

# How to build a computer with free will

March 14 2018, by Mark Hadley

---



Credit: AI-generated image ([disclaimer](#))

Do you have free will? Can you make your own decisions? Or are you more like an automaton, just moving as required by your constituent parts? Probably, like most people, you feel you have something called free will. Your decisions are not predetermined; you could do otherwise.

Yet scientists can tell you that you are made up of atoms and molecules and that they are governed by the laws of physics. Fundamentally, then –

in terms of atoms and molecules – we can predict the future for any given starting point. This seems to leave no room for free will, alternative actions or decisions.

Confused? You have every right to be. This has been one of the long outstanding [unresolved problems](#) in philosophy. There has been no convincing resolution, though speculation has included [a key role for quantum theory](#), which describes the uncertainty of nature at the smallest scales. It is this that has fascinated me. My research interests include the foundations of quantum theory. So could free will be thought of as a macroscopic [quantum phenomenon](#)? I set out to explore the question.

## Quantum free will

There is enough philosophy literature on the subject to fill a small library. As a trained scientist I approached the problem by asking: what is the [evidence](#)? Sadly, in some ways, [my research](#) showed no link between free will and [fundamental physics](#). Decades of philosophical debate as to whether free will could be a quantum phenomenon has been chasing an unfounded myth.

Imagine you are on stage, facing two envelopes. You are told that one has £100 inside and the other is empty. You have a free choice to pick one – yet every time the magician wins, and you pick the empty one. This implies that our sense of free will is not quite as reliable as we think it is – or at least that it's subject to manipulation, if it is there.

This is just one of a wide variety of examples that question our awareness of our own decision making processes. Evidence from [psychology](#), [sociology](#) and even [neuroscience](#) all give the same message that we are unaware of how we make decisions. And our own introspection is unreliable as evidence of how our mental processes function.

So, what is the evidence for the abstract concept of free will? None. How could we test for it? We can't. How could we recognise it? We can't. The supposed connection between our perception of free will and the uncertainty inherent to [quantum theory](#) is, therefore, unsupported by the evidence.

But we do have an experience of free will, and this experience is a fact. So having debunked the supposed link with fundamental physics, I wanted to go further and explore why we have a perception of being able to do otherwise. That perception is nothing to do with knowing the exact position of every molecule in our bodies, but everything to do with how we question and challenge our [decision making](#) in a way that really does change our behaviour.

## **Artificial free will**

For me as a scientist, this meant building a model of free will and testing it. But how would you do this? Could I mimic it with a computer [program](#)? If I were successful how would my computer or robot be tested?

The topic is fuelled with prejudice. You would probably assume without evidence that my brother has free will, but my computer does not. So I will offer an emotionally neutral challenge: if an alien lands on Earth, how would you decide if it was an alien being with free will like us, or a sophisticated automaton?

Strangely, the philosophical literature does not seem to consider tests for free will. But as a scientist, it was essential to have a test for my model. So here is my answer: if you are right handed, you will write your name holding a pen in your right hand. You will do so predictably almost 100% of the time. But you have free will, you could do otherwise. You can prove it by responding to a challenge or even challenging yourself.

Given a challenge you may well write with your left hand. That is a highly discerning test of free will. And you can probably think of others, not just finely balanced 50:50 choices, but really rare events that show your independence and distinguish you from an automaton.

Based on this, I would test my alien with a challenge to do something unusual and useless, perhaps slightly harmful even, like putting its hand near a flame. I would take that as evidence of free will. After all, no robot would be programmed to do that.

## **A computer with free will?**

And so I tried to model that behaviour in the simplest most direct way, starting with a generic goal seeking computer program that responds to inputs from the environment. These programs are commonly used across disciplines from sociology, economics and AI. The goal seeking program is so general that it applies to simple models of human behaviour, but also to hardware like the battery saving program in your mobile phone.

For free will, we add one more goal: to assert independence. The computer program is then designed to satisfy this goal or desire by responding to challenges to do otherwise. It's as simple as that. Test it out yourself, the challenges can be external or you can generate your own. After all, isn't that how you conclude that you have free will?

In principle the program can be implemented in today's computers. It would have to be sophisticated enough to recognise a [challenge](#) and even more so to generate its own challenges. But this is well within reach of current technology. That said, I'm not sure that I want my own personal [computer](#) exercising free will though ...

This article was originally published on [The Conversation](#). Read the [original article](#).

Provided by The Conversation

Citation: How to build a computer with free will (2018, March 14) retrieved 19 April 2024 from <https://techxplore.com/news/2018-03-free.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.