Consciousness and the Collapse of the Wave Function

David Chalmers [with Kelvin McQueen]

Two Questions

- What is the place of consciousness in nature?
- What is the reality behind quantum mechanics?

A Simultaneous Answer

- Consciousness collapses the quantum wave function
  - John von Neumann (1932)?
  - Eugene Wigner (1961)

Benefits

- This view potentially
  - solves the quantum measurement problem
    - gives a role for consciousness in nature
- But: the idea seems flaky, and has never been made rigorous.
- Can we do better?

Another Try

- This is an attempt to do better, using
  - a mathematical theory of consciousness
  - an account of quantum collapse dynamics
- Results are mixed, but interesting.

Outline

1. The problem of consciousness
2. Interpretation of quantum mechanics
3. Integrated information theory
4. Super-resistance and the Zeno problem
5. Pearle-style stochastic collapse
6. Empirical tests

Consciousness

- Consciousness: The subjective experience of the mind and the world
- A system is conscious if there is something it is like to *be* that system.

‘What is it like to be a bat?’

“… imagine that one has webbing on one’s arms, which enables one to fly around at dusk and dawn catching insects in one’s mouth; that one has very poor vision, and perceives the surrounding world by a system of reflected high-frequency sound signals; and that one spends the day hanging upside down by one’s feet in an attic. In so far as I can imagine this (which is not very far), it tells me only what it would be like for me to behave as a bat behaves. But that is not the question. I want to know what it is like for a *bat* to be a bat.”

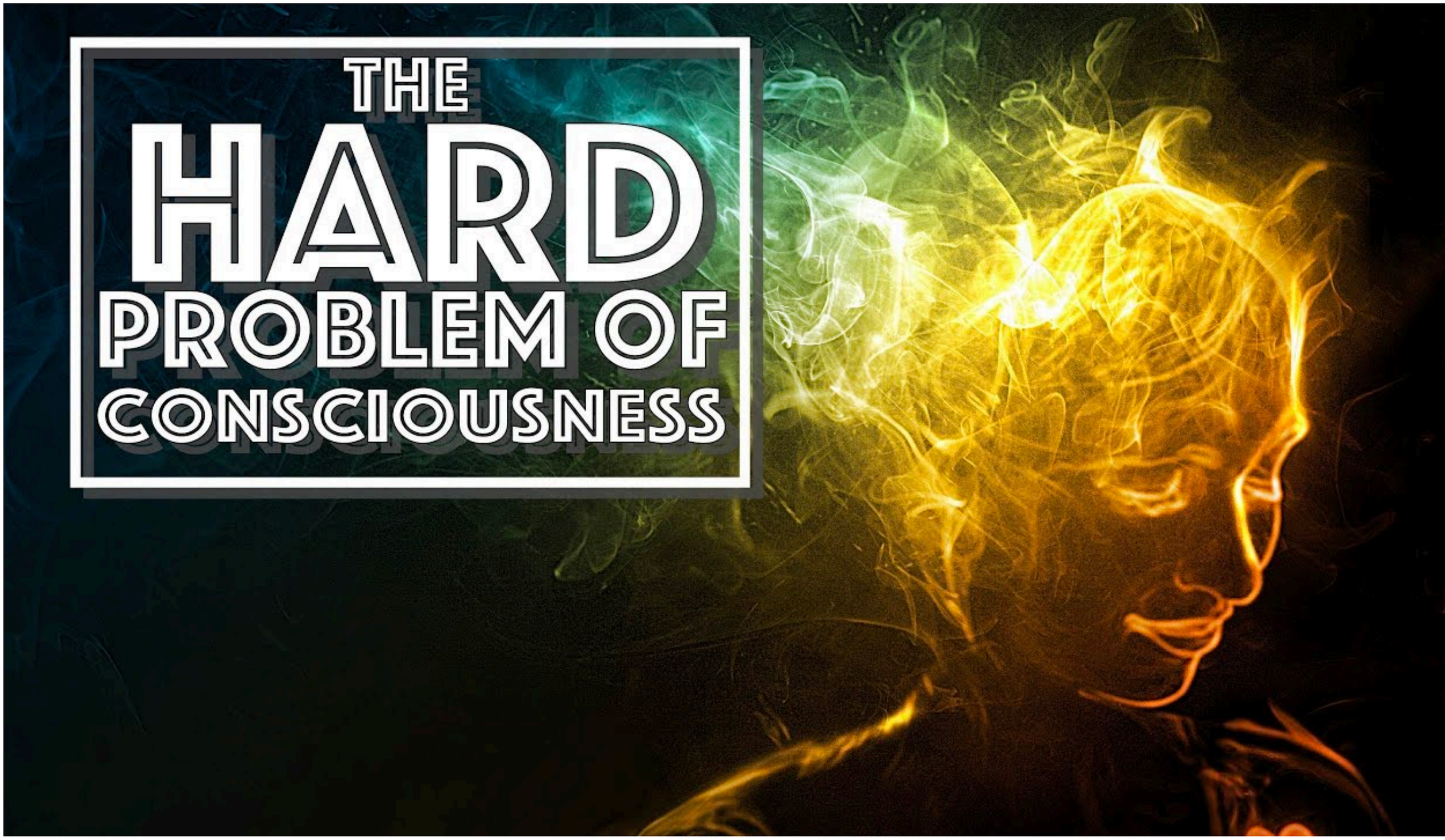
Thomas Nagel, ‘What is it like to be a bat?’ (1974)

Examples

Visual experiences.	e.g. color, depth, …
Other sensory experiences	e.g. sound, taste, …
Bodily sensations	e.g. pain, orgasm, …
Mental imagery	e.g. recalled visual images, …
Emotional experiences	e.g. happiness, anger, …
Stream of occurrent thought	e.g. reflection, decision, …

The Hard Problem of Consciousness

- How do physical processes in the brain give rise to conscious experience?


THE
HARD
PROBLEM OF
CONSCIOUSNESS

The Easy Problems

- The "easy" problems of consciousness: explaining behavioral/cognitive functions, e.g.
  - perceptual discrimination
  - integration of information
  - control of behavior
  - verbal report.

Functional Explanation

- We explain the easy problems by finding a neural or computational mechanism that performs the relevant function.

Explanatory Gap

- For the hard problem, explaining behavioral functions leaves open a further question:
  - Why is this accompanied by experience?
    - Why doesn't all this processing go on "in the dark", without consciousness?
- There seems to be an explanatory gap between physical processes and subjective experience.

Structure and Dynamics

- Problem: physical processes just explain structure and dynamics
- That suffices for the easy problems, but not for the hard problem.

Approaches to the Hard Problem

- Materialism (consciousness is physical):
  - Leaves an explanatory gap?
- Dualism: (consciousness is nonphysical):
  - How does consciousness interact?
- Panpsychism (consciousness is everywhere)
  - How does consciousness combine?
- Illusionism (consciousness is an illusion)
  - Denies a fundamental datum?

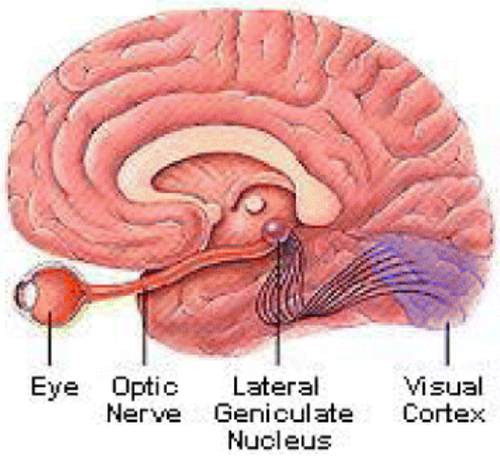
The Interaction Problem

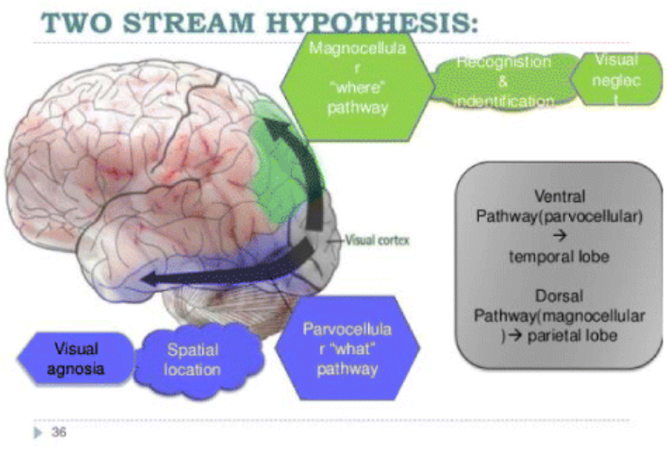
- How can consciousness affect the physical world?
- Physics seems to be causally closed, leaving no room for consciousness to play a role.

René Descartes and Princess Elisabeth

Interactionist Ontological Dualism

Violation of the causal closure of the physical world





Collapse to the Rescue?

- Potential loophole in causal closure of the physical world: wave function collapse.
- Can collapse provide a causal role for consciousness that is consistent with physics?

Quantum Mechanics

- Quantum mechanics postulates a wavelike reality where things don't have definite properties, but we experience a world with definite properties.
- How can this be explained?

Second derivative with respect to X

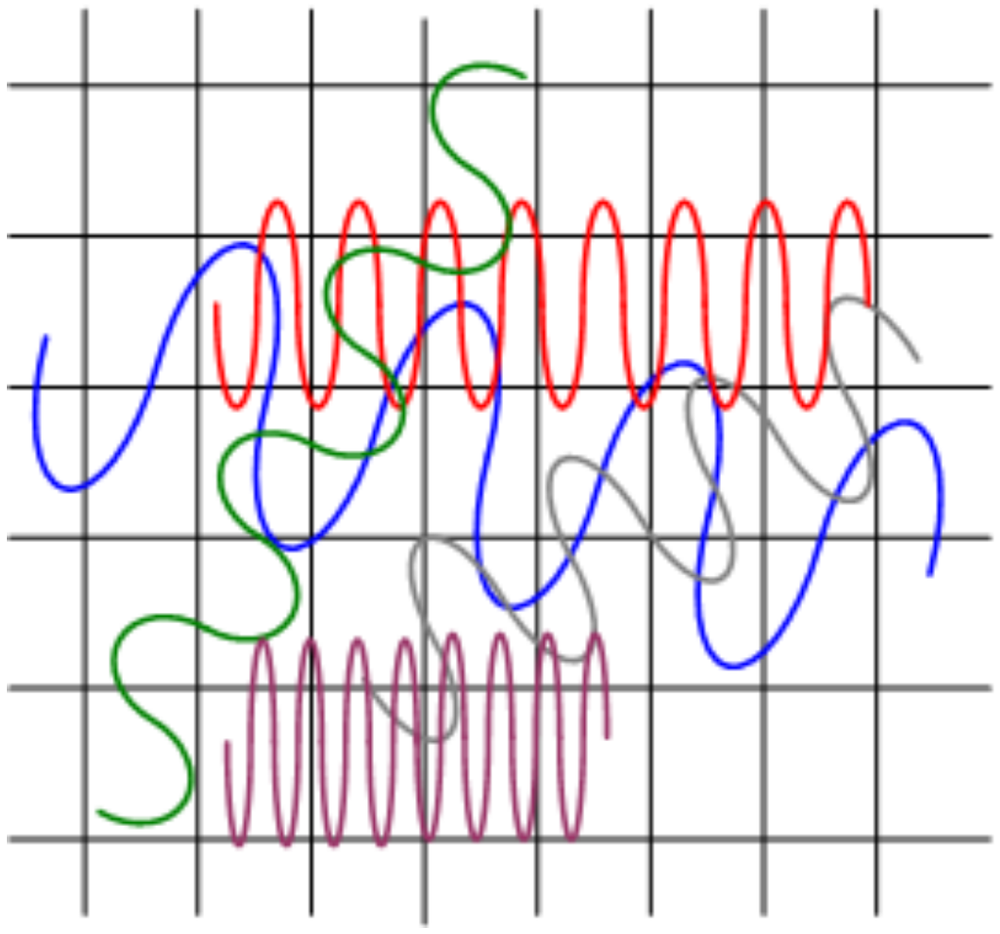
Shrodinger Wave Function

$$\frac{\partial^2\psi}{\partial x^2} + \frac{8\pi^2 m}{h^2}(E - V)\psi = 0$$

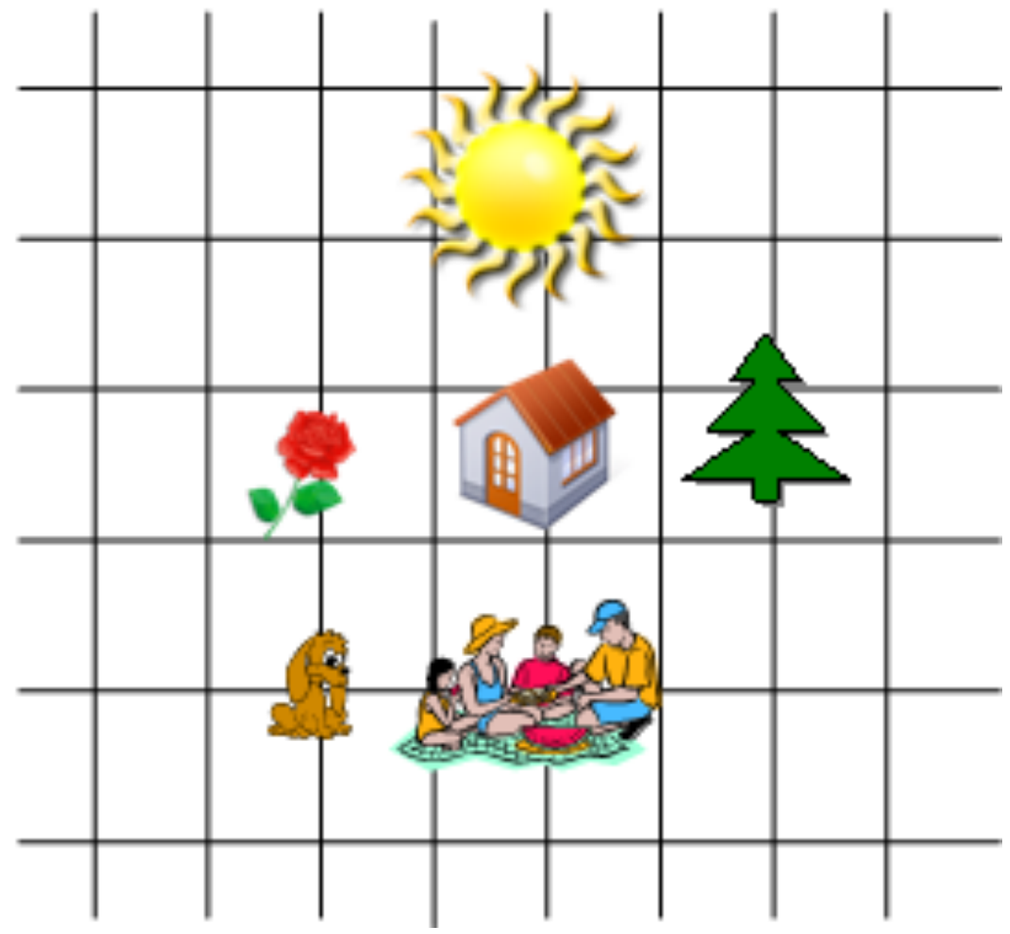
Position

Energy

Potential Energy

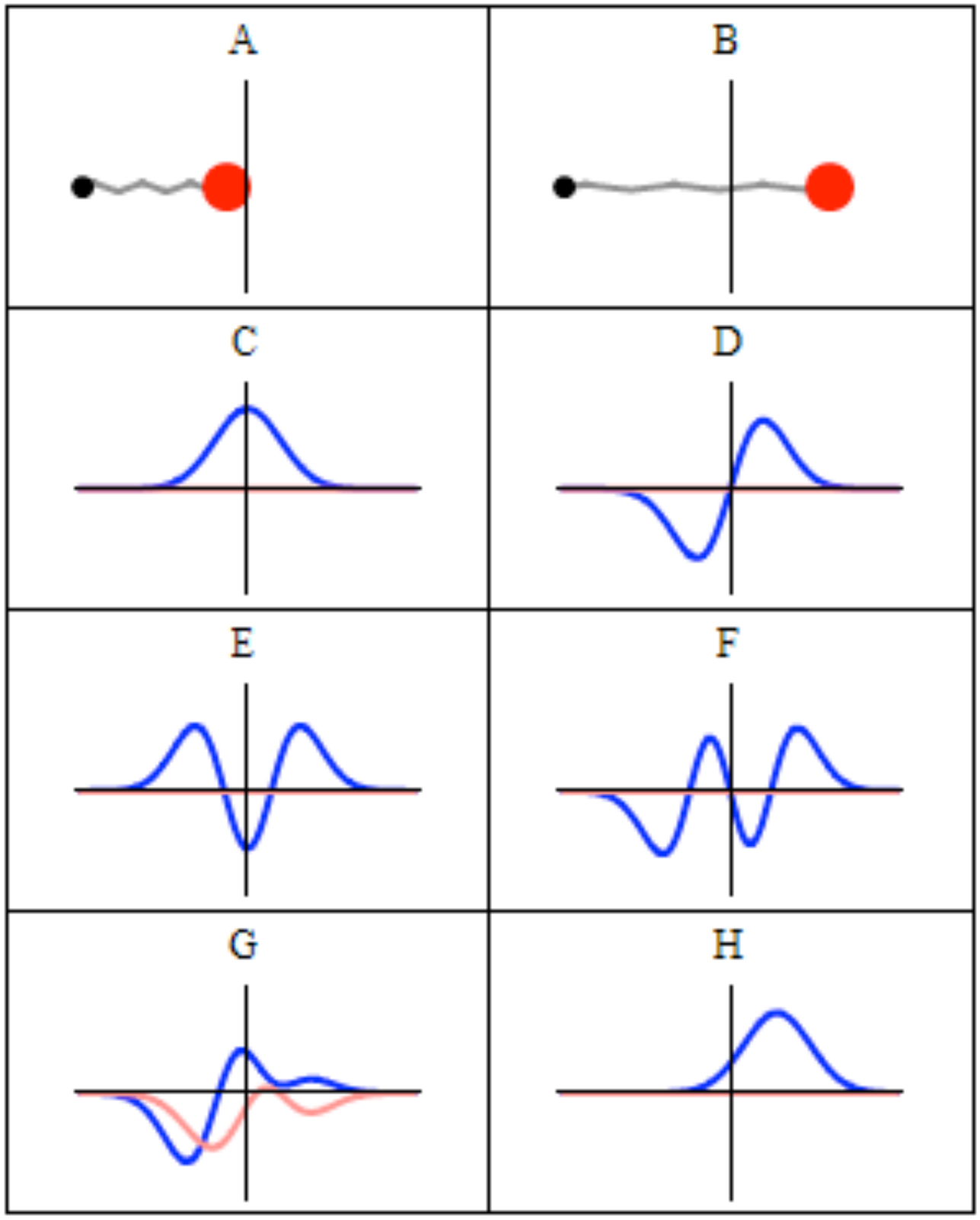


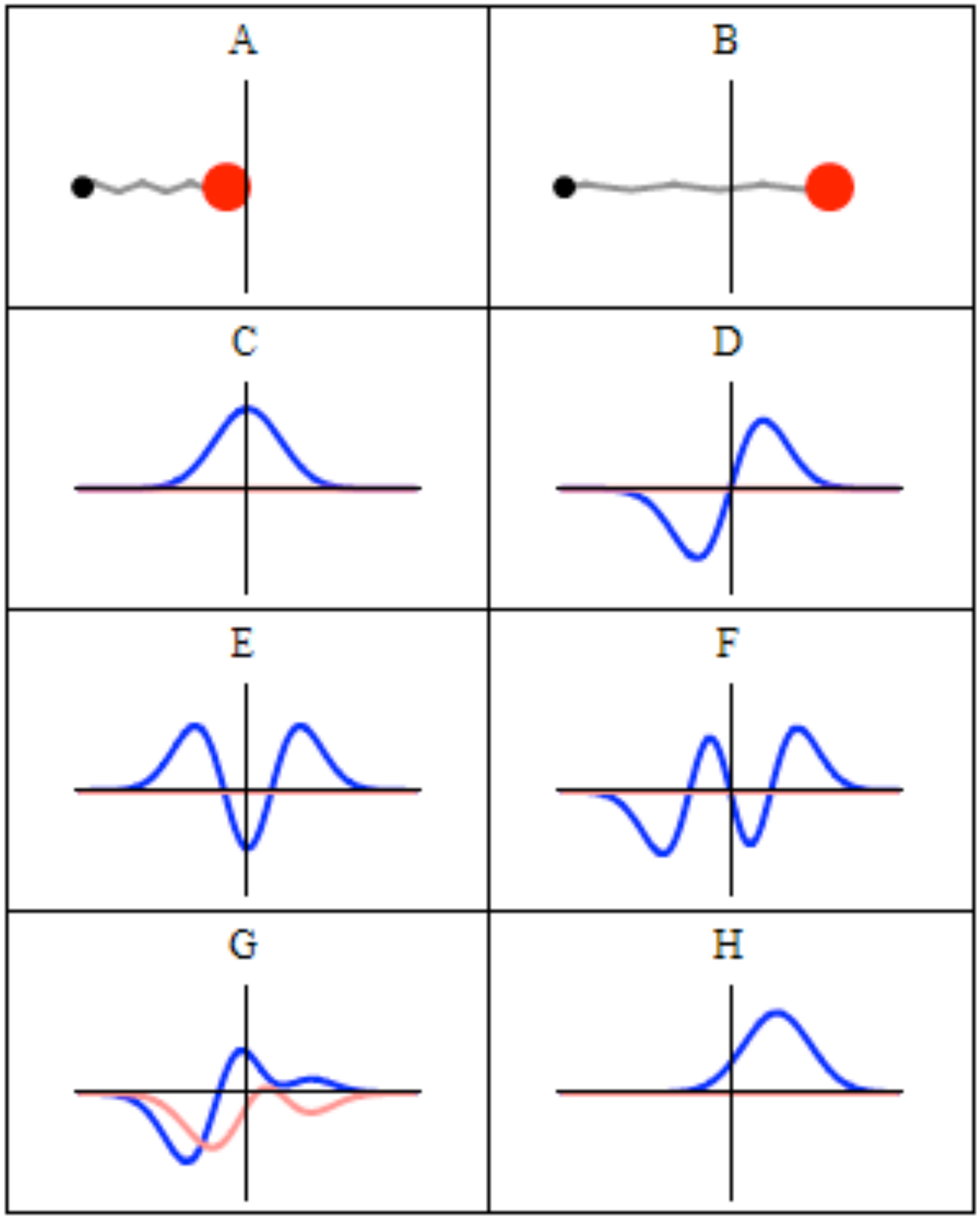
COLLAPSE OF THE WAVEFUNCTION



The Wave Function

- In classical physics, systems are described by definite values
  - A particle's position is specified by a definite location.
- In quantum mechanics, systems are described by wave functions.
  - A particle's position is specified by a wave function, with different amplitudes for different locations.


A
B
C
D
E
F
G
H

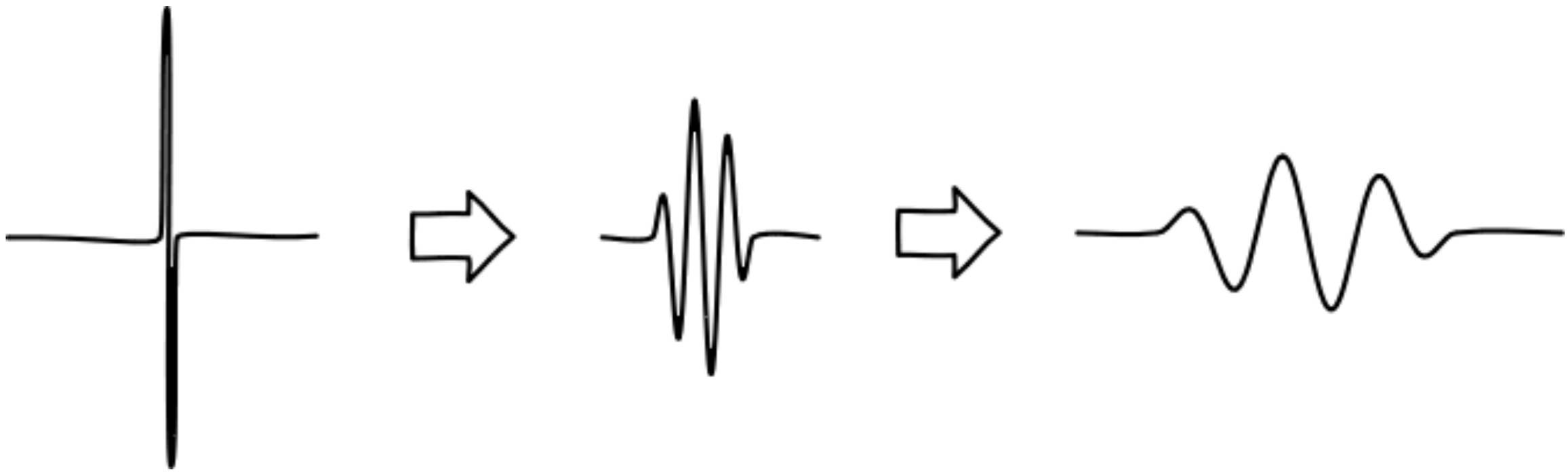

A
B
C
D
E
F
G
H

Superposition

- Typically a wave function will have nonzero amplitude at many positions.
- Then the particle is in a *superposition* of different positions.

The Schrödinger Equation

- The wave function usually evolves according to the Schrödinger equation
- Systems tend to evolve into superpositions.

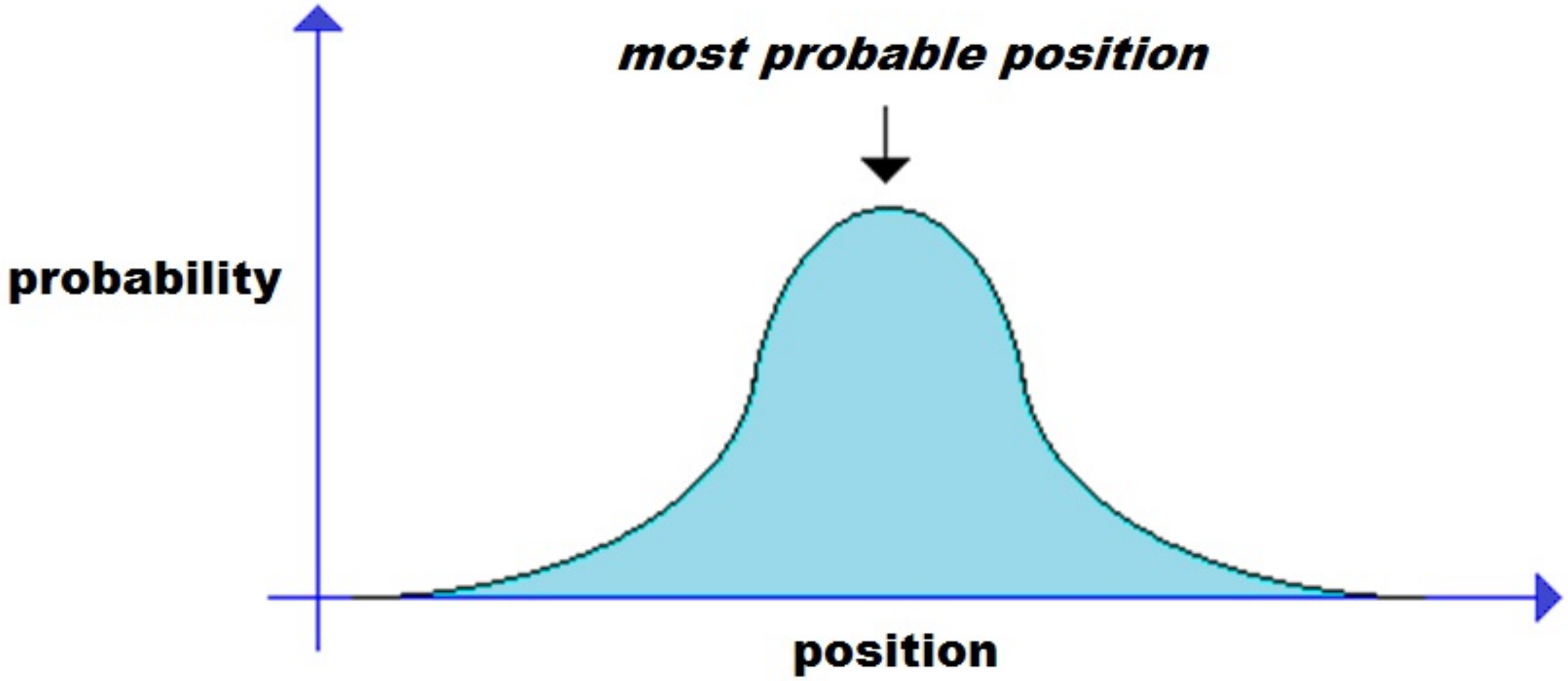


$$H(t)|\psi(t)\rangle = i\hbar\frac{\partial}{\partial t}|\psi(t)\rangle$$

Measurement

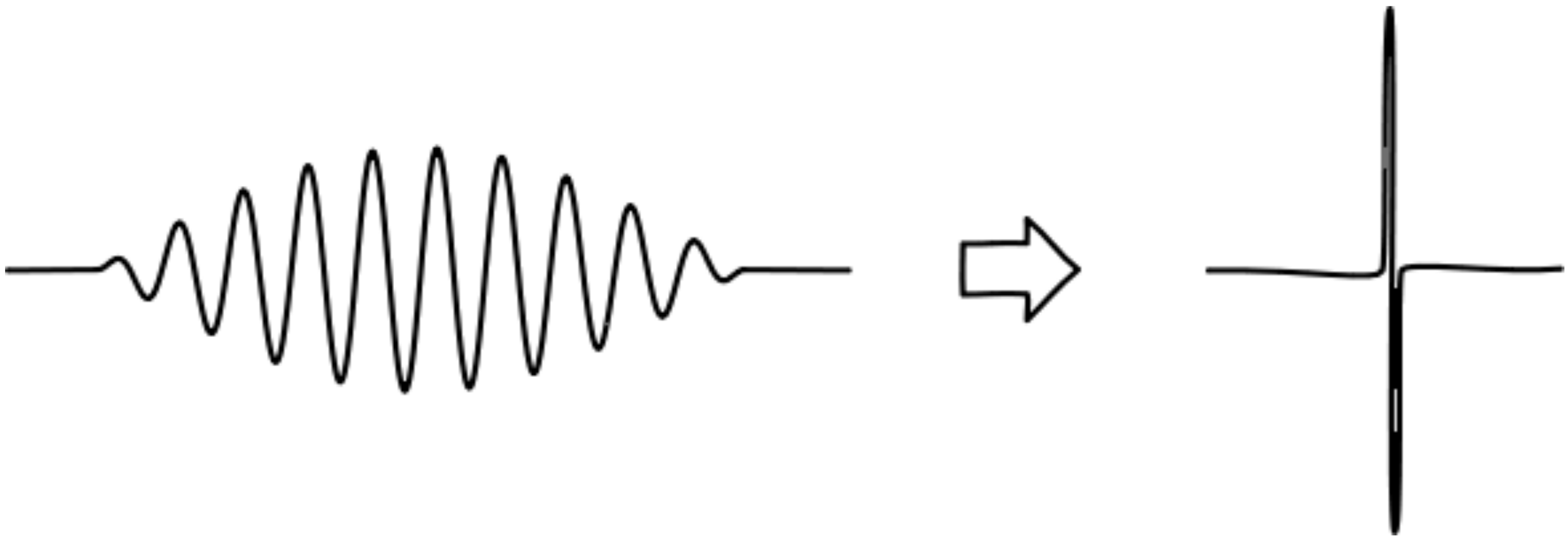
- When one measures a quantity (such as position), one always observes a definite result.
- When a system is in a superposition of values, the measurement may reveal any of these values, with probabilities given by the Born rule.

Quantum Wave Function



Collapse

- After measurement, the wave function enters a new state corresponding to the measurement result.
  - Initially: a superposition of position.
  - After: a definite position (an eigenstate).
  - This process is often called *collapse*.



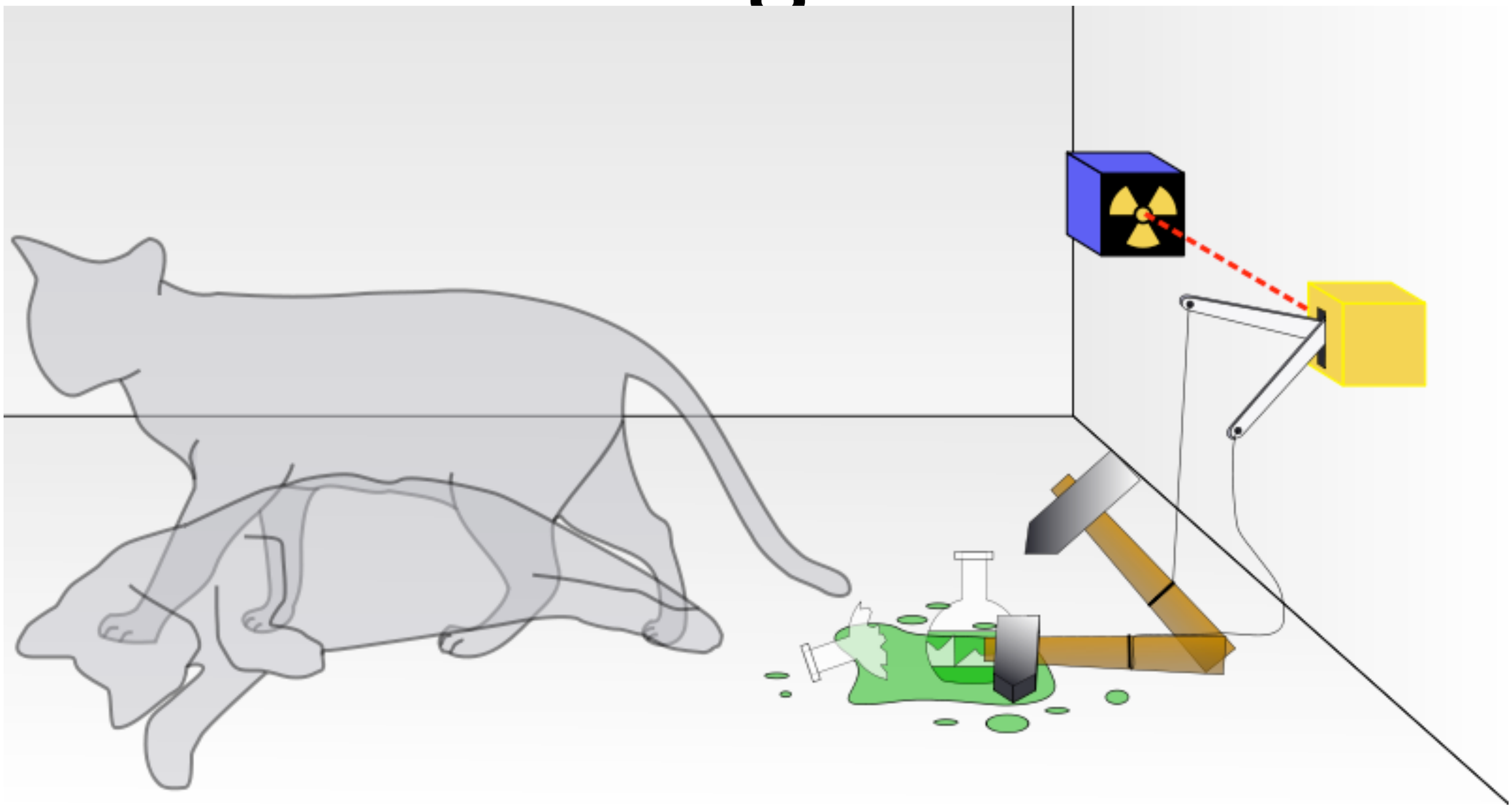
Formalism and Reality

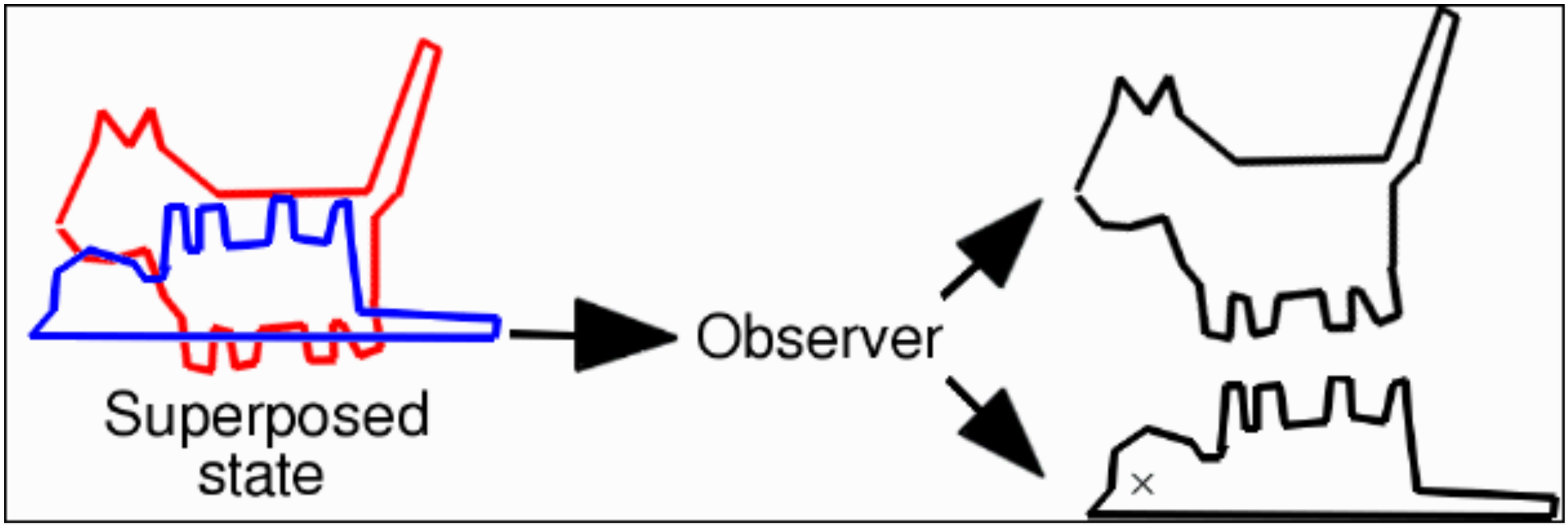
- Something like this story is the standard formalism for predicting measurement results in quantum mechanics.
- But what is really going on in reality?

The Measurement Problem

- The formalism says collapse takes place on measurement; but measurement is an imprecise notion.
- What is measurement?
- And how can it play a fundamental role in physical dynamics?

Schrodinger's Cat




Superposed
state
Observer

I SURVIVED

SUCK IT, SCHRODINGER

Alternative Interpretations

- Hidden-variables (Bohm):
  - Particles have definite positions all along
- Many worlds (Everett):
  - Even macro systems are in superpositions
- Spontaneous collapse (GRW):
  - Collapses happen randomly

Measurement Interpretations

- Collapses happen in reality, triggered by measurement events.
- One needs to precisify the notion of measurement and clarify the basic principles.

Consciousness-Collapse Interpretation

- Measurement = observation by consciousness.
- Consciousness triggers wave function collapse

Advantages

- Gives a precise trigger for collapse
- Gives a causal role for consciousness
- Fits the standard quantum formalism
- Explains why we always consciously experience a classical world

Disadvantages

- Criteria for consciousness are imprecise
- Dynamics of collapse is unclear
- Requires mind-body dualism

Project

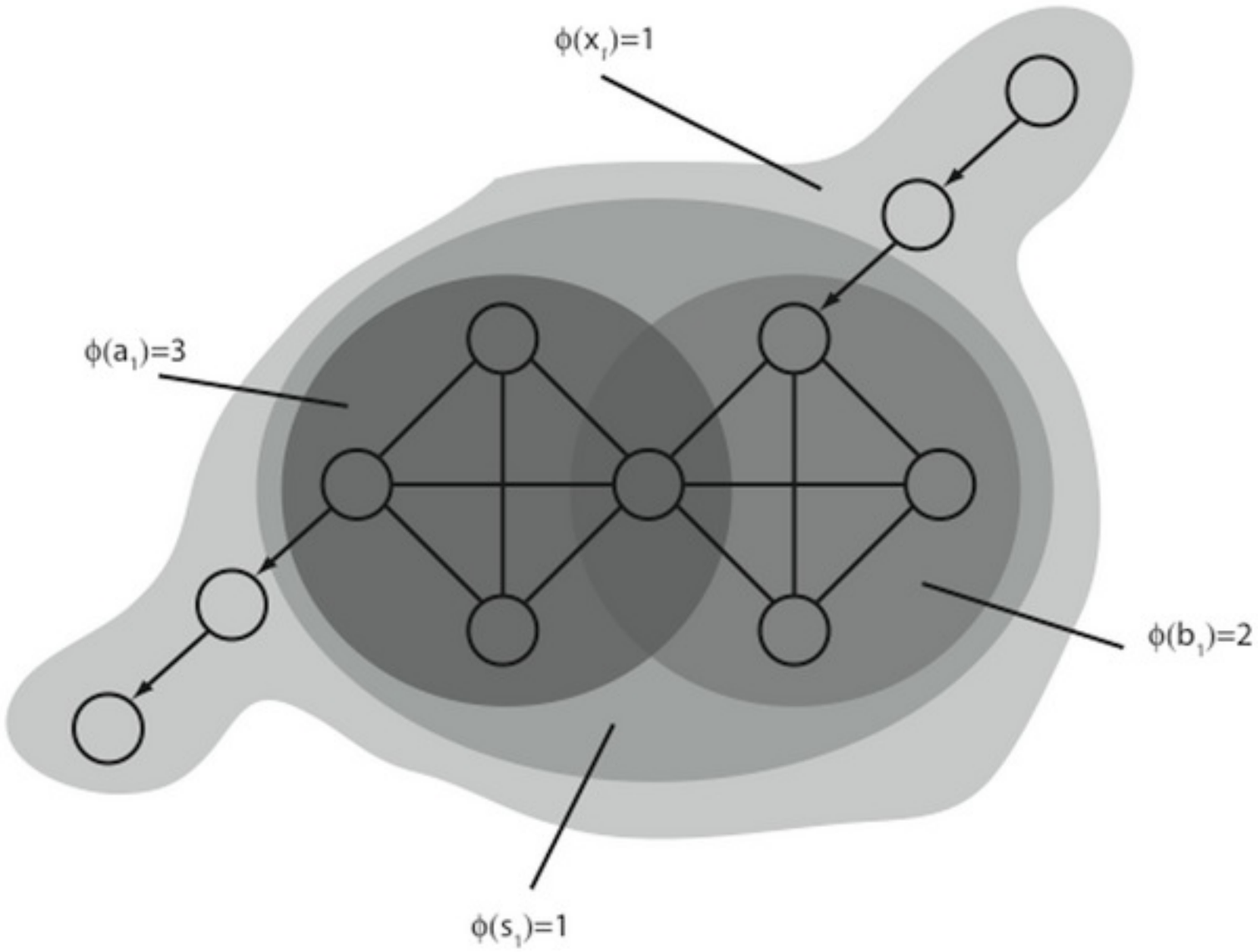
- Use a mathematical model of consciousness and a mathematical dynamics of collapse to minimize the unclarities
- Thereby evaluate the prospects for the consciousness-collapse thesis.

Physical Correlate of Consciousness

- The project needs a precise theory of the *physical correlate of consciousness* (PCC).
- One such theory: Giulio Tononi's integrated information theory.
- We'll use this theory to illustrate our approach, though others could also be used.

Integrated Information Theory

- Giulio Tononi: The physical correlate of consciousness is *information integration* in a physical system
- Φ = degree of consciousness



$$ei(x_1;\, P) = -\sum_{i=1}^{k}\sum_{\mu_0{}^{(i)}} p(\mu_0{}^{(i)}|x_1)\,\log p(\mu_0{}^{(i)}|\mu_1{}^{(i)}) - H(X_0|x_1)$$

$$\Phi(x_1) = \min_{P} \frac{ei(x_1;\, P)}{\nu_P}$$

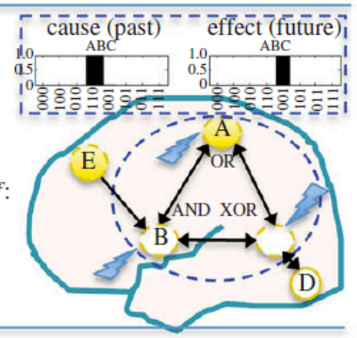
axioms
essential properties of every experience

postulates
properties that physical systems (elements in a state) must have to account for experience

intrinsic existense

consciousness exists intrinsically: each experience is real, and it exists from its own *intrinsic perspective,* independent of external observers (it is intrinsically real)

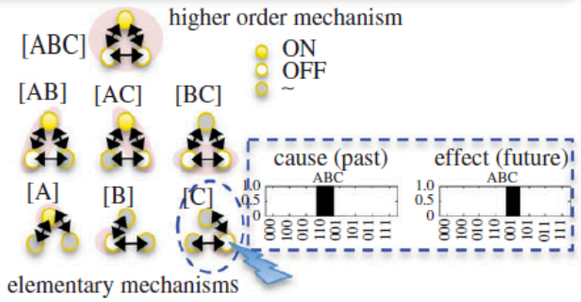
to account for experience, a system of mechanisms in a state must exist intrinsically. To exist, it must have *cause–effect power*; to exist from its own *intrinsic perspective,* independent of extrinsic factors, it must have cause–effect power *upon itself*: its present mechanisms and state must 'make a difference' to the probability of some past and future state of the system (its *cause–effect space*)



composition

consciousness is structured: each experience is composed of phenomenological distinctions, elementary or higher-order, which *exist* within it

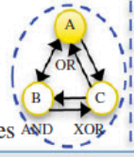
the system must be structured: subsets of system elements (composed in various combinations) must have cause–effect power upon the system



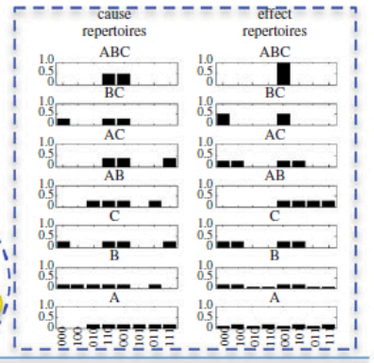
information

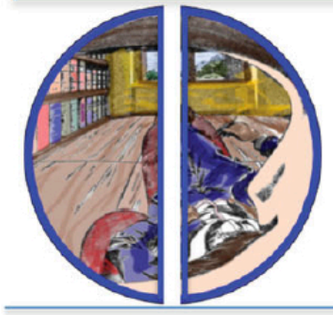
consciousness is specific: each experience is *the particular way it is* (it is composed of a specific set of specific phenomenological distinctions), thereby differing from other possible experiences (*differentiation*)

the system must specify a cause–effect structure that is *the particular way it is*: a specific set of specific cause–effect repertoires—thereby differing in its specific way from other possible structures (*differentiation*). A *cause–effect repertoire* specifies the probability of all possible causes and effects of a mechanism in a state. A *cause–effect structure* is the set of cause–effect repertoires specified by all subsets of system elements and expresses how the system gives an *actual form* to the space of possibilities



cause-effect structure

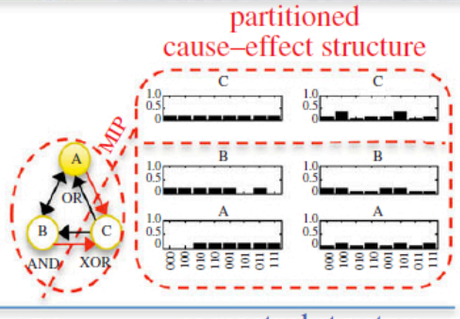




integration

consciousness is unified: each experience is *irreducible* to non-interdependent subsets of phenomenal distinctions

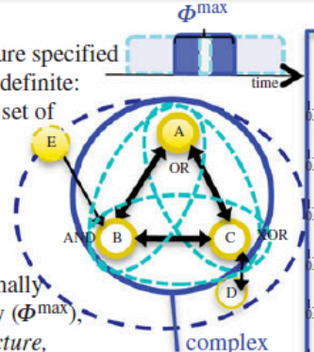
the cause–effect structure specified by the system must be unified: it must be intrinsically *irreducible* to that specified by non-interdependent sub-systems ($\Phi > 0$) across its weakest (unidirectional) link:
MIP = minimum information partition



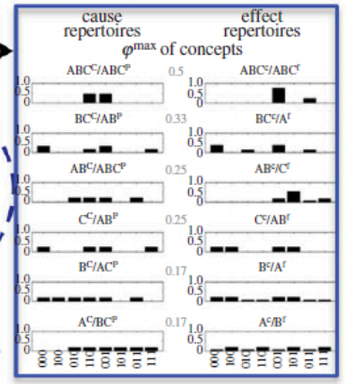
exclusion

consciousness is definite, in content and spatio-temporal grain: each experience has the set of phenomenal distinctions it has, not less or more, and flows at the speed it does, not faster or slower

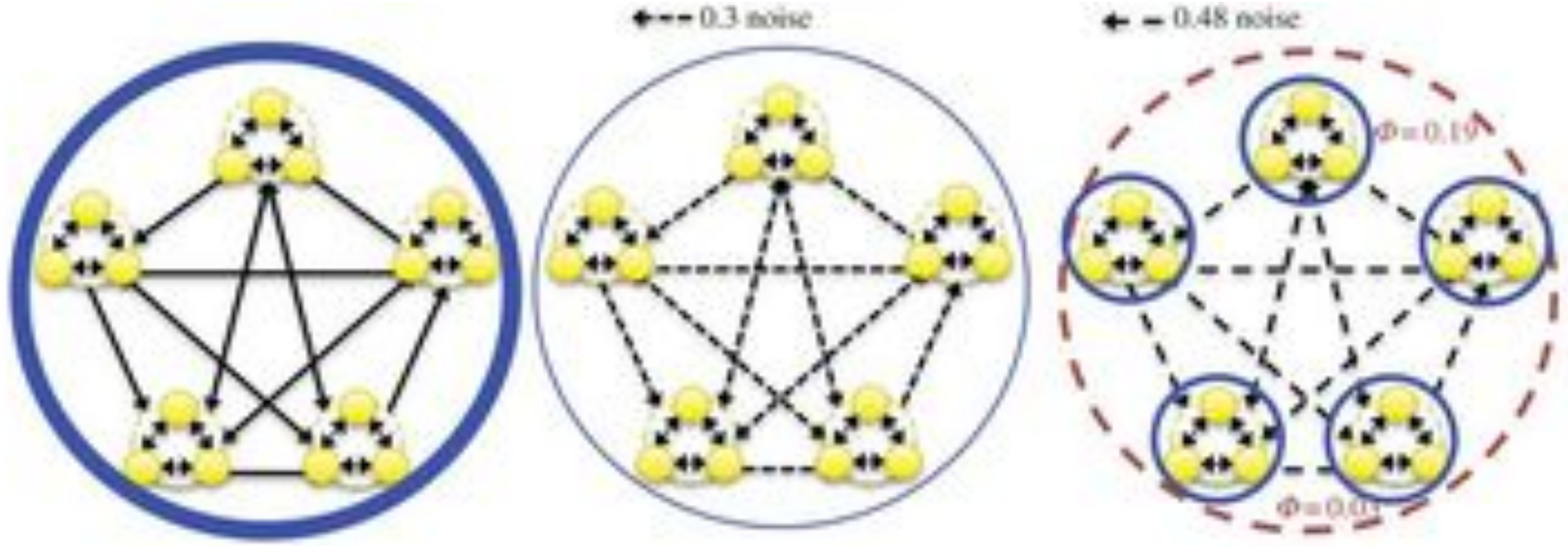
the cause–effect structure specified by the system must be definite: specified over a *single* set of elements—not lesss or more—and spatio-temporal grains—not faster or slower; this is a cause-effect structure that is maximally irreducible intrinsically (Φ^{max}), called *conceptual structure,* made of maximally irreducible cause–effect repertoires (*concepts*)



conceptual structure

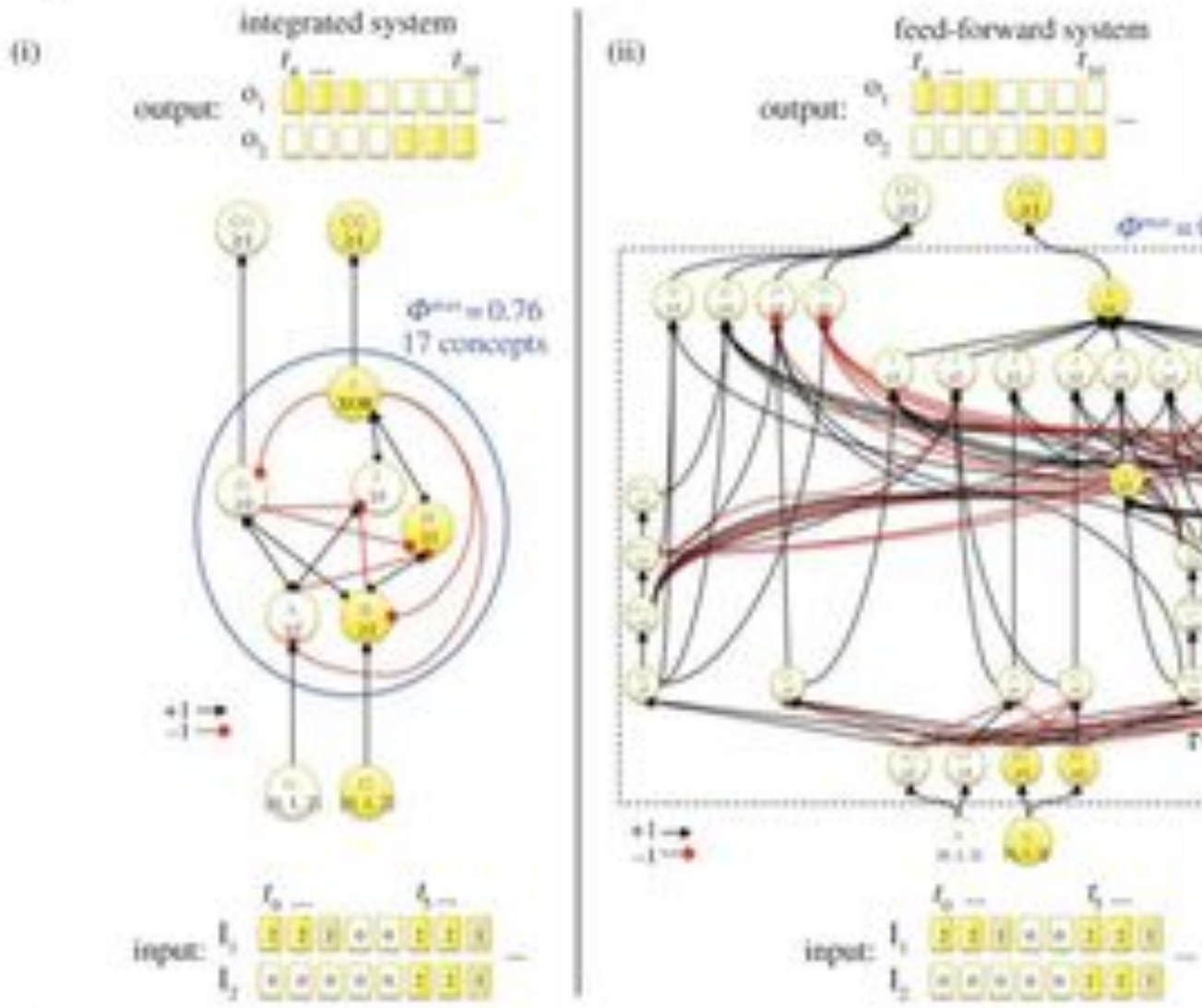


(a)



$\Phi^{max} \approx 10.56$ $\Phi^{max} \approx 3.22$ $\Phi^{max} \approx 0.19$

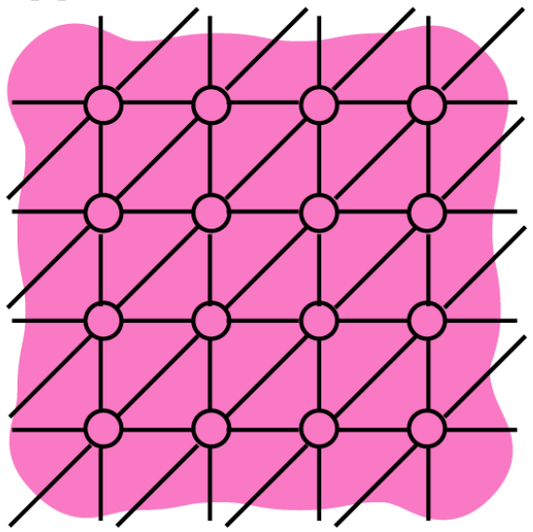
(b)



Measures of Consciousness

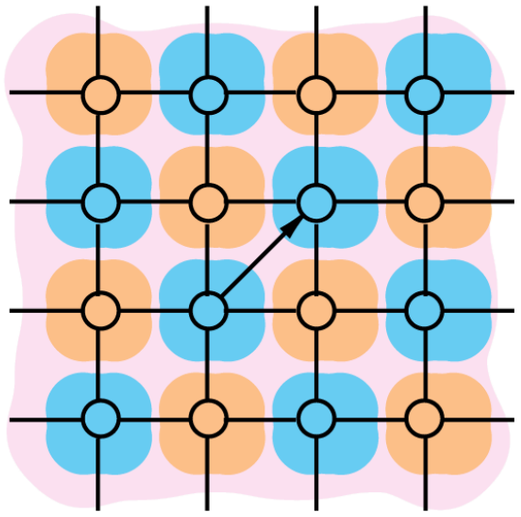
- Tononi gives mathematical definitions of
  - Φ (degree of consciousness)
    - Q-shape (state of consciousness)
- Q-shape is the physical correlate of consciousness.
- Fundamental principle: Q-shape ↔ consciousness

A



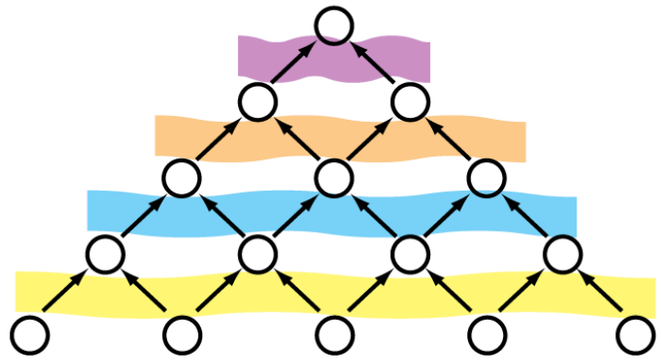
γ(all)=6.1 γ("orange")=.77

B



γ(orange)=2.53 γ(cyan)=2.46 γ(all)=.15

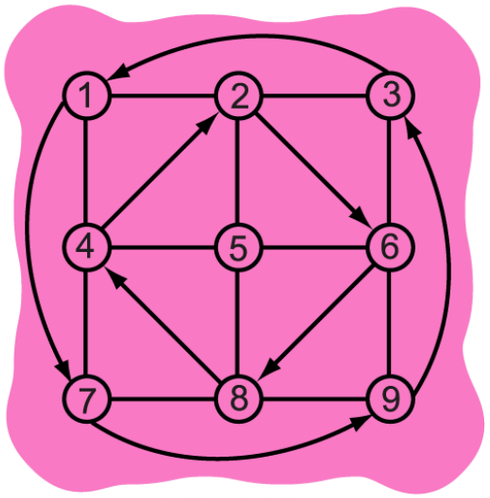
C



γ(purple)=.25 γ(cyan)=.14

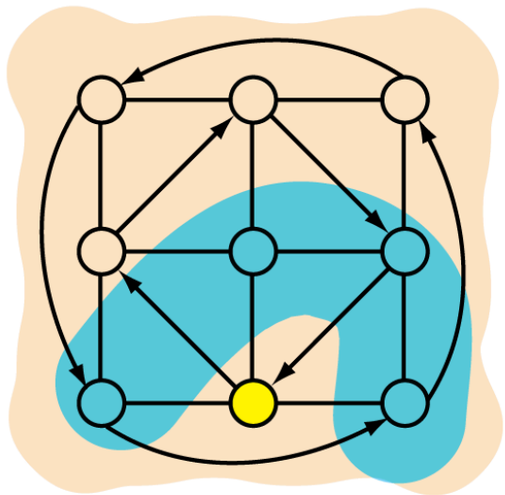
γ(orange)=.18 γ(yellow)=.16

D



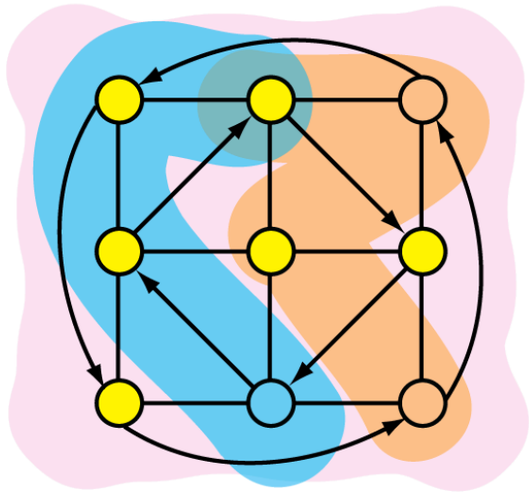
γ(all)=1.44

E



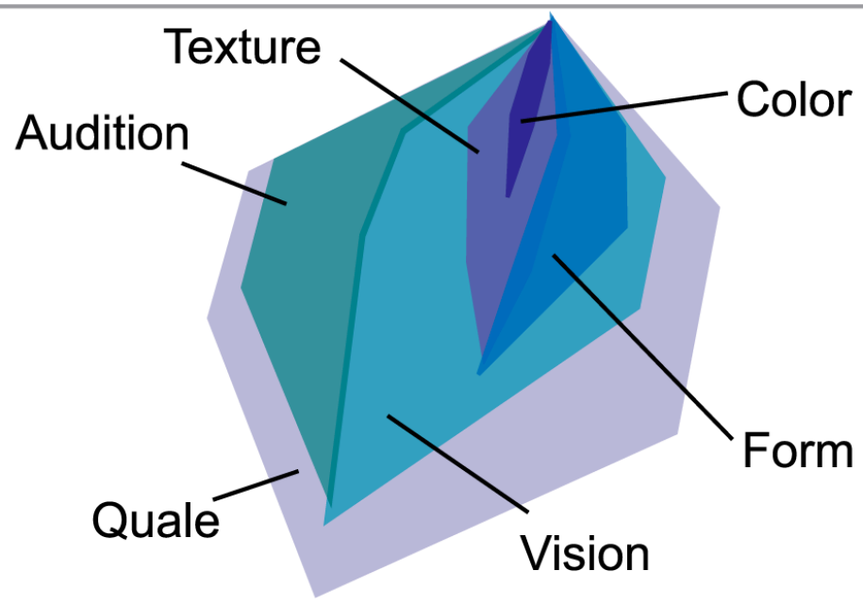
γ(cyan)=2.41 γ(all)=.23

F



γ(cyan)=.54 γ(orange)=.47 γ(all)=.89

G



Super-Resistance

- Key idea behind consciousness-collapse (Wigner, Albert, Chalmers, others): consciousness is *super-resistant*
- That is, consciousness (and its physical correlates) resist quantum superposition.
- Superpositions are either impossible or unstable.

Q-Shape as Super-Resistant

- Q-shape is defined classically, and can be treated as a quantum observable (or as a set thereof).
- Fundamental principle: Q-shape resists superposition
- Then: consciousness will collapse what it interacts with

Entanglement with Consciousness

- Take a superposed electron: $|S_1\rangle + |S_2\rangle$
- We consciously perceive it, potentially yielding $|S_1\rangle|Q(S_1)\rangle + |S_2\rangle|Q(S_2)\rangle$
- Consciousness collapses probabilistically e.g. to $|Q(S_1)\rangle$, electron collapses to $|S_1\rangle$
- Result: definite state $|S_1\rangle|Q(S_1)\rangle$

Superselection

- Simplest version: There is a *superselection rule* saying that the PCC (and consciousness) can never enter a superposition

Problem: Quantum Zeno Effect

- Quantum Zeno Effect (informal): if one measures an observable too often, it is slow to change.
- If one measures an observable continuously, it cannot change.
- E.g. if one continuous measures an electron's position, it will be frozen in place.

PERFORMANCE ANXIETY

QZE Applied to Measurement

- If the PCC can't superpose, then it's as if the PCC is being continuously measured
- If so, the PCC can never change.

Bad Consequences

- Bad consequences:
  - consciousness can never get started in the early universe
  - you can never wake up from a nap

Alternative: Stochastic Collapse

- Alternative: superpositions of the PCC are possible, but unstable.
- They collapse stochastically toward definite states.

Continuous Collapse

- Mathematical accounts of stochastic continuous collapse have been developed
  - Penrose/Diosi (gravitational collapse of spacetime superpositions)
  - Pearle (continuous spontaneous localization)

THE CONTINUOUS SPONTANEOUS LOCALIZATIONS (CSL) MODEL

G.C. Ghirardi, P. Pearle and A. Rimini, Phys. Rev. A **42**, 78 (1990).
A. Bassi and G.C. Ghirardi, Phys. Rept. **379**, 257 (2003).

$$d|\psi_t\rangle = \left[-\frac{i}{\hbar}Hdt + \sqrt{\gamma}\int d\mathbf{z}(M_{\mathbf{z}} - \langle M_{\mathbf{z}}\rangle_t)dW_t(\mathbf{z}) - \frac{\gamma}{2}\int d\mathbf{z}(M_{\mathbf{z}} - \langle M_{\mathbf{z}}\rangle_t)^2 dt\right]|\psi_t\rangle$$

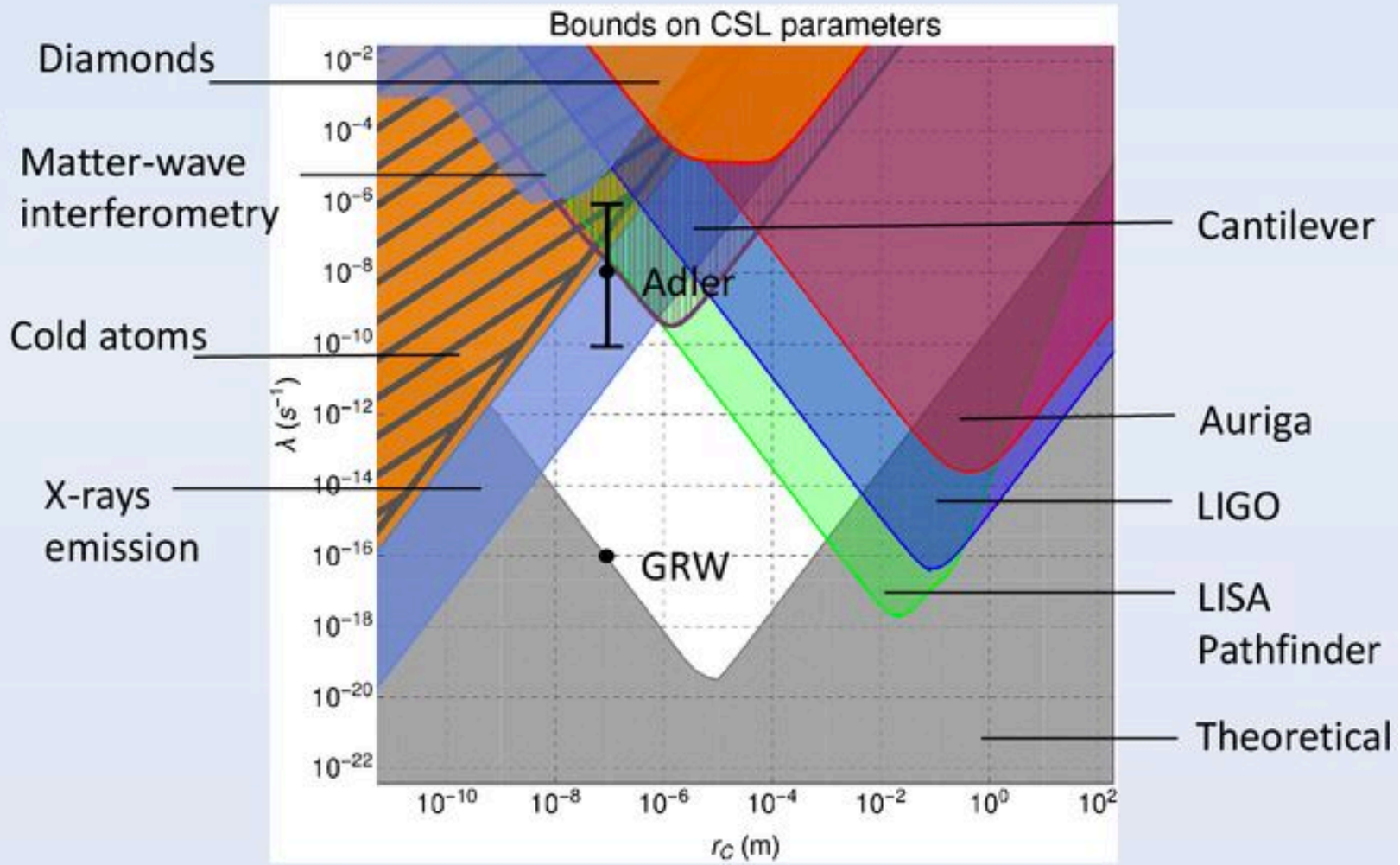
Schroedinger — collapse

$$\gamma = \lambda 8\pi^{3/2} r_C^3$$

$$\mathbb{E}[dW_t(\mathbf{x})dW_t(\mathbf{y})] = \delta(\mathbf{x}-\mathbf{y})dt$$

$$\hat{M}_{\mathbf{z}} = \sum_{j=1}^{N} \frac{m_j}{m_0} \frac{e^{-\frac{\left(\hat{\mathbf{x}}_j - \mathbf{z}\right)^2}{2r_c^2}}}{\left(\sqrt{2\pi} r_c\right)^3}$$

- Localization in space;
- Amplification mechanism;
- Predictions depend on λ and r_C.

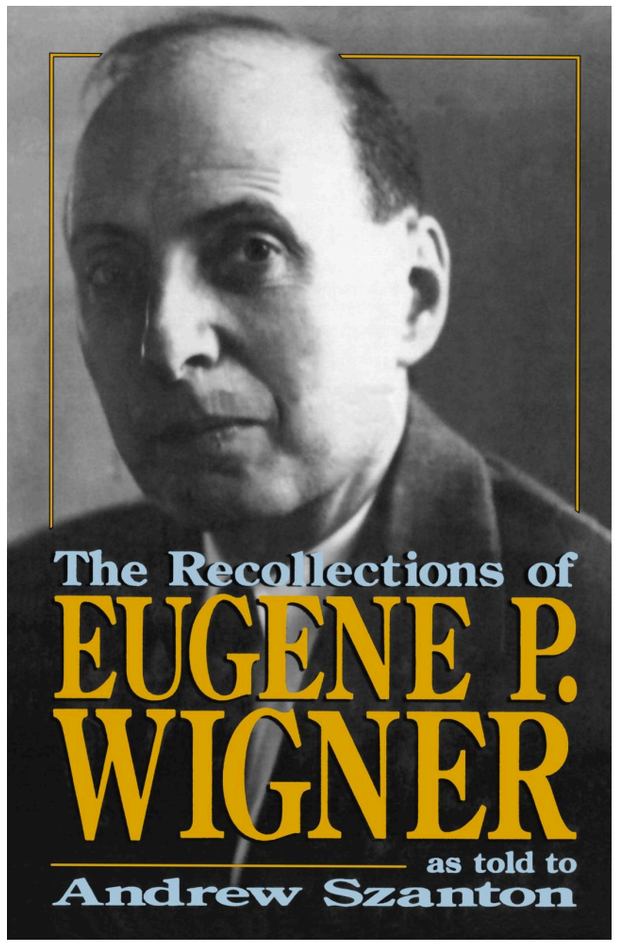


IIT Meets CSL

- We can combine IIT with Pearle's continuous spontaneous localization theory
  - Pearle: position stochastically localizes
  - IIT plus CSL: Q-shape (or PCC observable) stochastically localizes
  - Collapses toward definite Q-shape with Born probabilities

Stochastic Collapse of PCC

- Superpositions of PCC (Q-shape) yield superpositions of consciousness.
- Large superpositions of consciousness tend to stochastically collapse toward definite states (using distance metric between Q-shapes)



- Principles:
  - Q-shape ↔ consciousness (Tononi)
  - consciousness stochastically resists superposition (Wigner/Pearle)

Observation

- When someone observes a superposed particle (position A or B), their PCC will become increasingly superposed
- Consciousness will stochastically collapse to definite state, bringing about collapse of PCC and of entangled physical processes.
- Causes actions, e.g. “Particles in position A”
- Causal role for consciousness!

Questions and Objections

- How can consciousness be superposed?
- Does consciousness play the right sort of causal role?
- Is this compatible with relativity?
- What about the early universe?
- Can the theory be empirically tested?

Empirical Test

- In principle, different hypotheses about the locus of collapse lead to different empirical predictions.

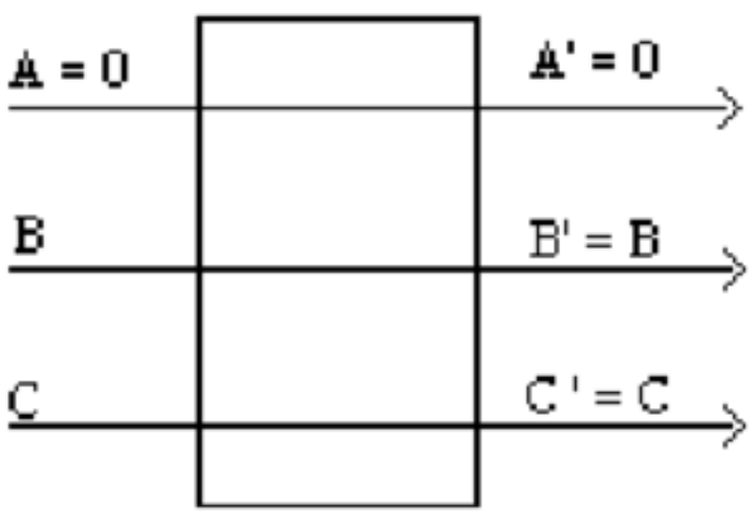
Interferometer Experiments

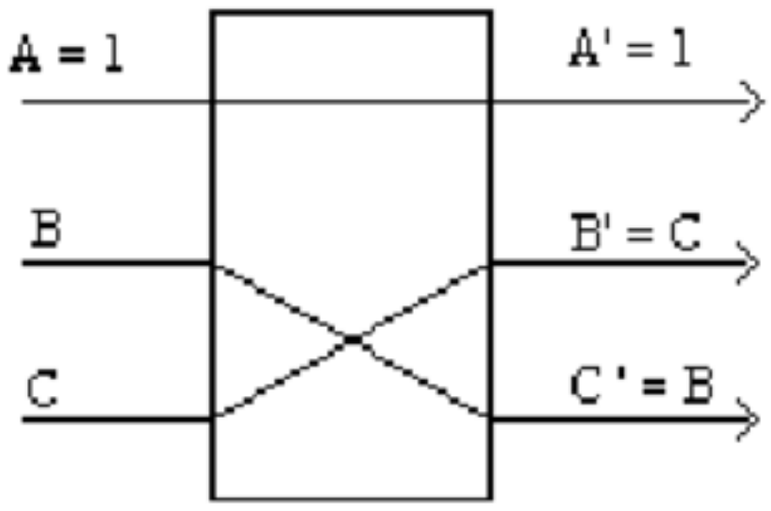
- Q: Does X collapse superpositions?
- A: See whether states of X produce superposition-based interference effects (interferometer experiments)
- Very difficult in practice! Most complex to date (Fein et al 2019): 2000-atom molecules.

IIT Experiments

- We can already test the hypothesis that Φ or Q-shape are super-resistant and lead to wave function collapse.
- Very simple two-unit systems have nonzero Φ and nontrivial Q-shape.
- Test using quantum computing technology.

Fredkin Gates





- Left: Φ(BC) = 0. Right: Φ(BC) = 2.
- Superpose A (the control bit): then B-C should enter superposition of Q-shapes.
- Q: Does this superposition collapse?

Answer

- Answer: no. Quantum computing Fredkin gates work as expected without collapse (Patel et al 2016).
- So: superpositions of Q-shape do not in general lead to wave function collapse.
- A simple Q-shape consciousness-collapse theory is falsified!

Successor Theories

- A more plausible consciousness-collapse theory will require consciousness to exist only in more complex systems.
  - Q-shape in systems with $\Phi > k$?
  - Modify IIT so simple systems have $\Phi = 0$?
  - Non-IIT based theory?

Empirical Program

- There's an empirical program of seeing which if any properties resist superposition.
- If we eventually find such properties:
  - we will have strong evidence for a collapse interpretation.
  - we will have empirical evidence about the physical locus of consciousness
- A longshot, but worth exploring!

Overall

- Consciousness-collapse theories can be made relatively precise.
- There is little direct evidence for them now, but some indirect support.
- They're empirically testable in principle.
- Intriguing possibility: an empirical resolution of both the problem of quantum measurement and the problem of consciousness.