



*To arrive at the edge of the world's knowledge, seek out the most complex and sophisticated minds, put them in a room together, and have them ask each other the questions they are asking themselves.*

<https://www.edge.org/response-detail/25491>

Printed On Wed January 11th 2023



## 2014 : WHAT SCIENTIFIC IDEA IS READY FOR RETIREMENT?

[In the News \[ 60 \]](#)

[Contributors \[ 177 \]](#) | [View All Responses \[ 184 \]](#)



**[Tom Griffiths](#)**

*Henry R. Luce Professor of Information Technology, Consciousness and Culture, Director of the Computational Cognitive Science Lab, Princeton University; Co-author (with Brian Christian), [Algorithms to Live By](#)*

### Bias is Always Bad

Being biased seems like a bad thing. Intuitively, rationality and objectivity are equated—when faced with a difficult question, it seems like a rational agent shouldn't have a predisposition to favor one answer over another. If a new algorithm designed to find objects in images or interpret natural language is described as being biased, it sounds like a poor algorithm. And when psychology experiments show that people are systematically biased in the judgments they form and the decisions they make, we begin to question human rationality.

But bias isn't always bad. In fact, for certain kinds of questions, the only way to produce better answers is to be biased.

Many of the most challenging problems that humans solve are known as inductive problems—problems where the right answer cannot be definitively identified based on the available evidence. Finding objects in images and interpreting natural language are two classic examples. An image is just a two-dimensional array of pixels—a set of numbers indicating whether locations are light or dark, green or blue. An object is a three-dimensional form, and many different combinations of three-dimensional forms can result in the same pattern of numbers in a set of pixels. Seeing a particular pattern of numbers doesn't tell us which of these possible three-dimensional forms are present: we have to weigh the available evidence and make a guess. Likewise, extracting the words from the raw sound pattern of human speech requires making an informed guess about the particular sentence a person might have uttered.

The only way to solve inductive problems well is to be biased. Because the available evidence isn't enough to determine the right answer, you need to have predispositions that are independent of that evidence. And how well you solve the problem—how often your guesses are correct—depends on having biases that reflect how likely different answers are.

Human beings are very good at solving inductive problems. In finding objects in images and interpreting natural language are two problems that people still solve better than

computers. And the reason is that human minds have biases that are finely tuned for solving these problems.

The biases of the human visual system are apparent in many visual illusions—images that result in a surprising discrepancy between our biased guesses and what's actually in the world. The rarity of visual illusions in real life is testimony to the utility of those biases. By studying the kinds of illusions the human visual system is susceptible to, we can identify the biases that guide perception and instantiate those biases in algorithms used by computers.

Human biases in interpreting language are demonstrated in the game of Telephone, or when we misinterpret the lyrics of a song. It's also easy to discover the biases that have been built into speech recognition software. I once left my office for a meeting, locking the door behind me, and came back to find a stranger had broken in and typed a series of poetic sentences into my computer. Who was this person, and what did the message mean? After a few spooky, puzzling minutes, I realized that I had left my speech recognition software running, and the sentences were the guesses it had produced about what the rustling of the trees outside my window meant. But the fact that they were fairly intelligible English sentences reflected the biases of the software, which didn't even consider the possibility that it was listening to the wind rather than a person.

Things that people do well—vision and language—depend heavily on being biased towards particular answers. Algorithms that solve those problems well have similar biases. So we shouldn't be surprised to discover that people are systematically biased in other domains. These biases don't necessarily reflect a deviation from rationality—they reflect the difficulty of the problems that humans need to solve. And one way to make computers better at solving these problems is understanding exactly what human biases are like for different problems.

In arguing that bias isn't always bad, I'm not claiming that it is always good. Objectivity can be an ideal that we strive for on moral grounds—say, when assessing other people. The more information and time we have available, the closer we can get to this ideal. But this kind of objectivity is a luxury, at odds with reaching the right answers in limited time from small amounts of evidence. When solving inductive problems, it can be rational to be biased.

[Return to Table of Contents](#)

