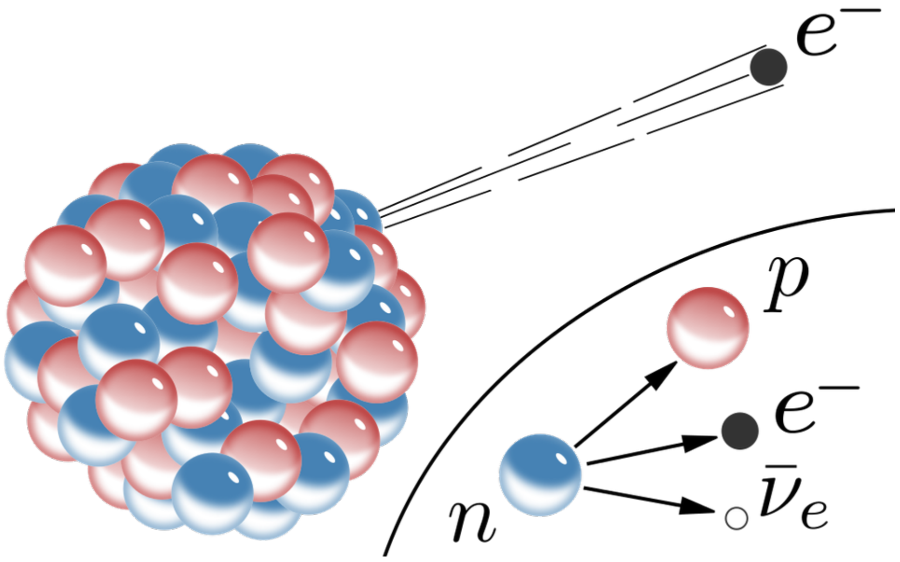
Deterministic process: A (fully) deterministic process is one for which there is only one possible outcome.

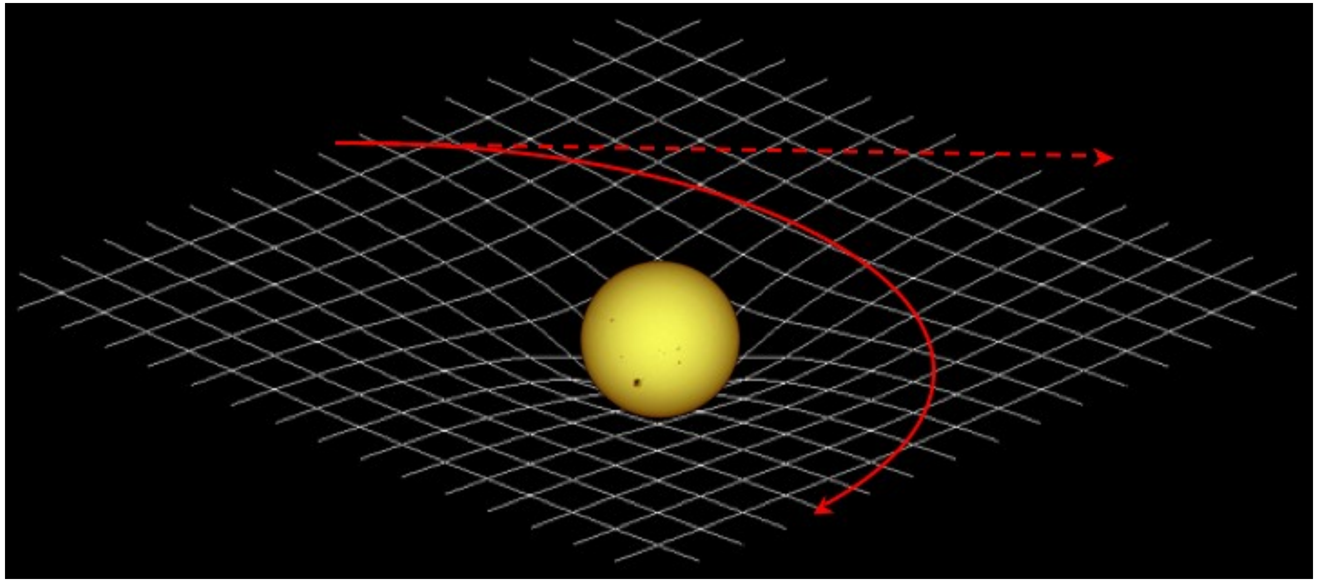
Random process: A (fully) random process is one for which there is more than one possible outcome and all possible outcomes are equally probable.

The essence of free will

- Free will is based on an agent's ability to make certain choices freely.
- A choice can only be said to be free if the agent can make some judgment about all possible choices in order to weigh them against one another, i.e. they must *mean* something to us. Otherwise the choice is random (and thus meaningless).
- The agent must have a high degree of confidence that the choice will be realized.
- An agent for whom a certain percentage of choices is deemed to be free, is said to have free will.

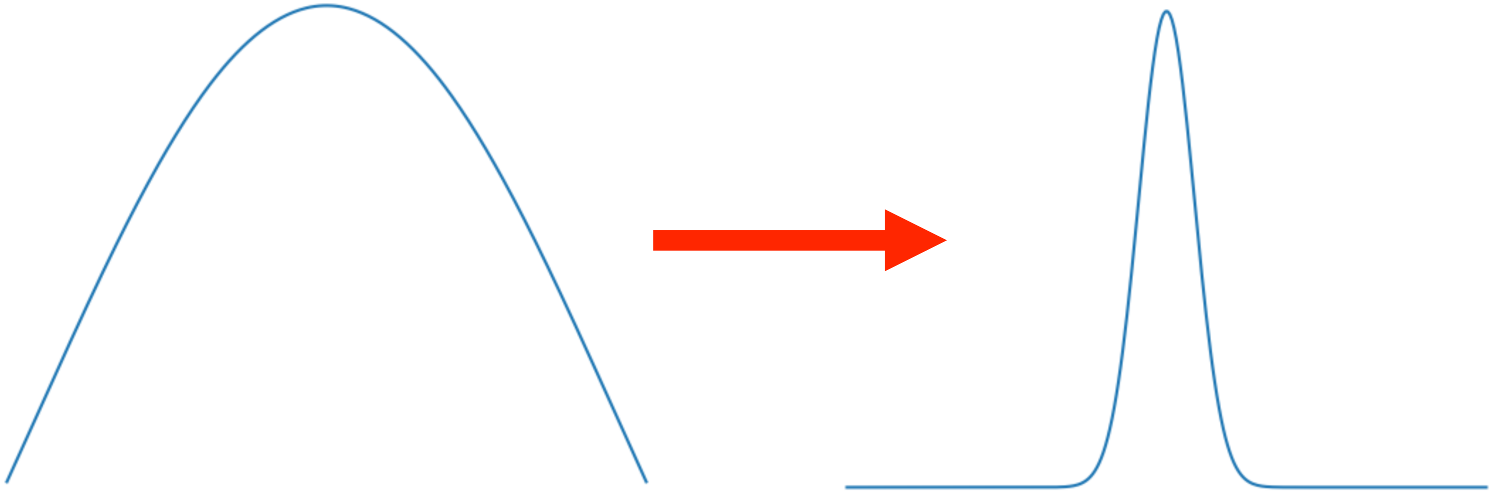
Reality: a continuum of models





Random → Deterministic

Determinism *from* randomness



Interacting systems of random processes can converge to deterministic macrostates if the multiplicity is sufficiently high.

Building a model: assumptions

- The most fundamental systems are irreducible to other systems, i.e. they contain no interactions and cannot be partitioned.
- The microstates (possible configurations) of a fundamental system are all equally likely in the long run.
- A (possible) choice is a (macro) process that takes a system from one macrostate to another; *different choices represent different processes*.
- A system's macrostates are formed from interacting microprocesses.

Building a model: assumptions

- The probability that a choice will lead to its macrostate is arbitrarily high if the choice is free, i.e. a free choice is nearly deterministic.
- The number n of possible choices that a system has must be small enough to be read into the system's memory in a finite time.
- The choices available to a system are determined by a combination of environmental forcing and internal dynamics.
- The choices do not change prior to the decision by the agent.

Building a model: formalism

Consider a set of choices with distribution functions

$$P_1(\mathbf{x}; \boldsymbol{\mu}_1, \Sigma_1), \ldots, P_n(\mathbf{x}; \boldsymbol{\mu}_n, \Sigma_n)$$

where $\boldsymbol{\mu}_i$ is the mean and Σ_i is the variance.

The choices can be represented as a mixed distribution

$$F(\mathbf{x}) = w_1 P_1(\mathbf{x}; \boldsymbol{\mu}_1, \Sigma_1) + w_2 P_2(\mathbf{x}; \boldsymbol{\mu}_2, \Sigma_2) + \cdots + w_n P_n(\mathbf{x}; \boldsymbol{\mu}_n, \Sigma_n)$$

convex

with weights $w_i \geq 0, \quad w_1 + w_2 + \cdots + w_n = 1.$

Building a model: analysis

The distance between any two choices in the distribution is given by the Mahalanobis distance

$$d_M(\boldsymbol{\mu}_i, \boldsymbol{\mu}_j, \boldsymbol{\Sigma}) = \sqrt{(\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_i)}$$

where $\boldsymbol{\Sigma}$ is the covariance matrix.

Assumption: the larger this value, the more distinct the choices

A measure of free choice

Given a finite number of possible choices *n*, the ‘freedom’ of choice *i* is given by the function

$$\zeta_i\left(d_M, \tau(n), \Sigma_i\right) \propto \{d_M\}_{\min}\tau(n)^{-1}\Sigma_i^{-1}$$

where $\tau(n) < \infty$ is the time it takes to read all *n* choices into the system’s memory and $\{d_M\}_{\min}$ is the minimum distance to another choice among all such choices in the ensemble.

A measure of free will

Given some number of processes that result in choices, a system's free will is given by a partition function

$$Z(\zeta) = \prod_{i=1}^{n} \zeta_i$$

i.e. the level of free will depends in some way on the freedom of the choices under consideration.

Expectations

We expect that the level of free will and thus the value of Z will increase with increasing system complexity.

Questions

1. Can this model produce real, measurable results?
2. Does this model necessarily presuppose a dualist view of consciousness?
3. Is this model compatible with existing formal models of consciousness?
4. Does this model make predictions that are consistent with the results of Bell inequality tests?

Acknowledgements and links

- Thanks to Robert Prentner for encouraging this project.
- Thanks to participants of the Modelling Consciousness Workshop (Dorfgastein, 2019) for helpful discussion.
- Thanks to Nate Durham for adding a key point on meaning.

- Early ideas:

  - "God's Dice and Einstein's Solids," in A. Aguirre, B. Foster, and Z. Merali (eds.), *Wandering Towards a Goal*, Springer, 2018.
  - "Boundaries of Scientific Thought," in I. Durham and D. Rickles (eds.), *Information and Interaction*, Springer, 2017.

Grazie!

SOCIETY FOR MIND-MATTER RESEARCH

SILICON VALLEY community foundation®

FQXi FOUNDATIONAL QUESTIONS INSTITUTE

SAINT ANSELM COLLEGE 1889