# Chapter 3

# Induction and Innateness

One of the deepest problems in philosophy concerns how we learn about the world, and whether there are right or wrong ways to go about it. In this chapter I introduce this problem—the "problem of induction"—and describe its relevance to understanding learning in intelligent agents, and brains in particular. One consequence of the problem of induction is that there can be no such thing as a universal learning machine; it is not even possible that brains could enter the world as blank slates equipped with universal learning algorithms. The goal of the chapter is to provide a kind of solution to the problem of induction, and also to put forth something I call a theory of innateness. The latter would be a mathematical framework in which we are able to make sense of the kinds of structures that must be innately generated in a brain in order for that brain to have its own innate way of learning in the world. I present a theory called *Paradigm Theory* (Changizi and Barber, 1998) that purports to do these things.

## What is induction?

"John is a man. All men are mortal. Therefore, John is mortal." This argument from two premises to the conclusion is a *deductive* argument. The conclusion logically follows from the premises; equivalently, it is logically impossible for the conclusion not to be true *if* the premises are true. Mathematics is the primary domain of deductive argument, but our everyday lives and scientific lives are filled mostly with another kind of argument.

Not all arguments are deductive, and 'inductive' is the adjective labelling any non-deductive argument. Induction is the kind of argument in which we typically engage. "John is a man. Most men die before their 100th birthday.

Therefore John will die before his 100th birthday." The conclusion of *this* argument can, in principle, be false while the premises are true; the premises do not logically entail the conclusion that John will die before his 100th birthday. It nevertheless is a pretty good argument.

It is through inductive arguments that we learn about our world. Any time a claim about infinitely many things is made on the evidence of only finitely many things, this is induction; e.g., when you draw a best-fit line through data points, your line consists of infinitely many points, and thus infinitely many claims. Generalizations are kinds of induction. Even more generally, any time a claim is made about more than what is given in the evidence itself, one is engaging in induction. It is with induction that courtrooms and juries grapple. When simpler hypotheses are favored, or when hypotheses that postulate unnecessary entities are *dis*favored (Occam's Razor), this is induction. When medical doctors diagnose, they are doing induction. Most learning consists of induction: seeing a few examples of some rule and eventually catching on. Children engage in induction when they learn the particular grammatical rules of their language, or when they learn to believe that objects going out of sight do not go out of existence. When rats or pigeons learn, they are acting inductively. On the basis of retinal information, the visual system generates a percept of its guess about what is in the world in front of the observer, despite the fact that there are always infinitely many ways the world could be that would lead to the same retinal information—the visual system thus engages in induction.

If ten bass are pulled from a lake which is known to contain at most two kinds of fish—bass and carp—it is induction when one thinks the next one pulled will be a bass, or that the probability that the next will be a bass is more than $1/2$. Probabilistic conclusions are still inductive conclusions when the premises do not logically entail them, and there is nothing about having fished ten or one million bass that logically entails that a bass is more probable on the next fishing, much less some specific probability that the next will be a bass. It is entirely possible, for example, that the probability of a bass is now *decreased*—"it is about time for a carp."

## What the problem is

Although we carry out induction all the time, and although all our knowledge of the world depends crucially on it, there are severe problems in our understanding of it. What we would *like* to have is a theory that can do the following. The theory would take as input (i) a set of hypotheses and (ii) all the evidence known concerning those hypotheses. The theory would then assign each hypothesis a probability value quantifying the degree of confidence one logically *ought* to have in the hypothesis, given all the evidence. This theory would interpret probabilities as *logical probabilities* (Carnap, 1950), and might be called a theory of logical induction, or a theory of logical probability. (Logical probability can be distinguished from other interpretations of probability. For example, the *subjective* interpretation interprets the probability as how confident a person actually is in the hypothesis, as opposed to how confident the person ought to be. In the *frequency* interpretation, a probability is interpreted roughly as the relative frequency at which the hypothesis has been realized in the past.)

Such a theory would tell us the proper method in which to proceed with our inductions, i.e., it would tell us the proper "inductive method." [An *inductive method* is a way by which evidence is utilized to determine *a posteriori* beliefs in the hypotheses. Intuitively, an inductive method is a box with evidence and hypotheses as input, and *a posteriori* beliefs in the hypotheses as output.] When we fish ten bass from the lake, we could use the theory to tell us exactly how confident we should be in the next fish being a bass. The theory could be used to tell us whether and how much we should be more confident in simpler hypotheses. And when presented with data points, the theory would tell us which curve ought to be interpolated through the data.

Notice that the kind of theory we would like to have is a theory about what we *ought* to do in certain circumstances, namely inductive circumstances. It is a *prescriptive* theory we are looking for. In this way it is actually a lot like theories in ethics, which attempt to justify why one ought or ought not do some act.

Now here is the problem: *No one has yet been able to develop a successful such theory!* Given a set of hypotheses and all the known evidence, it sure *seems* as if there is a single right way to inductively proceed. For example, if all your data lie perfectly along a line—and that is *all* the evidence you have to go on—it seems intuitively obvious that you should draw a line through the data, rather than, say, some curvy polynomial passing through each point. And after seeing a million bass in the lake—and assuming these observations are all you have to help you—it has just *got* to be right to start betting on bass, not carp.

Believe it or not, however, we are still not able to defend, or justify, why one really ought to inductively behave in those fashions, as rational as they seem. Instead, there are multiple inductive methods that seem to be just as

Springer

http://www.springer.com/978-1-4020-1176-4

The Brain from 25,000 Feet
High Level Explorations of Brain Complexity, Perception, Induction and Vagueness
Changizi, M.A.
2003, XXVI, 330 p., Hardcover
ISBN: 978-1-4020-1176-4