# *Causal Models*

## How People Think about the World and Its Alternatives

Steven Sloman

OXFORD
UNIVERSITY PRESS
2005

# Contents

# 1

# *Agency and the Role of Causation in Mental Life*

## *The High Church of Cognitive Science: A Heretical View*

How do people think? The effort to answer this question is the domain of cognitive science, a field of study that includes cognitive psychology and parts of computer science, linguistics, anthropology, and philosophy. The field emerged in lock-step with the development of the computer. After all, the computer is at heart an information-processing device, and it would seem that people are information-processing devices extraordinaire. Computers have input devices (e.g., keyboards, mouses, microphones), and people have input devices (e.g., eyes, ears, skin). Computers store information in memories; people store information in memories. Computers compute; they do calculations by transforming symbols in languages they understand (like Java and binary code); people compute by transforming symbols in languages that we understand (like English and arithmetic). Computers have output devices (e.g., screens, speakers, disk drives), and people have output devices (e.g., mouths, hands, feet). Indeed, it would seem, and it did seem for many years (it still does to some), that cognitive science would answer the question "how do people think?" by programming a computer to behave like a person. The program would be the answer. This was essentially the conclusion offered by Alan Turing in a famous essay called "Computing Machinery and Intelligence."[1] He carefully developed a clever way known as the Turing Test to decide if a computer could think by fooling a judge into believing that it was a man and not a woman.

Now there are contests modeled on the Turing Test. The Loebner Prize offers a gold medal and $100,000 for the first computer whose responses are indistinguishable from a human's. Each year an annual prize of $2000 and a bronze medal is awarded to the most human computer. Some machines have proven very clever, but nobody has built a machine that can come close to passing the test as Turing envisioned it. Nobody has yet won gold. One reason may be the complexity of thought or the huge amount of knowledge required to mimic even a young child. Think about how much knowledge is required to understand something as simple as a chair. You need to understand sitting (and that requires knowledge about the human form). You need to understand something about materials (chairs cannot be made out of powdered sugar). You can't even really understand what a chair is without understanding something about fatigue, the benefits of rest, and perhaps the importance and ritual of breakfast, lunch, and dinner. Knowledge is interrelated and therefore a critical mass is required even to understand the simplest things.

Disillusionment with the view that the computer is the best metaphor of mind, sometimes called the High Church Doctrine of Cognitive Science, has been widespread and has deep roots. People differ from computers in critical ways. For one, we compute differently. Traditionally, computers perform one operation at a time; they compute sequentially. People are able to perform many operations at a time; certain functions (like memory) involve the simultaneous operation of billions of simple units, resonating together to retrieve memories, in the way the entire body "remembers" how to ride a bike. Some recent computer designs involve a limited amount of parallel processing, but nothing like the parallelism that operates in the mind.

A second difference is that, unlike computers, people have emotional lives directly tied to the chemical composition and physiological processes of our bodies. We may be machines, but we're machines made out of meat, and that changes everything.

A third difference, and the one that I'm going to focus on, is that we're not passive consumers of information, blindly transforming symbols as requested. We're agents. We actively pursue goals, be they the need for food, oxygen, or love or the desire for entertainment, education, or liberty. Whether people have free will is a question far beyond the scope of this book. But it seems clear that we talk, think, and act as if we do. Therefore, to understand the mind requires a way of representing agency.

### *Agency Is the Ability to Represent Causal Intervention*

Understanding what it means to be an agent may seem to require answers to questions like "what is consciousness?" "how is intentionality coded in the brain?" and "what does it mean to be aware?" But agency can be understood in much more mundane, yet comprehensible, terms. Agency can be treated as nothing more than the ability to intervene on the world and change it. And I'm not even going to claim that people have that ability, although they probably do.

All I'm going to claim is that people *represent* the ability to intervene in the world and change it. Think about human knowledge. What do you know? One kind of thing you know is how things work in the world. You know about mechanisms. You know how to operate some machines; you know the buttons to press, the levers to push. You know how it behaves (it hums, it rattles, it lights up, it glides). You have some ideas about what makes people go, their motivations and behaviors. You know Johnny is driven by his stomach and Nancy by her pride. You know about the mechanisms that drive political systems, about power and economics.

This knowledge is all causal in the sense that it's about the mechanisms that bring about effects from causes. It's all about events that take place in time and describes not only how and when objects and events appear together but why. The answer to this why question comes in the form of descriptions of how things could be otherwise, not only which effects follow which causes but which effects would follow if the causes had been different. As the 18th century Scottish philosopher David Hume put it, causality is about "an object followed by another, . . . where, if the first had not existed, the second had never existed." This is precisely the kind of knowledge required to predict the effect of action, how behavior changes the world.

What do we really understand when we think we understand a mechanism? Presumably, at minimum, we have some idea about which inputs produce which outputs. We understand how the choice of inputs determines the outputs and that the reverse does not hold. The choice of outputs does not determine the value of inputs. This special and structured kind of knowledge requires that we understand that (1) changing X is likely to end up with a change in Y; (2) causes and effects are asymmetric: changing Y won't budge X; (3) causes and effects go together over time; and (4) Y does not occur before X. Believing that heat causes expansion requires believing that (1) changing the temperature will change the volume (of a gas, say); (2) changing the volume won't change the temperature; (3) certain temperatures are associated with certain volumes; and (4) new volumes aren't observed before new temperatures.

I'm defining knowledge as a set of beliefs about change in the world and the mechanisms that support those changes. Knowledge is about how changes in some things lead to changes in other things. In other words, what we know about the world is how things could have been otherwise. Representations of causality allow us to describe how the world would have been—that is, another possible world—if some cause had had a different value, for then its effects would have been different. In cognitive science, this is the domain of counterfactuals, beliefs/statements about worlds that are not necessarily this one, like the world in which today is a balmy 18°C, with a light breeze blowing off a sky-blue sea (fortunately, sometimes counterfactuals are about the world as it actually is).

So agency concerns intervention on the world to change it, to see how things might have been otherwise. But even more important, agency is about how we represent intervention, how we think about changes in the world. Because by being able to represent it, we are able to imagine changes in the world without actually changing it. And that ability opens up the possibility of imagination, fantasy, thinking about the future, thinking about what the past might have been like if only. . . .

## *The Purpose of This Book*

This book is about how to understand causal systems. But causal systems can't be understood without an analysis of cognition, for a causal claim is intrinsically a claim about beliefs, not merely a claim about the way the world is. Causality isn't just a figment of our imagination entirely independent of the actual world (if it were, then we'd be free to impose any causal relations we wanted, between, say, touching things and having them turn into gold). But causality is a form of construal. We impose a causal frame on the world to understand it, a frame that tells us about the mechanisms that not only produced the world as it is but that (counterfactually) govern the world however it had turned out.

Causal frames allow us to understand certain things (the mechanisms that take us from one state to another), including very important things like why people act the way they do and how to build an aqueduct. They may even be important for understanding why people are so often enamored of magic and where some religious belief comes from. This book will argue that they're critical for understanding how people reason and make judgments and decisions, certain aspects of language, and how we think about moral, legal, scientific issues and more. But they're not necessarily helpful for allowing us to understand noncausal formal systems like counting or other aspects of the way we represent the world, such as the images we use to try to remember the color of our front door or whether any of our high school teachers had a moustache.

The book was inspired in large part by the development in computer science, statistics, and philosophy of a general framework for representing causal relations, powerful enough to express the diversity of causal relations, whether probabilistic or deterministic, necessary or sufficient, weak or strong, direct or indirect, actual or background.[2] The framework is largely based on a mathematical framework for representing probability known as Bayesian networks. My plan is to discuss the developments mainly at a conceptual, not mathematical, level. For mathematical introductions, I refer readers to some excellent technical books.[3] What's unique about the causal model framework is that it gives us a way to think about the effect of action in the world. It makes the claim that the cognitive apparatus that people use to understand the world has a specialized operation that encodes certain changes in the world as the effect of agency, of intervention. These representations differ from other representations of change in being easily reversible. They invoke localized changes to causal models that are simple but have enormous effects. And what's so useful about them is not only that they give us a way to represent action but that they also give us a way to represent imagined action, how we think about the way the world could or would be if such and such were different.

The framework is extraordinarily rich but highly technical, though I'll discuss technical details only in a cursory way. What I'll do instead is to draw out the implications of the framework for cognition, what it says about how people reason, decide, judge, imagine, classify, talk, and how we learn to do all this in just a few short years of life. To foreshadow: in large part, we don't learn to do these things ourselves but depend very much on the people and institutions around us.

## *Plan of the Book*

This book will be a long argument that causation is central in how humans understand the world. The first part will focus on the theory of causal models. Chapters 2 and 3 will provide a conceptual introduction to the significant ideas. Chapter 2 focuses on why cause matters and chapter 3 on what causes are. Chapters 4 and 5 provide the technical meat of the book; chapter 4 is about causal models generally, and chapter 5 is about the representation of intervention. Although I've tried to keep these chapters as light as possible, too light for more mathematically sophisticated readers, less technical readers are liable to find them hard going. The rest of the book can be read fruitfully without them. Nevertheless, I encourage all readers to struggle through them, because the excitement that ideas about causation and intervention have engendered in so many psychologists and other scientists, philosophers, and applied mathematicians can be shared only after having glimpsed the richness and detail of the ideas.

The second part of the book applies the theory to various domains of everyday life: how we reason (chapter 6); how we make decisions (chapter 7); how we judge (chapter 8); how we categorize objects (chapter 9); how we induce properties of the world (chapter 10); how we use aspects of language (chapter 11); and how we learn about causal structure and the strength of causal relations (chapter 12). These chapters are ordered according to how naturally they follow from the theoretical discussion of part I and how smoothly they follow one another. None of them is particularly difficult, but reading the earlier chapters first should make the later ones easier to fathom. But it may not, and the reader should feel free to pick and choose the chapters whose topics suit his or her fancy. Much of the scientific work on causal models has focused on learning. Because that topic is the most complex, I've left it to the end. Keep in mind that my intent is not to provide exhaustive surveys of each subfield but to give a sense of the central role that causal models play in each.

The final chapter, chapter 13, attempts to draw out the general lessons of causal and interventional principles for an understanding of people, their situations, their problems, and solutions to their problems.

# 2

## *The Information Is in the Invariants*

After locating food, a honeybee will return to its hive and perform an elaborate dance to signal the food's location to its cohort. Most bee experts believe, following Nobel Prize–winning work by Karl von Frisch, that a bee that has located a source of nectar more than about 100 yards away will return to the hive and repeat a figure eight pattern that includes a straight movement in which the bee waggles or vibrates its abdomen for a few seconds and then circles back first to one side and then to the other. The orientation of the straight movement relative to vertical tells the other bees the direction of the nectar relative to the sun; the duration of the waggles reveals the distance of the flower patch from the hive. The bee's dance focuses on only two critical elements for finding food; it presupposes that the only thing the other bees care about is the direction and distance to the food source. They don't care what the color of the terrain they'll be flying over is or whether it's made of rocks or sand. Neither of these things matters when a bee flies through the air to a flower patch. This example illustrates that animals are selective in what they attend to. They attend to the properties that serve their goals as best they can, while ignoring the properties that are irrelevant.

A useful skill for all moving organisms is the ability to predict how long it will take before an approaching object (like a wall) is bashed into. This is called time to collision. Being able to estimate time to collision can help one avoid obstacles when flying (if you're a bee or a pigeon), jumping (if you're a locust), or driving (if you're a human). It can also help you catch or hit a ball when it is the object that is moving and you are stationary. It turns out that when objects are neither speeding up or slowing down but moving at a constant velocity, time to collision can be estimated through an incredibly simple calculation based on the fact that the image of an object on your retina gets larger as you get closer to the object (or it gets closer to you). To estimate time to collision specifically, all you need to know is how big the image of the object is on your retina and how fast the size of the image is expanding. Divide the image size by its rate of expansion and that's it; that calculation gives you an excellent estimate of how long it will be before a big bang. This estimate is called tau (it was first derived in a science fiction novel by astronomer Fred Hoyle[1]). A fair amount of evidence suggests that locusts, birds, and people are all sensitive to tau when performing tasks that require estimates of collision time (like jumping out of the way of a looming entomologist, in the case of a locust). In fact, a type of neuron in locusts has been identified that is tightly tuned to the time to collision with an object approaching in a direct collision course.

Tau exemplifies how the environment can give us rich information in a simple form about something important. By ignoring almost everything about an object except the size of its image on the retina and how that image is changing, we render vital information immediately available for making contact (when catching) and avoiding contact (when navigating). Tau is so useful because it refers to a relation that doesn't change, that is *invariant*. The size of an image relative to its rate of expansion consistently indicates time to collision in many situations.

When learning to drive a car, a human being will appreciate that the steering wheel turns the wheels and that the brake will cause deceleration. They might notice the color of the upholstery, but they will also be aware that the color has no effect on the car's motion. When getting into a car they have not driven before, a person won't flinch in the face of new upholstery (even if it's green), but they'll be at a complete loss if there is no steering wheel. In this case, we have selective attention to invariants that reflect causal relations that we are aware of. We know that steering wheels cause changes in direction and brakes cause deceleration.

Taken together, these examples show three properties of behavior. First, animals are selective in what they attend to. Second, animals attend to what is stable—to invariants—because that's where the crucial information is for helping them achieve their goals. Third, at least for humans (and probably for other animals), invariants can take the form of causal relations. Causal relations carry the information that we store, that we discuss, and that we use for performing everyday activities that change the state of the world. Let's consider each of these three properties of behavior.

## *Selective Attention*

The world is infinitely rich and complicated According to William James, babies experience the world "as one great blooming, buzzing confusion." James's contention would have been right, and would be applicable to adults, too, if perception didn't impose structure, so that we do not pay equal attention to every sensation and fact.

Imagine taking note of every fact around you at the moment, every shadow, every illuminated surface, every object, and the parts of each object. And then there are the relations among each of these things. Which surfaces are obstructing the view of which other surfaces? Which objects are supporting which other objects? What is the purpose of each object? Why is it there? If we took note of each of these things at every moment of every day, we'd collect enough information to fill several libraries every month. But who would care? What use would it serve anybody to take note of the fact that yesterday a dirty glass was beside the microwave and not the toaster before it was washed? Most facts are useless, and taking note of them would merely clog our minds.

The famous Russian psychologist Alexander Luria had a patient ("S") who had a fabulous memory.[2] He could reproduce a table of 50 digits after studying it for only 3 minutes. He could memorize long equations even though they made no sense to him, and he could memorize poetry in Italian, a language he did not speak. He could remember countless arbitrary facts based largely on a rare talent for constructing and maintaining vivid, detailed images from multiple senses: vision, audition, even taste. Yet he had trouble finding structure and meaning. Given a series of numbers that progressed in a simple logical order:

1 2 3 4
2 3 4 5
3 4 5 6
4 5 6 7,

he could reproduce the series yet not notice the logic of the sequence! Indeed, he would take as long to memorize a sequence like this as a sequence consisting of random digits. He complained that the vividness of the images that he evoked would clog his mind and make it hard for him to understand text and speech. The onslaught of images would confuse him and made it hard for him to maintain a job, except as a professional mnemonist in theaters. The ability to remember is useless without the ability to pick and choose what is important and to put the useful pieces together in meaningful ways.

Organisms must be selective; they must attend only to those things that carry the information relevant to their general interests and current goals to guide them through the world. A bee looking for nectar focuses on direction and distance from the hive, and a pigeon trying to avoid a tree focuses on the image of the tree and how fast it's expanding. Similarly, if you want to make friends and influence people, you should attend to their attitudes, behavioral responses, and their names. You can feel free to ignore how many freckles they have on their noses and, usually, whether they've passed through Toledo, Ohio. A person looking for entertainment would do well to head toward a stereo or bookshelf rather than the broom closet. Even the broom closet can be ignored from time to time.

### *Selective Attention Focuses on Invariants*

So what do we select to attend to? Obviously, the answer is ''it depends.'' If we're looking for shoes, we attend to shoe shops and if we're looking for hats, we attend to hat shops. More abstractly, we look for the places, the objects, the events, the information that will satisfy our current goals. We can think of selective attention as solving a problem. We want to get somewhere, to satisfy some desire or requirement, so we have to find those aspects of the environment that hold the solution so that we can limit our attention to them.

Generally, we know just what to do because we've done it before or because somebody tells us what to do. We already know how to start a car so we immediately attend to the location of the key, the brake, the accelerator pedal, and in some cars the clutch. We know how to purchase something in a department store, so we look for the cash register. And if we don't know these things (as we might not in a foreign country), we ask. In each case, what we need to know is something that holds not just in a single case but in all, or at least many, relevant cases. The great value in knowing about car keys and accelerator pedals is that so many cars have them. The utility of knowing about cash registers is that they have the power to give us ownership of a product in so many different stores.

We have a kind of expertise in car starting and consumerism that derives from knowing what *doesn't change* across instances and across time, that is, from knowing what's *invariant*.

Expertise inevitably involves the ability to identify invariants. Experts can do far more than list facts about their area of expertise. They can tell you what's important and what's more peripheral because they know what explains the way things are and they know what predicts the way things will be. An electrician focuses immediately on what's connected to what and how they're connected, ignoring whether the connection is visible or hidden behind walls; a plumber looks at the angle (or *fall*) of a pipe to make sure it's steep enough to allow water to flow but not so steep that the water will flow too fast, ignoring the design of the pipe manufacturer's trademark. A psychological counselor helping someone who is depressed will focus on the human relationships in that person's life, ignoring what their favorite color is.

In each case, the expert has picked out the properties that explain why the system (electrical, plumbing, or emotional) is in its current state and that predict the state of the system in the future. What allows the expert to do so is that these critical properties are the same from occasion to occasion. It's not the case that plumbing systems sometimes depend on the trademark design and sometimes on fall, and it's not the case that electrical circuits sometimes depend on how things are connected and sometimes don't. The properties that matter are invariant across situations and across individuals; otherwise, they wouldn't be learnable. So the knowledge that gives us expertise is knowledge about those aspects that are durable, that don't change with the wind, but that represent the stable, consistent, and reliable properties that hold across time and across different instantiations of a given system.

This is not to say that picking out invariants is always easy for an expert or that experts don't sometimes make mistakes. And experts can pick out different invariants over time because experts continue to learn and because domains of expertise change as discoveries are made and theories change. The point is that, at a given time, experts are usually faster and more accurate than novices in identifying critical invariants in their domain. When they're not, you should ask for your money back.

Chicken-sexing provides a fascinating example of experts' ability to pick out invariant properties. Deciding whether a baby chick is male or female can be very important. In some locales, there may be regulations against owning roosters within city limits. Also, one may want to limit the number of roosters one has because they sometimes fight. Or a farmer may want only hens and the eggs they produce. Therefore, farmers often want to know the sex of newly hatched chicks so they can keep the females, and "sacrifice" the males. To do so, they hire a chicken-sexer, an expert in distinguishing male and female chicks. This is not an easy task. It is an art developed in Japan. The chick has an external opening called the *cloaca,* which serves digestive, urinary, and reproductive purposes. This opening is closely examined for a degenerate penis, which is found in all males but also 15% of females. The developed skill of a professional chicken-sexer is in determining the sex of this 15%. It requires effort and many hours of training and practice. Novices are not much better than chance at determining a chick's sex.[3] But experts get faster and faster. The best sexers in the world are able to sex about 800 chicks an hour with 99% accuracy. They have so developed their ability to distinguish the male from female pattern that they can do it instantly. In other words, they have so trained themselves to pick out what's invariant about chicks of each sex that they can make a decision based on that invariance immediately.

The search for invariants is not limited to experts. It is something that we all do and can be observed at the lowest levels of perception. Perception involves discovering the cues that consistently signal things of interest to distinguish them from noise. This is why the tau variable is so useful. It refers to an invariant relation between moving agents and the obstacles they could encounter and so provides all the information that an organism needs to negotiate its environment without bumping into things. A variety of variables of this type have been identified by psychologists interested in how we use information in the environment to control our behavior.

The father of this approach to the psychology of perception was James J. Gibson.[4] He referred to these aspects of the environment that we attend to and that control our behavior as "affordances." Gibson believed that the world has invariant properties and relations that provide the necessary information to afford organisms the opportunity to fulfill their perceptual goals. Gibson talked about perception being "direct." Rather than conceiving of organisms as processors of information who use sensory mechanisms to grab a variety of evidence that is then processed to extract what is relevant (recall the High Church view of cognition from chapter 1), Gibson believed that organisms are built to respond directly to relevant invariants in the perceptual stream. He believed that organisms can be compared to radios. Radios don't compute; rather, they are tuned to transmissions in the environment. Radio signals are out there ready to be received. For this analogy to work, you must believe that adequate information is already in the environment to allow people to accomplish their goals, that this information is invariant, and that people are somehow already tuned to pick up this and only this information.

This book does not endorse Gibson's approach as a general theory of how the mind works. For one reason, adequate information for achieving goals is not always in the environment but often must be constructed on the basis of memory and inference (imagine trying to fly a modern jet airplane without memory and inferential procedures!). But this book will endorse the idea that people search for invariants relevant to satisfying their goals and that those invariants are used to guide thought and action.

The focus on invariants can be observed beyond the perceptual system. In general, prediction requires identifying the variables whose behavior is constant over time so that their future behavior can be derived from their present values. Predicting the weather requires focusing on wind direction, pressure gradients, and so on because these are consistently related to changes in the weather. Explanation, like prediction, involves assimilating an observation or phenomenon to a process or representation that applies generally, that emanates from or instantiates relations that are regular. Explaining someone's behavior involves appealing to personality traits of that person that persist over time or to forces in the person's situation that would cause most people to behave that way. Beyond prediction and explanation, control requires knowing the systematic relations between actions and their outcomes, so the right action can be chosen at the right time. A good politician will know who is motivated by greed and who is motivated by larger principles in order to discern how to solicit each one's vote when it is needed. In all these cases, the secret is to identify and use invariance, the constant, regular, systematic relations that hold between the objects, events, and symbols that concern cognition.

## *In the Domain of Events, Causal Relations Are the Fundamental Invariants*

Where should we look for invariance? It seems to be hiding. All around us, we see change. Patterns of light and sound come and go constantly as we move about, as the sun shifts, as someone speaks. People change over lifetimes, species over evolutionary time. The sun has even set on the British empire by some accounts. Everything that has a physical realization is transient. So those things that are physically realized, objects and events, do not wear invariance on their sleeve.

But it must be there. We wouldn't be as certain as we are that (say) money buys goods, gravity holds things to the ground, war will occur again somewhere, or that we'll take our next breath if we couldn't rely on some things not changing. In none of these cases does our certainty derive from direct observation by our senses. It's not the world itself that doesn't change; everything physical changes. It's the generating process that produces the world that doesn't change. For example, the process that produces trade doesn't change. People have desires and needs and the intelligence and social structure to realize a means to satisfy some of those desires, such that mutual benefit through trade is common. This process is causal; it involves causal mechanisms that turn supply and demand into trade. In the case of gravity, too, the invariant is the causal mechanism that produces gravity. It takes mass and distance from the center of a body and produces a force as an effect. It is constant and it is pervasive, though physicists debate whether it holds only in this universe or whether it holds in every possible universe. Either way, it is a good process to bet on if you are a mere mortal. It is as reliable an invariant as the laws of mathematics, though only the former is obviously causal.

So relations of cause and effect are a good place to look for invariance. The mechanisms that govern the world are the embodiment of much that doesn't change. They don't embody all invariants: mathematical relations are invariant without necessarily being causal. But they embody a good part. The physical world, the biological world, and the social world all are generated by mechanisms governed by causal principles.

The invariants picked out by the experts mentioned earlier all concern the operations of mechanisms. The electrician cares about how things are connected because the causal pathways that govern electricity are determined by electrical connections. Similarly for plumbing. The fall of a pipe matters because the mechanism that controls drainage involves gravity operating within the walls of a pipe. A psychologist focuses on human and not material relationships because our emotional health is always a function of our relationships to other people and related only weakly and indirectly to our material possessions. In later chapters we'll see that prediction does not always require appeal to causal mechanisms because sometimes the best guess about what the future holds is simply what happened in the past. But sometimes it does, especially when there is no historical record to appeal to. And we'll see that explanation and control depend crucially on causal understanding.

That causal relations are invariant should come as no surprise. After all, much of science is devoted to discovering and representing invariant causal structure (e.g., force changes acceleration, demand increases price, structure determines function, life requires energy, power breeds corruption). The causal principles that govern mechanisms are so useful because they apply so widely. They apply across time (yesterday, today, tomorrow), and they apply across large numbers of objects. Every physical body in the universe is subject to inertia, and every living organism consumes energy, now and always.

In sum, people are exquisitely sensitive to the invariance of causal relations, the relations that govern how things work. Notice it is not the working mechanisms themselves that are invariant but the principles that govern them. A car engine is a causal mechanism designed to be reliable and in that sense invariant, but—as we all know—that is a pipedream. Shadows are cast by objects that block a light source like the sun, but the precise form of the shadow-casting mechanism may change continuously as the earth rotates and clouds shift. The invariant is not the form that the mechanism takes but rather the principles that govern its operation. The drive shaft of a car is not invariant, but the principle relating torque to force is. And the sun-cloud mechanism that produces shadows is changing all the time, but the principles that govern how light and objects produce shadows are constant. Causal principles that govern how events affect other events are the carriers of information, and it is those principles, not the mechanisms that they govern, that persist across time and space. They are the most reliable bases for judgment and action we have.

One might argue, along with philosophers like Bertrand Russell,[5] that invariant laws aren't causal at all but rather mathematical. Metaphysically, this may indeed be correct. Perhaps the world is nothing but a flow of energy. Perhaps there is no will, no agents, no intentions, no interventions, just the transformation of energy following certain eternal mathematical rules. Perhaps there is no root cause of anything, and perhaps there is no final effect either. Perhaps everything we misconstrue as cause and effect is just energy flow directed by mathematical relations that have determined the course of history and will determine our destiny in a long chain of events linked by the structure of energy in time and space.

Maybe. But this book isn't about metaphysics. It's about representation. It's about how people represent the world and how we should represent the world to do the best job of guiding action. The central idea of the book is that the invariant that guides human reasoning and learning about events is causal structure. Causal relations hold across space, time, and individuals; therefore, the logic of causality is the best guide to prediction, explanation, and action. And not only is it the best guide around; it is the guide that people use. People are designed to learn and to reason with causal models.