'Life hates surprises': Can ambitious theory unify biology, neuroscience and psychology?

August 15 2022, by Ross Pain, Michael David Kirchhoff and Stephen Francis Mann

Credit: AI-generated image (disclaimer)

In the early 1990s, British neuroscientist Karl Friston was poring over brain scans. The scans produced terabytes of digital output, and Friston had to find new techniques to sort and classify the massive flows of data.

Along the way he had a revelation. The techniques he was using might be similar to what the *brain itself* was doing when it processed visual data.

Could it be he had stumbled upon a solution to a data engineering problem that nature had discovered long ago? Friston's eureka moment led to a "theory of everything," which claims to explain the behavior of the brain, the mind, and life itself.

As we discovered when we put together a collection of papers, the theory—known as the "free energy principle"—is controversial among scientists and philosophers.

Re-engineering nature

Friston's initial idea was appealing because the problem facing the brain is similar to that facing an experimental scientist. Both must use the data they have to draw conclusions about events they cannot observe directly.

The neuroscientist uses scan data to infer facts about brain processes. The brain uses sensory input to infer facts about the external world.

The algorithm Friston used to draw conclusions from his data—a mathematical operation called "minimizing free energy"—was based on existing techniques in statistical analysis.

Friston (and others such as computer scientist Geoff Hinton) realized artificial neural networks could easily carry out this operation. And if artificial neural networks could do it, perhaps biological neural networks could too.

But Friston didn't stop there. He reasoned that the problem of drawing conclusions from limited information is a problem faced by all living things.

This led him to the "[free energy principle](#)": that every living thing, everywhere, minimizes free energy.

The free energy principle

But what, exactly, is free energy? Why might all living things minimize it?

Start with a simpler idea: every organism is trying to minimize how surprising its experiences are. By "surprising," we mean experiences that have not been encountered previously by the organism or its ancestors.

Your ancestors were successful enough to produce a lineage that eventually produced you, so what they experienced must have promoted survival. And your own experiences so far have resulted in you still being alive.

So experiences you have not had before—surprises, in other words—may be dangerous. (The ultimate surprise is death.)

We can dress this idea in mathematical clothes by defining surprise in terms of probabilities. The less probable an experience, the more surprising.

And that's where "free energy" enters the picture. It's not energy as we would usually think of it—in this situation, free energy measures how improbable your experience would be *if* a certain unobserved situation were true.
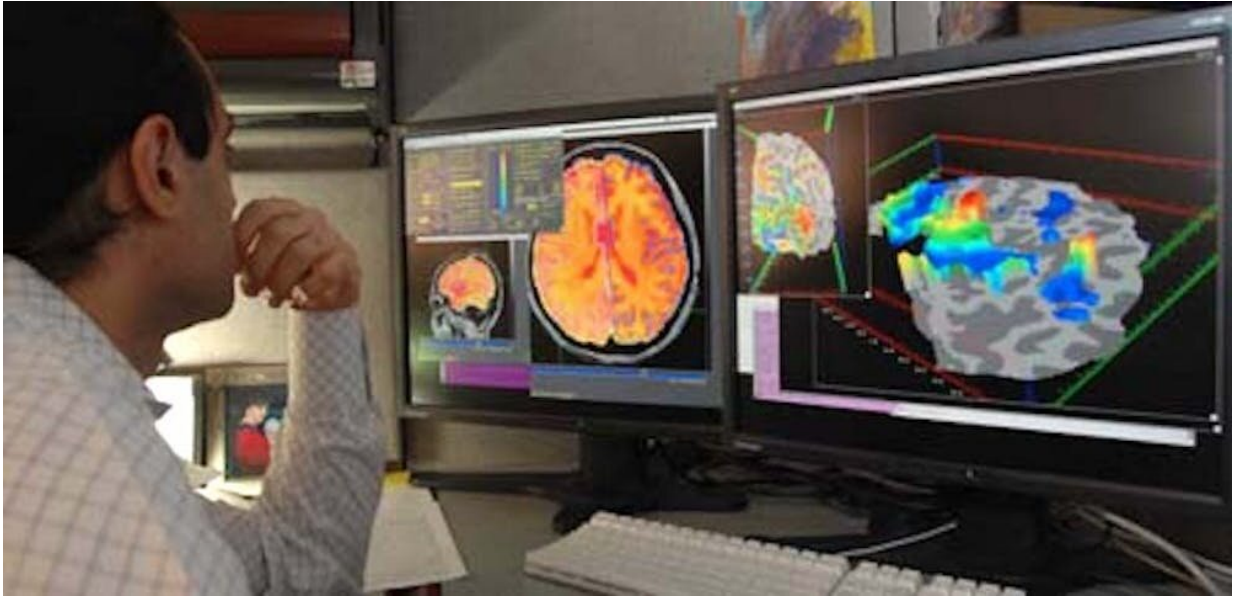
Karl Friston thought the method he used to interpret brain scans might be the same as the method the brain itself used to interpret visual data. Credit: NIMH / Wikimedia

No surprises?

Minimizing free energy means choosing to believe in the unobserved situation that makes your observations least surprising.

Here's an example: imagine you are picnicking in the park, watching two friends kick a football to and fro. Your view is occluded by a tree, so you don't see the full trajectory of the kicked ball.

Now, it is possible that there is a third person behind the tree, who catches the ball each time it passes them and then throws on a spare ball they have handy.

However, there is no evidence for the existence of this third person, so

their existence would be very surprising. So you can minimize your surprise by believing there is no secret third person behind the tree.

Minimizing free energy can help guide our actions, too. According to the free energy principle, you should do things that will change the world in such a way that your experiences are less likely to be surprising!

Seen from this perspective, we eat to avoid the surprise of extreme hunger, and we seek shelter to avoid the surprise of being cold.

How much does a 'theory of everything' actually explain?

So the free energy principle is a "theory of everything" spanning neuroscience, psychology and biology! But not everyone is convinced it's a useful idea.

Some of the skepticism concerns whether or not the theory is true. An even bigger concern is that, even if it is true, it may not be very useful.

But why would people think this?

The American population biologist Richard Levins famously outlined a dilemma facing scientists who study biological systems.

These systems contain a huge amount of potentially important detail, and when we model them we cannot hope to capture all of it. So how much detail should we attempt to capture, and how much should we leave out?

Levins concluded there is a trade-off between the level of detail in a model and the number of systems it applies to. A model that captures a lot of detail about a specific system will be less informative about other,

similar systems.

For instance, we can model the technique of an Olympic swimmer in order to improve their performance. But that model will not faithfully represent a different swimmer.

On the other hand, a model that covers more systems will be less informative about any particular system. By modeling how humans swim in general, we can design swimming lessons for children, but individual differences between children will be ignored.

The moral is that our models should fit our aims. If you want to explain the workings of a particular system, produce a highly specific model. If you want to say things about a lot of different systems, produce a general model.

Too general to be useful?

The free energy principle is a highly general model. It might even be the most general model in the life sciences today.

But how useful are such models in the day-to-day practice of biology or psychology? [Critics](link) argue Friston's theory is so general that it is hard to see how it might be put to practical use.

Proponents claim [successes](link) for the [free energy](link) principle, but will it turn out to be an enormous breakthrough? Or do theories that try to explain everything end up explaining nothing?

This article is republished from [The Conversation](link) under a Creative Commons license. Read the [original article](link).