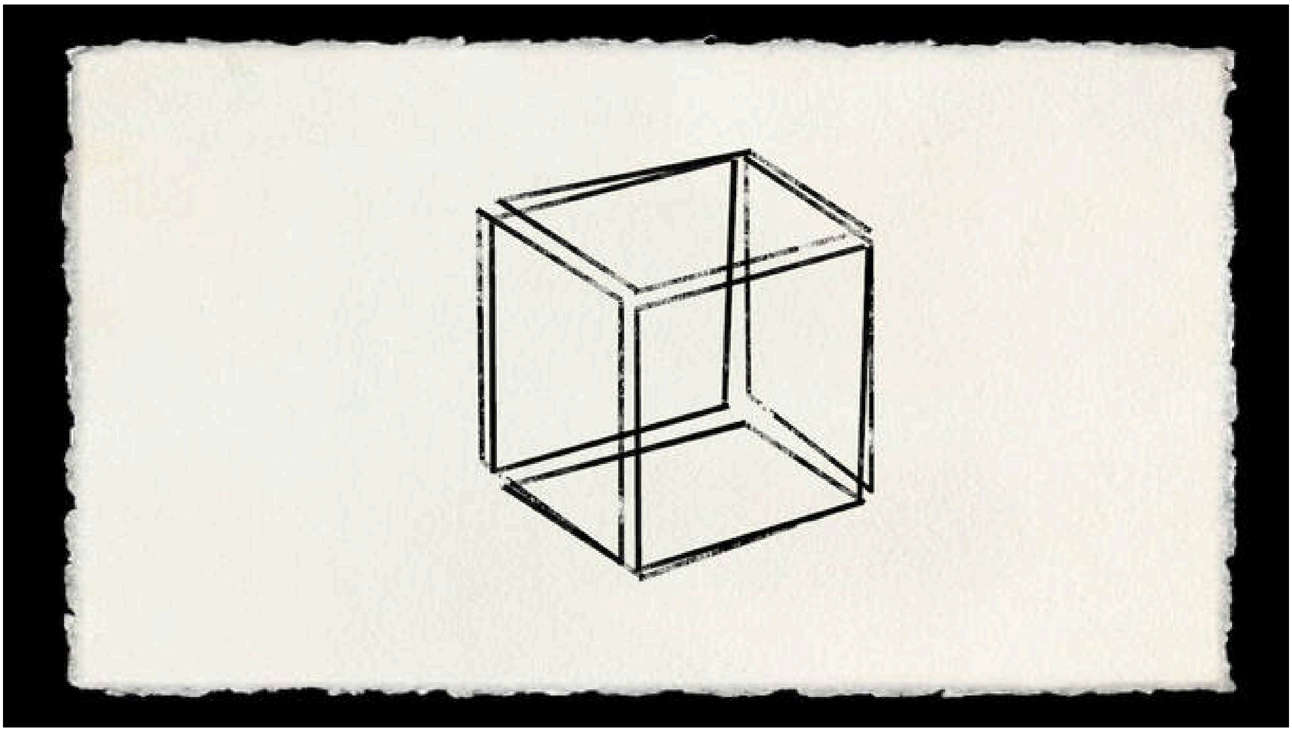
11. The Art of Modelling



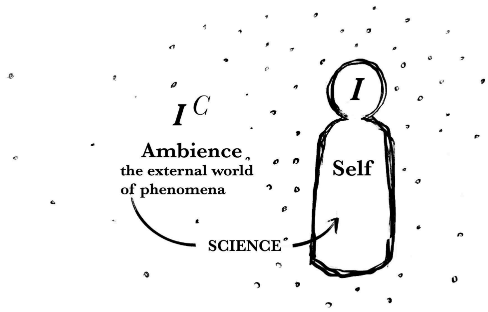
Pattern

You may wonder: where is this formal model of doing scientific research that we promised? Well, we're getting there — step by step. In fact, we've already started introducing it. Perhaps you haven't noticed?

In the last chapter, we focussed on the *primaeval epistemic cut*, the basic *logical disjunction* that separates *self* from *ambience* — Rosen's *first dualism*. This cut is absolutely necessary for any limited being to pick out *differences* from the ambience *that make a difference*. But it also separates and isolates our first and third person perspectives, which is why we tend to repeatedly relapse into dualist ways of seeing reality.

The good news is: as living beings, we are uniquely capable of *bridging the epistemic cut* by achieving *semiotic closure*. Getting to know the world means to *bring* relevant parts of the ambience *inside* our selves. Only organisms can do this; machines (including algorithms) cannot, because they have no self. Science is just a particularly robust way for humans to bridge the epistemic cut, which enables particularly sound decision-making and coherent action. It is an adaptive process that continues indefinitely. We constantly get better at learning new things about the large world we live in. All of this explains the undeniable and incomparable *empirical success* of science. It is one of the most useful and powerful practices humanity has ever invented — absolutely essential for our continued survival. And this, of course, means we need *more* science, and the *best* science we can possibly get. But we may be repeating ourselves here...

So then, let's visualise this first fundamental distinction to reveal its formal nature:



This diagram *formalises* the discussion of the previous chapter by representing the "self" as an abstract mathematical structure: a *set*. We call this set "*I*" (as in "*I* experience therefore *I* am"). We'll talk more about sets in the next chapter (see also the appendix). For now, simply think of a set as an abstract bag of entities, a conceptual container for its *members* or *elements*, as indicated by the circle in the diagram. These members completely define what the set is: same elements, same set (mathematicians call this the *axiom of extension*). In other words, a set is a *collection* of specific entities (possibly including other sets).

But what do we aim to achieve by depicting the self as a set? In essence, a set represents the claim that the contents and boundaries of the self are properly definable (at least in principle). Actually, the self plays a big role in defining itself! We'll have a lot more to say about this in the rest of the book. In contrast, note that the ambience is *not* a set, because we *cannot* properly define its contents, nor can we delimit its extent, because it is impossible to quantify (or even recognise) our *unknown unknowns* — the *semantic residue* of our large world. This is why we have drawn no boundaries around the ambience. It is only defined negatively, as the *complement* of set *I*: that which is *not I* (which mathematicians write as I^C).

Remember the *exclusive or* of this logical disjunction: something is part of the ambience if (and only if) it is *not* part of our self. This is how we formally define the primaeval epistemic cut. But does it make sense? Clearly, self and ambience physically *interact*. Energy, matter, and information are constantly exchanged between them. Otherwise, we could not bridge the epistemic cut. But this is not the point.

Our unambiguous distinction between self and ambience should not be interpreted in *physical* terms. Instead, it is an *epistemic* distinction, necessary for a properly grounded *theory of knowledge*. It is about how we identify ourselves as *epistemic agents*. Whenever we make an observation, whenever we perform a measurement, *we* divide the world, *we* rely on the epistemic cut, so that nothing can be part of the self and the ambience at the same time, *by definition*. And we really have no choice in this: this is how we *have* to set ourselves up in order to get to know our world. Otherwise we can't know *anything*! Both first and third perspectives are necessary. There is no way to gain knowledge from nowhere. The self is our *locus*, our *home base*, the place *from* which we can have a perspective *on* the world.

Now that we have anchored our perspective in the primaeval epistemic cut, we can start thinking about *how* to *bridge the cut*. How can we build a sound and reliable *nomological order* for our selves? This, of course, is where science comes in as our most handy tool. It boosts our ability to *structure* that part of the ambience that we *can* and *do* experience. We've clarified earlier that our *reality* consists of those *phenomena* in the ambience that we perceive, and that somehow have an *impact* on our selves. To get to know the world means to structure this experienced reality — the part of the ambience that affects us but is *not* under our own control. The question is: where do we even begin with such a formidable task?

As already established, it all starts with the *problem of relevance*: the first step in getting to know our world is to pick out *relevant* phenomena from the ambience. This is what it means for an *agent* to delimit their *arena*, to bring forth a *landscape of affordances*. These affordances present themselves to us as *opportunities* and *obstacles* — as (positive or negative) aspects of *problems* we need to solve*, decisions* we need to make — in the pursuit of our *goals*. Let us put it like this: knowledge generation is *affordance landscaping* first, and *problem solving* second. As limited beings in a large world, we must first *frame* our problems, we must properly *define* them, before we can solve them to make decisions and pursue a coherent set of actions.

This is the part of reality that *is* under our own control: it is totally up to us to identify the appropriate affordances, to build the right kinds of practical and conceptual tools, and to pick a coherent course of action given those affordances and tools. In this sense, we are the landscapers of our own destiny! But always within boundaries over which we, as limited evolved beings, *do not* have full control. This is why it is so important to recognise our limitations for what they really are.

As a logical next step, therefore, we need to figure out *how* we can structure our ambience in a way that yields a manageable and effective arena which helps us survive and thrive. Every living organism has to meet this challenge in their very own way. Human science can achieve it (as we've also said before) through the *skillful modelling* of specific *phenomena* in the ambience. Simply put, science is nothing but a particularly systematic and rigorous approach to *subdividing* the ambience into smaller (preferably well-defined) chunks, to pick out and identify relevant features, and to model them in order to understand how they come about and to predict their future behaviour. But what are those features? What is it that we are modelling? And what exactly do we mean by a specific "phenomenon?"

One option is to propose that we subdivide reality into ever more fine-grained *objects*, indivisible *atoms* perhaps, or even more fundamental *particles* (when atoms turned out to be fissile after all). This is the path that fundamental physics initially chose — followed by the mainstream of most other branches of science. It also seems to be the approach that most modern humans intuitively tend towards. George Lakoff and Mark Johnson, in their "*Metaphors We Live By*," call it the psychological *doctrine of containment*. It is the view that the ambience is a kind of container for objects that can move around and change properties over time. These objects are themselves containers whose properties can be explained by the smaller objects they contain. And so on. We could call it the *Tupperware model* of reality: boxes within boxes — all the way down. Objects cause things to happen mainly by bumping into each other. These bumpings are what lies at the root of the phenomena we perceive, the events we can experience.

Does this sound familiar? Well, it *is*. The doctrine of containment is nothing but the psychological foundation for the machine view of the world. Remember James Ladyman's "*microbangings*" from the last chapter? Reality explained as a game of billiards. We can see now why this is so appealing and so deeply ingrained in us. Early on in our lives, as toddlers already, we learn how to pick out particular *objects* from the ambience. But not only that: the most widely used human languages, and the logical formalisms that underlie our mathematics and sciences, are also suffused by this doctrine. In fact, set theory, with which we describe the epistemic cut above, is a good example: sets are the quintessential containers that contain containers. Philosophers call this the *substantivist stance*. On this view, the world primarily consists of static *things* (based on some kind of physical *substance*). Substantivism primes us to ask: "*what is the world made of?*" Events and phenomena are secondary, caused by stable objects bumping into each other.

It should come as no surprise that we don't like this view. Not at all, in fact. And there are good arguments against it. For one, an object that does not interact with *anything* cannot impact us either, and thus cannot be real in the sense in which we have defined the term previously. Obviously, what is not real cannot be fundamental either, right?. Therefore, it makes no sense to us to assume that the existence of an inert object somehow precedes the existence of its interactions with other objects.

A related problem is that substantivism requires us to define an object in terms of its *intrinsic properties* only — what John Locke called *primary qualities*. Think of the atomic number of gold as a chemical element, or the mass of an object more generally. Yet, no object in the universe is truly isolated, and object properties always depend on context. In particular, it depends on the relations an object has with *other objects* — what Locke called *secondary qualities*. Mass, for instance, manifests as the weight of an object (which is what we measure), and weight depends on the gravitational field it is currently exposed to. And if you think about this a bit more, then you realise that all primary qualities *must* be *relational* (*i.e.*, *secondary*) in the sense that they *have to* be defined through *our interaction* with an object in the first place.

On top of all this, no object is ever truly static. Literally nothing in the universe lasts forever. The most stable thing we know of today is a proton, and it also decays, although with a ridiculously long half-life of at least 1.67 x 10^{34} years, which is about 1.21 x 10^{24} times the current age of the universe! In practice, this means that *very* few protons have decayed since their formation, just a few seconds after the Big Bang. Yet, even these very stable entities are not eternal. They *do* have a determinable expiry date, even if it lies on a time scale our human brain cannot even begin to comprehend.

One last problem with the substantivist stance is that it often has to attribute some mysterious kind of *motive force* or *agency* to the objects that constitute the world. What makes them move around? Why do they affect each other when they interact? Biology, for example, likes to explain many of its phenomena through the *action* of *genes*. But a gene is merely some kind of inert DNA sequence (biologists are still quibbling about its precise definition). You often hear that there is a gene "for" some trait or behaviour, be it blue eyes, or lactose intolerance, or aggressive behaviour, or intelligence, or whatever. But what this means is not clear: we have already discussed how genes do not *do* anything outside the context of a living cell. We may as well say it is the cell that reads the gene, rather than the gene that makes the cell.

Contemporary physics deviates from the substantivist view even more radically. Quantum field theory (as its name says) explains the behaviour of fundamental particles in terms of *fields*. But the concept of a particle-as-field does not resemble our traditional idea of an object at all. Instead, fields are abstract descriptions of particle *behaviour*. They are *not* made of substance in any straightforward way. And they do not have to be precisely localised, like an object, to a specific location in space and time. Instead, the properties of a field are diffuse, dynamic, often probabilistic, and manifest in a *relational* manner only, through interactions with other fields. In addition, quantum fields can be in a state of superposition (*i.e.*, in more than one state at once), and they can be entangled over large distances, due to their common past. It is all rather weird and counterintuitive at the scale of the very small, not like Tupperware at all.

For all these reasons, we adopt an alternative *processual stance* here, considering *processes*, *activities*, *flows*, *relations*, and *interactions* (not *objects*) as fundamental. This makes us ask "*what is happening in the world?*" And objects become just a special kind of process, one that is slow and stable compared to the time scale we are looking at. This stance is not only compatible with what we know about fundamental physics, but also with what we have argued so far, and will continue to argue in the rest of the book.

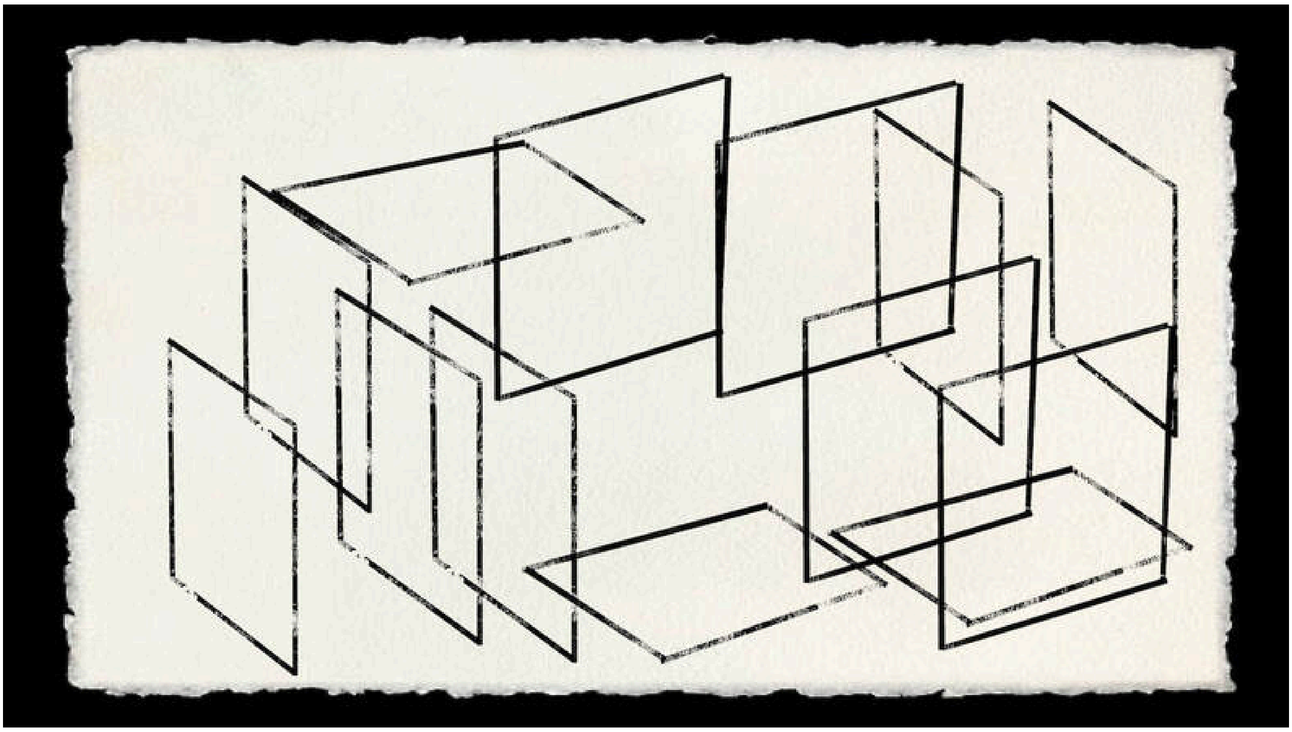
In practice, it means that we focus on *patterns*, rather than objects, in the ambience. The features we pick out are *distinguishable* and (at least to some extent) *reproducible regularities*. Conveniently, this includes conventional objects, but also more insubstantial, transient, and dynamic *phenomena* such as quantum fluctuations, solar flares, candle flames, electromagnetic and ocean waves, evolving populations and biological species, fluctuating weather conditions, human civilizations, philosophical schools, social relations, mood swings, or the thoughts that go through your mind while you are reading this. The last two are particularly interesting: you could think they belong to the domain of the self, but they don't if we treat them as the *object* of our study. Again, the distinction is *epistemic*, not based on physical nature or location. And we shouldn't forget the most interesting ambient patterns of them all: living organisms! They are quintessential *processes*, as we explain in detail in the last two parts of the book.

Of course, the classical approach to the *dynamics* of natural patterns is the Newtonian one. Indeed, we could say that classical physics picked out all the most obvious and most general patterns first to formulate its laws. But there are many more regularities to be found in the ambience, most of which are neither obvious, nor general or lawlike. Many are specifically situated, temporary, or only approximately reproducible in different contexts. Regularity comes in many forms and degrees, and patterns interact and influence each other in a myriad of ways, all of the time. And so, as Robert Rosen suggests, we need to extend Newton's idea of *natural law*, especially when dealing with complex phenomena such as those in the biological or social realm, which do *not* behave in simple, mechanistic ways.

Rosen's revised and extended idea of *natural law* is based on two basic principles. The first is the assumption that the ambience exhibits a certain degree of *orderliness* or *regularity*. It implies that each phenomenon appears for a reason, as the *effect* of one or many *causes*. Remember weak determinism? Without it, there could be no science, no natural language, and in fact no sanity, as Rosen rightly points out. Bill Wimsatt said it best: no living being could exist, survive, and evolve in a world entirely without order, a world that is completely random. In addition, Rosen postulates a second prerequisite for science: humans must be able to perceive, to grasp, and to comprehend these *causal patterns* or *regularities*. In Rosen's words: the orderliness of the ambience must be *discernible* and *articulable* by the human self, that is, it must be possible to make our *explanations* of the world congruent somehow with the *flow of causation* among ambient phenomena. Again, we've talked about this before — from quite a different angle though — when we first introduced the criteria for robust knowledge and coherent action.

Thus, Rosen's idea of *natural law* is that there are (more or less) reproducible and generalisable *patterns* in the ambience that we can identify as *phenomena* and understand in terms of *scientific explanation*. We'll come to what "understanding" and "explanation" mean in this context shortly. For now, let us note that Newton's kind of natural laws are both more broad (in terms of how general the regularities have to be) and more narrow (in the range of natural phenomena they apply to). Remember: Newton's master trick was to separate the box from its contents — to separate the rather haphazard initial and boundary conditions that delimit a specific phenomenon or pattern, from the underlying *universal laws* that determine its dynamics. Rosen's "natural law," in contrast, can be very localised. In fact, it is a bit similar to what contemporary philosophers call a "mechanism." But this is *really* confusing, because not all of Rosen's "laws" are mechanistic (*i.e.*, machine-like) in the specific sense we've introduced earlier.

There is surely a lot to unpack here, and it will take us several chapters — indeed the whole rest of the book — to do so. To begin with, let us simply agree *not* to use the term "natural law" any further. The word "law" has such strong Newtonian connotations that we immediately think of the equations of fundamental physics that underlie *all* of our reality. But we've rejected this idea already: even the most fundamental principles of physics are nothing but very broad models (limited to their specific domain and open to future revision) that pick out particularly general regularities or patterns that we observe in the ambience. They differ from more local causal models of less regular phenomena only by degree, but not in kind. They are just particularly robust and wide-ranging *scientific perspectives*. So why give them a distinct name? Instead, we will go with Ron Giere and imagine a science *without* laws. What Rosen calls "natural law" — the formal description of any kind of regular pattern, whether general or not — can be captured satisfactorily by the concept of a *formal system*, a particular kind of *scientific model*.



System

We now go a step further towards defining what a *formal system* is and how it relates to the phenomena that populate and make up our ambience. In the last section, we've mentioned that a *phenomenon* is some kind of *salient pattern* we detect in the ambience, a recognisable reproducible regularity that (for some reason) catches our attention and thus appears in our *arena* as an *affordance*. We achieve this by *relevance realization*. In addition, *events* occur (at least partly beyond our control) and impact our selves. *Phenomena* and *events* are closely related. They both arise transjectively, from our active engagement with the ambience, and they both influence the further course of action we select in pursuit of our goals.

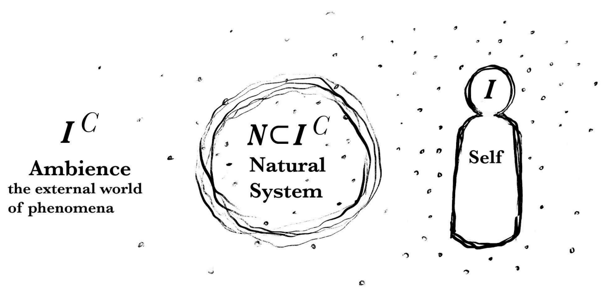
It is our ability to *realize relevance* that allows us to subdivide an ambiguous and ever-changing ambience into well-defined discrete chunks called *natural systems*. These systems stand for the particular phenomena and events that we want to model and understand. Rosen calls this subdivision his *second dualism*. Again, we want to avoid any confusion with the Cartesian kind: this second "dualism" just means that it is possible for us as *epistemic agents* to draw a boundary between a *system* and its *environment*.

But what exactly *is* a *natural system*? And how does it connect to *phenomena* and *events*? Well, first we need to make clear that there are two basic types of systems. A *formal system* is an abstract mathematical object that exists within the horizon of the self. In contrast, a *natural system* is some concrete delimited part of the ambience. Rosen defines it, somewhat vaguely, as a "collection of percepts that seem to us to belong together." There are many, more specific, definitions available from a wide range of sources. In essence, they all boil down to a natural system being some bounded assortment of parts that interact to form a whole. Here is one of our favorite definitions of this kind: Aloisius Louie (Rosen's student), in his book "*More Than Life Itself*," describes a *natural system* as "a subset of the external world … a collection of qualities, to which definite relations can be imputed." Let's parse that, from back to front.

Remember: the *pleroma* — the external world as it exists without any observers — is full of happenings, but has no particular form. It is our human interaction with it that turns it into *creatura*, the world of phenomena. This is our ambience, full of differences that could potentially make a difference to us. We *impute* our distinctions and meanings *onto* the large world we inhabit and engage with. In the case of a natural system, we express these distinctions as *relations*. In common parlance, this means that we attribute some kind of *connection* to the collection of qualities that make up a pattern or phenomenon in the ambience. Think of thunder arriving a little while after lightning. According to David Hume's account of *causation*, we cannot experience cause and effect *directly*. Instead, we infer it through what he called the *constant conjunction* of events. If thunder habitually occurs after lightning, we come to assume that lightning is the *cause* of thunder. There is a *causal relation* between the two. This, of course, is not (yet) sufficient for a robust scientific explanation, but it is not bad as a first step in the right direction.

But what about the "subset" part? What does it mean for a *natural system* to be "a subset of the external world?" Didn't we just say that the ambience is *not* a set? Indeed, we did. But this is not what matters here. Just as we can precisely delimit the extent of the self in contrast with the ambience, the whole point of defining a natural system is to draw a *specific boundary* that separates system (a set) and environment (*not* a set). This is not optional: as is the case for the self, a system can be *open* — in constant exchange of matter, energy, and information with its environment — but we still need to decide what is and isn't part of the system if we want to properly structure our ambience. Only if we draw a clear-cut border can we formalise a system, as we will see shortly. And yet again, drawing such a boundary serves our *epistemic purposes.* It is not an objective feature of the physical world, independent of our attempts to comprehend it. Instead, it is something that we transjectively impose on reality because it is useful (and, in fact, necessary) for our *understanding* of our large world as limited beings.

We can now draw another formal diagram to summarise our argument so far, with the formula reading "natural system N is a *subset* of the ambience I^C."



This diagram, however, is far from satisfactory. It leaves two important questions unanswered and, in fact, unasked. The first concerns the idea that a system forms some kind of a *whole*. How does it do that? This problem turns out to be closely connected with how we draw boundaries around a system. It is the question of *organisation*, which is central to our whole argument. The second is the question of how to turn a natural system into a formal one. Let's tackle the problem of organisation first.

When we say that a system consists of relations between a set of qualities, we mean that the system has identifiable *parts* that *interact* to produce some kind of *organised overall behaviour*. But what are these parts? In fact, they are bounded natural systems themselves! We call them *subsystems* or *components* in what follows. And this is an important distinction to make: system parts cannot simply be indivisible *things* or *objects*. Otherwise we end up in Tupperware World again, with inert objects as primary ingredients and interactions of secondary importance only. Instead, we insist that every system and all its components are *processes* that must, ultimately, be described in terms of *what they do*, not what they *are composed of*. Our approach is not only processual, but also fundamentally *relational.* A component that does not interact with other components is not a component. It's as simple as that. This means that the parts of a system are *defined* by the relationships they have, the role they play, *within a whole.* One comes with the other.

This may sound weird, ungrounded, and disconcertingly circular to you, especially if you have been conditioned to a rigid mechanical view of the world. It is certainly not as intuitive as being a substantivist. For one, you may ask: what is at the bottom of all this? What is the equivalent of an indivisible fundamental particle in this kind of description? What is the base level of reality? The bad news is: there simply is no such ultimate level of description. It's process all the way down! But this is not entirely true either, because there is no real "up" or "down" here, only processes that interact across time and space. Granted, they do this in a *hierarchical* manner, which is something we explicitly acknowledge when we talk about systems and their subsystems. But this hierarchy is only ever local and temporary. There is no global "top" or "bottom," because a system-level process can, at a later point in time, become a component of one of its current subsystems. Disorientating, isn't it? This kind of fluid nested organisation is called a *holarchy*, and we'll have lots more to say about it later.

Another problem we need to address is how the "whole" can influence the behavior of its parts. The good news is: there simply is no issue here if we consider a world of processes! We've talked about this already when we introduced the idea of *constraints* and their *closure*: component processes mutually restrict each other in their behaviour when they interact within the context of a system. Mystery solved. We'll come back to this too, later on. For now, let's just say that the idea of being a "whole" has something to do with *coherence*. For a phenomenon or pattern in the ambience to stand out, it needs to exhibit some degree of *unity* or *integrity*. It must do something that is recognisable and (at least to a certain extent) reproducible. This is, after all, what we mean by *organisation*: it is what underlies and coordinates the orderly and orchestrated behavior of a collection of parts and their interactions.

For this reason, organisation is also what allows us to draw specific borders around a system. Think of the weather, for example: a *cold front* is difficult to delineate precisely in terms of its extension across time and space, but it is much easier to list the factors that are required to account for its behaviour. It involves a cool body of air at the ground level replacing warmer air above it, which causes the characteristic wall of clouds to form, with rain (or other forms of precipitation) coming in beyond the front. Each component of this system consists of enormous numbers of interacting subcomponents ("air molecules", wind gusts, rain drops, and so on), which are highly dynamic, resulting in diffuse and shifting physical boundaries. And yet, the basic pattern of such a front is quite simple, and easily recognisable if you know the least bit about meteorology and are paying attention to what is going on.

A cold front is not an object, not a static thing, but it is definitely a *phenomenon*, a natural system we can identify and characterise. But what then is an *event*? We haven't explained this concept yet, because it requires us to first understand what a system is (which we now do). An event, then, is what occurs when two natural systems *interact*. When your toe hits the door jamb, for example, causality *is* directly (and painfully) knowable. We think that Hume should have gotten out of his armchair more often to experience this. Similarly, it's not the fall that kills you, but the sudden stop at the end — your interaction with the ground. Air masses colliding, billiard balls bouncing off each other, bodies exchanging heat, and Maxwell's demon meddling with the shutter of its box, those are all instances of events. And so is a bacterium sensing toxins or food, or you, dear reader, absorbing and parsing these letters. Yes: perceiving a phenomenon is also an event — an interaction between natural systems.

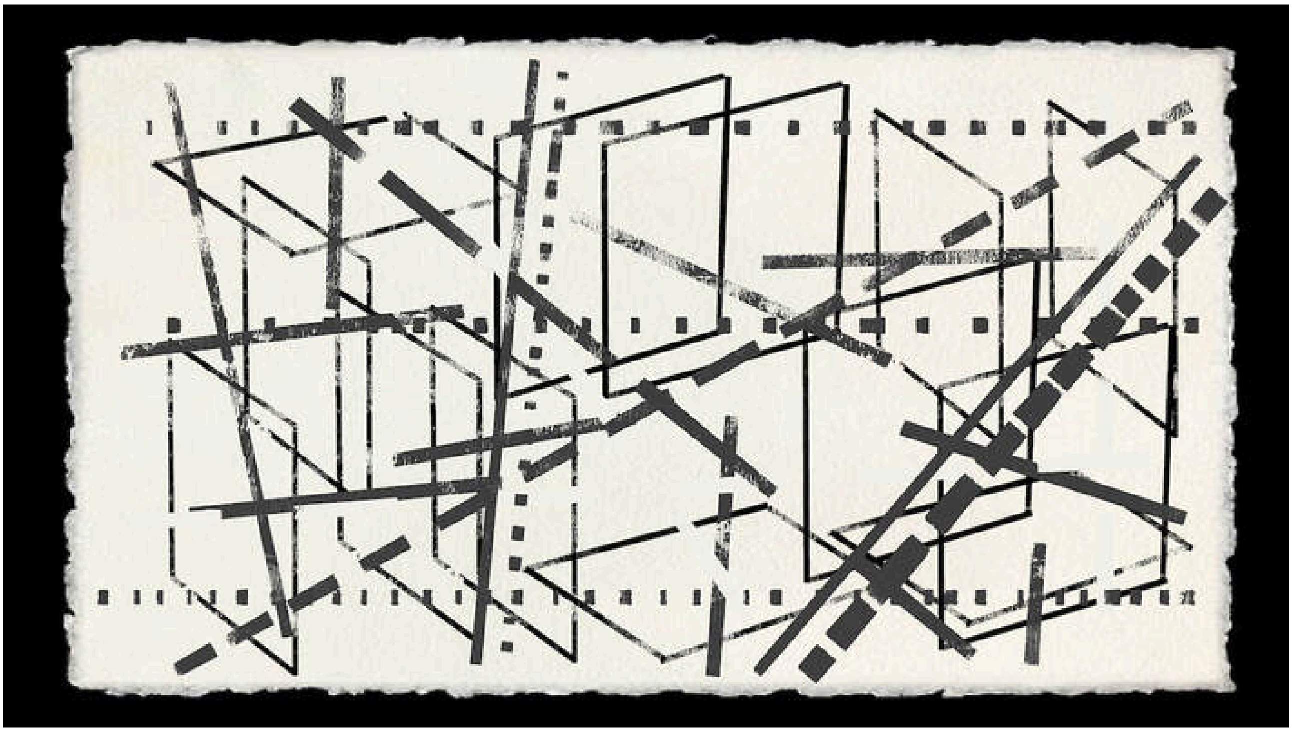
But, wait a minute, you say: doesn't all this implicitly presume the existence of some kind of material objects after all? If not, what are all these processes *made of*? You may have a point here. We are certainly not denying *material reality*. But, at the same time, you'd also be missing the point. On the one hand, we have mentioned earlier that objects can be considered processes that are too *slow* to appreciably change over the kind of time window we are looking at. We have no problem with the practical notion that there are *stable ingredients* to our systems. We are even okay with calling them "*objects*." We just don't think that inert objects are a useful foundation for our thinking about systems. On the other hand, components of a system can be exchanged at a *faster* time scale than the system process itself. Both faster and slower time scales are illustrated by our weather example: the molecules that comprise the air and rain drops in that cold front? They keep coming and going (quite fast) and they also form and fall apart (quite slowly), each at time scales different from that of the front as a phenomenon itself.

Where does that leave us? Well, we now have a reasonable definition of a *natural system*! It is our basic conceptual tool with which we subdivide our ambience into phenomena and events that we perceive or experience during our explorations of a large world. In the words of the great pioneer of systems biology, Paul Weiss: a *natural system* is "the embodiment of the experience that there are *patterned processes*" whose basic characteristic is an "*essential invariance* beyond the much more variant flux and fluctuations of its constituents." There is something more to a system than just the behaviour of its parts. Weiss goes on to point out that this is "exactly the opposite of a machine" where "the structure of the product depends crucially on strictly predefined operations of the parts." In a system, however, "the whole determines the operations of the parts." We wouldn't go quite so far: in our opinion, machines are natural systems too — just very peculiar, simple, and artificially constructed ones. This is why it is such a bad idea to try and explain the whole natural world using mechanistic metaphors.

But this does *not* mean that *mechanistic reductionism* isn't useful as a *method*. In fact, this approach seems to work fine (even if only as a rough approximation) across many situations, or otherwise modern science and engineering would not be as successful as they obviously are. But this is not the problem, which only really arises when the *method* becomes a *view of the world*. Then we start applying the mechanistic approach to systems where it is *not* a good fit — where (in fact) it fails catastrophically. Machine thinking becomes a misleading and dangerous *ideology*, a straightjacket for our thoughts. Subdividing reality into static objects makes it look as if it were much simpler than it really is. The machine view tempts us into hubristic delusions of predictability and control. And we lose our sense of perspective.

And, yet, the *doctrine of containment* is deeply rooted in our thinking, our culture, maybe even our biology. There is a good reason why process-based views remained marginal during the history of our sciences. It is much easier to think in substantivist terms. So we *should* when it's useful and appropriate. Imagine a language without nouns, as Borges does in his story "*Tlön, Uqbar, Orbis Tertius*," where he describes a fictitious planet (called Tlön) with two hemispheres. In the North, nouns are replaced by transient assemblages of monosyllabic adjectives. Instead of "moon," for example, you may say "round airy-light on dark," or "pale-orange-of-the-sky," depending on your situation. In such a language, there is no way to express stable intrinsic properties of an object, as adjectives are readily recombined. The Southern languages of Tlön are even more radical, relying on descriptions of heterogeneous activities only. For instance, "the moon rose above the river" is "upward behind the onstreaming it mooned." Obviously, it'd be difficult for any human being to develop a systematic analytic philosophy in any such language.

The basic point of this digression concerns, once again, the importance of recognising our boundaries: we are both *restricted* and *enabled* by the *conceptual tools* we build for ourselves. We need them to access reality. There is no other way. And it is perfectly fine to use nouns, and even machine-like models of the world, as long as we remember the intrinsic limitations and biases of such practices, and as long as we *don't mistake these maps for the actual territory* we are navigating. Our large world is never quite what we think it is. And it is always changing. We will probably never understand it completely. It is important to keep this in mind when developing *new conceptual tools*, which is what we will now be doing: it is high time to finally translate the natural systems we picked out from the ambience into *formal models*.



Formalisation

When we get to know the large world around us, we somehow absorb parts of the ambience into our selves. This is one side of what it means to achieve *semiotic closure*. And it is the reason why knowledge is neither subjective nor objective, but fundamentally *transjective*, arising through our experiences while exploring a reality beyond our own control. We have argued earlier that *all living creatures*, from bacteria to humans, are *anticipatory systems* that use *internal predictive models* of their world. But science has its characteristic manner of bringing the ambience into our selves. We've also covered this before: scientific knowledge is *naturalistic*, and strives to be *robust*, laying the ground for *coherent* action. But there is much more to it than that. And one core ingredient of the scientific approach we haven't even mentioned yet is *logical inference*. Scientific arguments are committed to follow a *rational* and *consistent* structure.

To understand what this means, we need to get back to the idea of *formalisation*. Remember: to formalise a problem means to pose it explicitly in such a way that we can actually solve it using some form of logical inference. Take the *general problem solving framework* as an example. In this framework, a problem solution consists of a precisely defined sequence of logical operations (an *inferential pathway*) that gets us from the initial problem statement to a valid solution, given a set of constraints on the operations that are allowed. This is precisely what Church and Turing meant by *computation*, and what we mean by an *algorithm* today: a problem is *computable* — and the algorithm *halts* — if we can reach the solution from any valid initial state within a finite number of consecutive logical operations. So far so good.

However, we know by now that the tricky bit is not how to *solve* a problem, but how to *define* it in the first place. This is what *relevance realization* does. It is also known as the *frame problem* in the field of artificial intelligence, or the *symbol grounding problem* in cognitive science. The term "grounding problem" makes the connection to semiotic closure explicit: framing a problem is the same as grounding our symbolic processes of reasoning in the phenomena they refer to in the physical domain of the ambience. This is what adds *semantic* significance to our *syntactic* logical inferences. Without it, inferences are just the rote execution of logical operations, without any purpose or meaning. And it is the grounding problem we have to tackle when we try to connect *natural* and *formal systems* to each other, when linking patterned processes in the ambience to their formal conceptualisations within our selves.

To recap: when exploring our large world, we realize relevance by subdividing the ambience into *natural systems,* which are perceived assemblages of qualities and their relations, and *events*, which are *interactions* between such systems. Thus, perception is the first step in bringing chunks of the ambience inside our self, turning physical phenomena into symbolic descriptions. We generate informal distinctions — *imputed* on the ambience — which we can now *formalise* (*i.e.*, turn into well-defined problems) with the conceptual tools at our disposal. We'll call this second step the *encoding* of a *natural system* (a perceived chunk of the physical ambience) into an *internal predictive model*, a symbolic object we call a *formal system*.

The main aspect of encoding a formal system, and this is crucial to understand, is an *act of abstraction*. Rosen tells us that a natural system consists of a group of *percepts*, while Louie calls it a collection of *qualities*. We now formalise both of these into an abstracted *set of observables*, which we can either monitor qualitatively, or measure quantitatively. As an example of the first, think of describing an object you hold in your hand. Let's say it's a piece of rock. By characterising its colour, composition, texture, and so on, you label it with certain qualities. You may then conclude it's a piece of granite, based on these qualities. As an example of the latter, think of measuring the volume and mass of the rock, or determining its precise chemical and mineral composition: how much of each ingredient is present?

Why do we call this an *act of abstraction*? Remember when we said that designing an experiment is also a kind of modelling? We justified this by pointing out that experiments foreground certain observables, while backgrounding others (by keeping them as constant as possible). Performing a measurement, or simply observing a phenomenon, fall into the exact same category. When we decide what to monitor in the ambience we are necessarily backgrounding a whole lot of other potentially relevant phenomena and events. The number of choices, as mentioned before, is *indefinite*. Going from natural to formal system, therefore, is not a formalisable process, because there is no well-defined starting point or search space. Instead, it *is* the *act of formalisation*, implemented through *abstraction*. To observe, to measure, means to transform *physical phenomena* into specific sequences of *symbolic labels* and *values*, and by focussing on one particular set of observables, we necessarily neglect and exclude others.

As Rosen points out, such an act of abstraction is the very foundation for any kind of scientific inquiry. Unfortunately, it is all too easy for us to forget this basic insight. The irony is that it is, *itself*, routinely backgrounded, when we think *with* (rather than *about*) our models of the world. And this is when we are most vulnerable to the *fallacy of misplaced concreteness*. It is why we're so prone to *mistake the map for the territory*. And even the most committed process thinker cannot avoid drawing boxes. To understand our large world, we have to delimit the phenomena *somehow*. Our criticism is not that we do this, but that we forget the boxes we have drawn, and that it is *us* who have *imposed* those boxes on the world. Let's remind ourselves, again and again, that no natural system in the universe has ever been truly isolated. The boxes we draw around natural systems are not only a form of abstraction, but also an *idealization*. We often deliberately distort and oversimplify our view of nature, excluding aspects from our considerations that we *know* to be important to the phenomena we are studying.

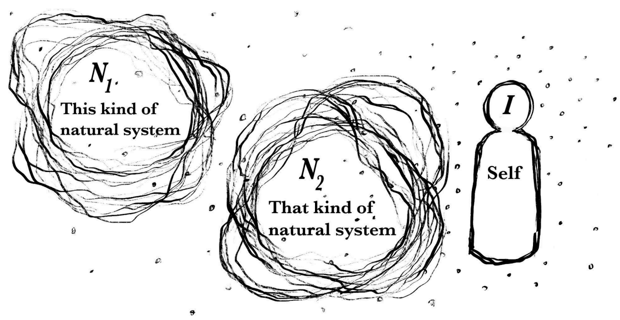
Therefore, our basic assumption should be that everything in the observable universe is interconnected. It is *us* who are doing the subdividing into phenomena and natural systems for our own (utterly human) purposes. And yet, because the world *must* exhibit some degree of orderliness for us to survive and evolve in it in the first place, it is also reasonable for us to assume that we *can* cut nature at its joints.

There is no real paradox here: our subdivisions of the ambience are far from arbitrary. But they are also not perfectly accurate, or universal. Neither are they objective, in the sense of being *independent* of our selves. The aim, after all, is to make sense of the world as it appears and occurs *to us*. The phenomena we pick out as relevant, and the observables we use to formalise them, *always* depend on our *situation* — as individuals, communities, and species. Relevance is *radically context-dependent*. There is no such thing as a universal class of phenomena that are relevant across *all* possible situations. Our needs, motivations, and priorities constantly change, and with them what is significant to us at any given moment.

In addition, our perception is obviously constrained and shaped by what our evolved senses allow us to discover. Moreover, our capacities to measure depend crucially on the current state of our technological evolution and the *measuring devices* it brings forth. Instruments such as rulers, scales, chronometers, thermometers, voltmeters, spectrometers, microscopes, and telescopes greatly increase robustness and precision when recording the *value* of a given observable, and also massively extend the range of observables we are able to detect. Microscopes and telescopes, for example, opened up whole new universes to us — those of the very small and the very large — that were undetectable, and thus unknowable, before the invention of these instruments. Spectrometers, in turn, allow us to measure and analyze regions of the electromagnetic spectrum that are usually hidden from our eyes, while the needle of our compass aligns with magnetic field lines we are utterly oblivious to.

The particular combination of capacities and motivations we have right now is crucially important for how we define a system, for how we subdivide the ambience. This is one main reason why scientific inquiry will never exhaust itself: as long as we continue to change through learning and evolution, we will keep on seeing the world with different eyes. In addition, viewing science as an ever-evolving *modelling activity* leads to another important realisation: any model we can possibly come up with as limited beings will be restricted to some specific domain. A model, however general in scope, only ever captures *certain parts* of the world, *certain aspects* of a natural system, and *certain interactions* the system can have with other systems. There always remains a *semantic residue: other factors* and *aspects* that could become relevant, *other interactions* that may occur, under different circumstances, elsewhere, in the future.

And this brings us to an absolutely central point of our argument: our kind of science no longer strives towards a universal set of laws, a "*theory of everything*." We see no reason to assume that such a theory is achievable. But, more importantly, it is also no longer *desirable*. Instead, what we want is a systematic and rigorous ongoing *classification* of different kinds of systems we come across in the ambience. In Rosen's words: "[t]he growth of science, as a tool for dealing with the ambience, can be seen as a search for special classes of systems into which the ambience may be partitioned." And we've said this before: this search is unlikely to ever end. Reality, basically, is *an infinite zoo of distinct species of natural systems*, as depicted in the following diagram, with *classes* of systems shown as bubbly collections of similar sets.



As we continue to explore the ambience, the number and diversity of systems we encounter steadily *increases*, and even more so do the different ways in which these systems can interact. At the same time, we may be able to *unify* different types of classified systems (and their interactions) into increasingly larger *(super)classes*. Like other open-ended adaptive processes, the evolution of scientific knowledge is driven by ever more complex cycles of integration and diversification.

This sounds reasonable, you may think. Or maybe not? At least it's an interesting way to look at science, we think. But why this digression? What does it all have to do with *formalisation?* Well, a *formal approach* is necessary for any rigorous characterisation and classification of natural systems. In the next chapter, we introduce some powerful mathematical tools from *category theory* that serve exactly this purpose. But for these to apply, we require a *formal definition* of what we mean by a *system*. So let us end this section by making a first (and rather superficial) pass at this, which we develop further in the following chapters.

The most common way to formalise a system is through a mathematical description of its *internal states*. A *state*, in fact, is nothing but a specification of what a system is like at any given moment in time. The idea of a *formal system*, therefore, is traditionally captured by its *set of states*. Let's call this set S, with states s as its elements. A mathematician would write it like this (see also the appendix)

$$S = \{s\}$$

where *curly braces* indicate the immaterial "container" or "bag" for the various s that are the members defining the set. The braces work quite well, visually, to provide a holding space for those little s states.

This abstract set is called the *state space* or *phase space* of the system (depending on whether you are more of a computer scientist or physicist). It precisely circumscribes what the system can do, since it contains all the possible states the system can be in. And what a system can do is what a system *is*. The whole point of a *state description* is to follow changes across time, as the system transitions from state to state, and plots out a *trajectory* (a sequence of states) on its journey through its own abstract state space S. These changes are governed by *transition rules*. In the case of a Newtonian mechanism, these rules are given by its underlying *laws of motion*. In the case of a computer program, they consist of the sequence of *logical operations* that make up the algorithm the program is implementing. We'll come back to that.

However convenient and commonplace this framework may be, there is a serious practical problem with the concept of a *system state*. It is itself an *abstraction* rather than an empirical notion. So, if we're serious about grounding our science in experience, defining systems through *states* is not a great starting point. When we pick out a chunk of ambience as a distinct natural system, we perceive it as a collection of related *qualities*. But these qualities do *not* correspond to system states in any direct way. Remember our example of thunder and lightning? In this case, the system states are made of myriad molecules in the charged areas (on the ground and in the clouds) between which the electrical discharge occurs, and of equally many molecules that create the pressure waves we perceive as the subsequent thunder. Lightning and thunder are *observable phenomena*, yet these molecules are invisible to an unaided observer who *derives* them from the model on which her understanding of meteorological phenomena rests.

Clearly, there is a problem: we seem to require a model of the system *before* we can derive the states of the system. This is obviously impossible. Instead, our initial model must be built around *observables*, not states. But the observables we have available, those we *can* perceive or measure based on our current motivations and capabilities, are often not ideal for our purpose of understanding the behavior of the system under study. There is a basic conflict here, two constraints that counteract each other: on the one hand, we want to define our states such that they are maximally informative about the dynamics of the system; on the other hand, we also want states that can be inferred from the observables we have at hand. It should be evident that these two aims do not naturally align. And for this reason, the connection between observables and states can be quite complicated. Plus: the picture we get of the internal states of a system from our observations or measurements generally remains incomplete.

The central idea here is that observables are not necessarily connected to states in any straightforward way. But at least they are related in a more or less specific manner. To each state s we therefore associate one or more *values*, which we will denote by y, with Y the set of all possible observable values, *i.e.*, an alphabet from which we draw them). This simply captures our very basic presupposition that a system in a particular state will present itself to us in a particular way. It is one of Rosen's two preconditions for doing science. We can now make this vague idea precise by defining a *relation* formally as a *set* that contains *ordered pairs* of states associated with observable values. Or, as a mathematician would write

$$\{\langle s, y \rangle\} \subset S \times Y$$

which looks complicated but basically means "a bag full of ordered state-value pairs (indicated by *angle brackets*) that are drawn from (or a *subset* of) a much bigger bag of all *possible* combinations between states and values (the $S \times Y$ part of the formula). Put simply: only *some* (but not all) observable values are associated with specific states, so we can (sort of) guess what the latter are from the former.

Now, we can make our lives a lot easier by assuming (for the sake of argument) that our measurements or observations are *perfectly accurate*. Yes, we know this is another *idealisation*, but it's okay for now, or at least as long as we remember that we've made it. With perfect accuracy, there is exactly *one* value of the observable associated with each state, and our relation turns into a *mapping*. Mathematicians also call this a function, which is confusing because of the completely different way in which biologists use the term. That's why we'll only use "*function*" in its biological sense in what follows. In mathematical notation, a *mapping* is much easier to parse than a general relation. We write

$$s \mapsto y$$

with this particular kind of arrow reading as "*s* maps to *y*". Note that it is still possible that multiple states map to the *same* value *y*, which can complicate matters (see below), but at least, there is only *one* specific *y* that belongs to each specific state *s* now. We can then take another leap of abstraction and look at this mapping at the level of whole sets, instead of the individual elements. Thus, we write

$$S \to Y$$

which indicates that the mapping takes us from its *domain*, which is the *set of states*, into its *range*, the *alphabet* of *values* for our observable. Again, this reflects that states express themselves as observables.

This captures the *process of measurement* quite aptly. As an example, think of an old-fashioned mercury thermometer, which is itself a natural system (according to our definition). The height of its column is determined by how much the volume of mercury expands due to the kinetic energy of its component molecules. We can now read this off as a specific temperature value from the scale of the thermometer. This is how the device makes the state of its mercury observable to us: mapping it to a particular number. We call natural systems that do this *meters*, and each observable is measured by a specific meter.

We hope the mathematical formulas are not putting you off? They are (with the possible exception of the one for the relation) deceptively simple and (we hope) illustrative of the concepts we are trying to introduce. But a simple arrow can hide a lot of complexity. Luckily, there is a lot of solid mathematical theory about mappings that helps us sort out such complications. It tells us, for example, that we can only get a *complete* and *fully resolved* picture of the system states, if our mappings are *invertible*, *i.e.*, if we can go *backwards* from observables to states. The problem is: *invertible mappings* are exceedingly rare among all possible mathematical relations. In particular, they require a strictly one-to-one association between states and their observable values. This means that either there must be a special reason why most associations between observables and states are of this particular kind or that we have to get incredibly lucky to be able to infer the states of a system accurately and completely with the observables that we have available. Since, in fact, we *don't* have any reason to assume a special relationship, we conclude that most of the time we just won't get the full picture, but only *piecewise approximations* to reality.

Actually, this is one of the reasons why it is so difficult to obtain robust and exhaustive knowledge of the world just by looking at it. Lucky for us, we *can* improve on that: systems not only communicate with their environment through their observables, whose values we can think of as the system's *output*. Obviously, the state of a system also responds to *inputs* from its environment. Let's call such inputs *x* (and the *alphabet* from which we draw them *X*). Think of a computer program, as an example: any such program that does anything interesting depends on some kind of interaction with a user, or on input data read from storage (or some sensor). Clearly, the algorithm will perform differently on different inputs. The same applies to a Newtonian mechanism: its workings can drastically differ based on what we called its *initial* and *boundary conditions* — representing inputs that come from *outside* the system.

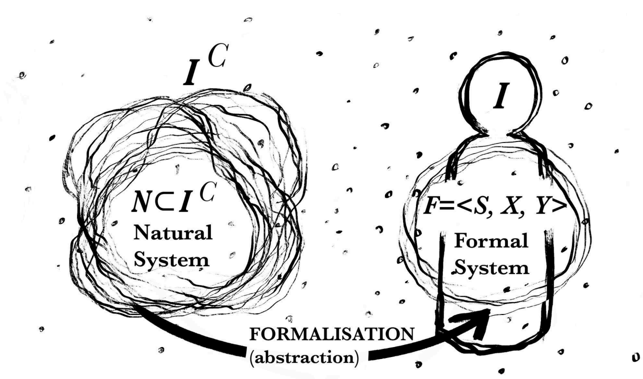
This suggests, of course, that we can not only observe, but also *manipulate* the inputs (as far as physically possible) to alter the outputs. *Both* of these are *observable* and often *measurable properties* of a system. In this way, we can get the system to explore a much larger part of its set of states — its possible kinds of behavior — than if we are just passively observing it. This is, in a nutshell, a general model of how we do *empirical research*. Evidently, it is a much more robust way to infer the states of the system than the mere observation of regular associations in its output. And it reveals how the logical succession of states in a system is supposed to model *causal chains of events* in the ambience. But we're getting ahead of ourselves. For now, let's focus back on what all this means for our basic definition of a formal system.

In fact, it immediately suggests a *minimal model*, which we can now define as a *relation* or *mapping* that takes an input, in addition to its current system state, to generate some output. We write

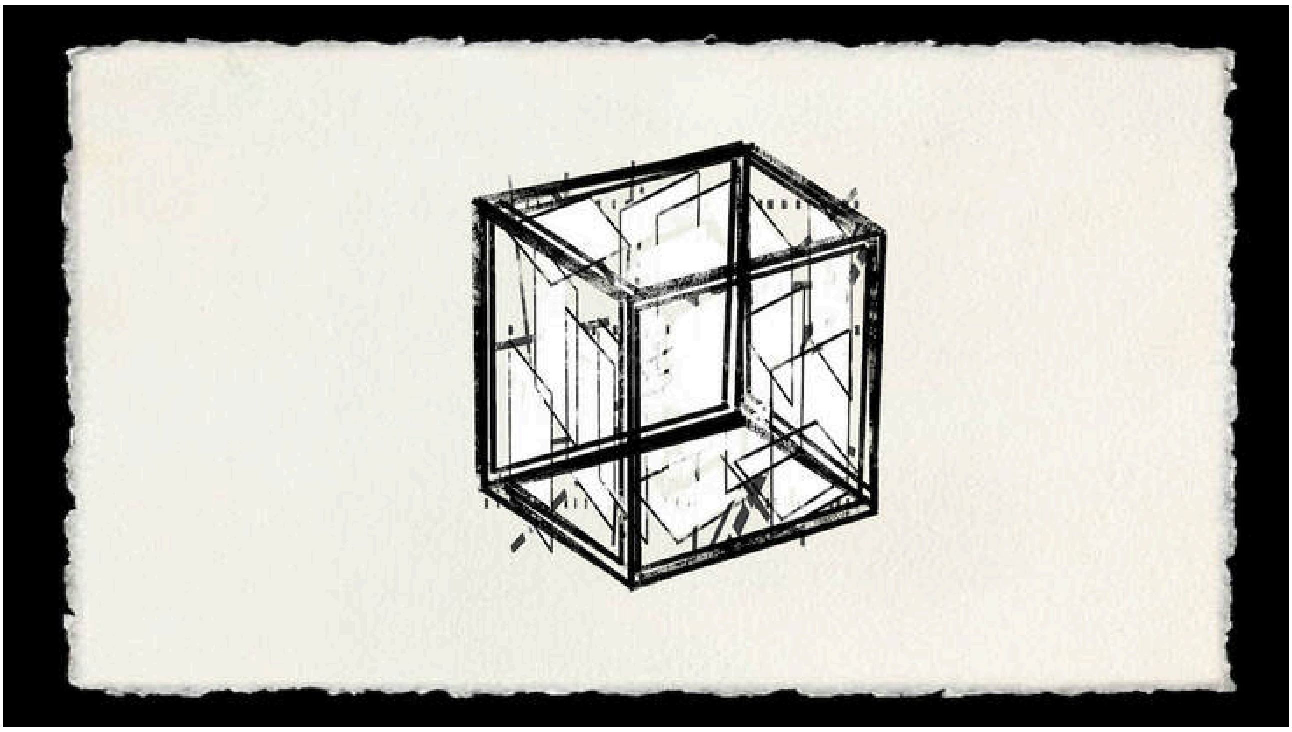
$$X \times S \rightarrow Y$$

meaning that we take an ordered pair of input x and state s, *i.e.* $\langle x,s \rangle$, which the system then maps to a specific output y. This kind of *formal system*, which lies at the heart of the *general systems theory* formulated by Mihajlo Mesarovic and Yasuhiko Takehara in the late 1960s and early 1970s, is still fundamentally based on its set of states S. But, on top of this, this set now has additional *mathematical structure* — in the form of well-defined *relations* to both its inputs X and outputs Y — which can be directly linked to our empirical research practice of performing *interventions* and *measurements*.

We end up with the following general picture, which sum up our argument in this section:



Looks neat! But again, we must emphasise that the description we get from such a formalised approach is *always incomplete* (by its very design) — as we hope we have made abundantly clear. Ironically, our attempt to build a formal model of what it means to do research gives us the insight that a formal model, in general, just gives us one (of potentially many) valid accounts of what a system can do. This self-undermining quality of the approach is a feature, not a bug! It reveals that the relation between natural and formal systems is quite nuanced and complicated, and we have to be careful when using such formal models as tools (or maps) to understand our large world (or territory).



Congruence

Now we come to a really central question: what do we mean when we say a formal model is *congruent* with a natural system? Our aim as skillful modellers, as we've said multiple times before, is to learn to find our way around the large world we live in. We use models to gain a robust understanding of the causes of phenomena and events in the ambience. This understanding, in turn, allows us to make predictions of what is going to happen next, or to recognise when such predictions are hard to come by. The practical purpose of science is to support our decisions with solid evidence, and to enable a coherent course of action towards our goals. On the longer run (and in the best case), a better grip on reality allows us to act in a sustainable manner — to be truly and comfortably at home in our world.

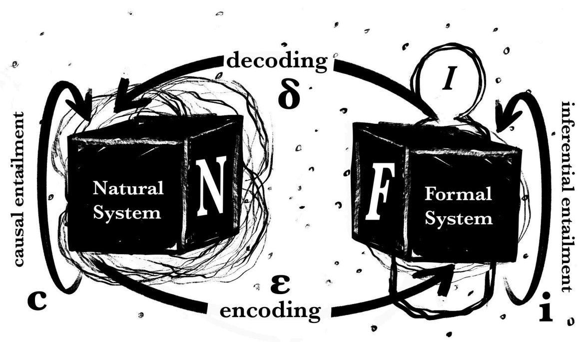
Considering the complexities of navigating our large world, it would be naïve to assume that formal systems are straightforward one-to-one *representations*, or (*mirror-*)*images* of the natural systems they model, that their main aim simply is to *reconstruct* or *reproduce the ambience within the self* — as accurately and exhaustively as possible. This is really *not* their point. We said it before (and so did Borges): the use of a map does not lie in the amount of indiscriminate detail it captures about the territory.

Remember: we are no longer after a "theory of everything." Instead, we judge the quality of a model by how well it works for us — right here, right now. Does it bring out features of the world that are *relevant* to our current task? Does it do so robustly, and reproducibly, across a wide range of different circumstances? Does it improve the coherence of our actions? Does it increase our chances of solving our problems? Scientific models are tools, primarily, before they are representations. And this is why it is a feature, not a bug, that they only ever give us a partial picture of reality.

As philosopher Angela Potochnik points out, idealised models *deliberately* misrepresent certain aspects of reality, and *willingly* ignore some others we know to matter in principle. They do this *in order* to be *useful* in other regards. Sometimes, it's better to take a pretty cartoonish approach, lest we lose the forest for the trees. As limited human beings, we constantly have to struggle with the tension between the complexity of the world, and our need for relatively simple explanations that we can actually understand. Models are always used by *someone* for some specific *purpose*. In particular, *our* models must serve *our* purposes. Just like in the case of symbols, there is a triad: model, modelled, *and* modeller.

But then, is there anything general we can say about the relationship between natural and formal systems? Or is modelling just one damn thing after another? A fresh start required in each situation?

Lucky for us, there *are* few things that most models have in common. For one, as we've mentioned above, models generally attempt to capture certain *causal features* of the ambience through the use of *logical inference*. There is a deep connection there. Rosen calls this the *modelling relation*. It is one of his most central concepts, and he visualises it as follows:



We'll discuss in detail what all these symbols and arrows mean. Easy to recognise are *natural system* N and *formal system* F (depicted as black boxes, without showing any of their internal workings). And we've already encountered the *encoding arrow* ε in the last section. It stands for the act of *abstraction* that gets us from N to F, leading us to translate putative *causal interactions* in the ambience into steps of *logical inference* in a formal system. This basic parallelism is what Rosen captures with his concept of *entailment*. To be *entailed* means to be "an inevitable consequence of something." In the ambience, it is *events* that *causally entail* each other (marked by self-referencing arrow c of N in the diagram); in a formal model, it is *states* that *inferentially entail* each other, *i.e.*, they follow logically from each other (marked by self-referencing arrow i in F). Both go hand in hand.

But what do we mean when we say a model "captures" a natural phenomenon? What does it mean in this context, as Michela Massimi puts it, to *get things right*? Simple *correspondence* doesn't really work, as we've explained above. Instead, Rosen says that natural and formal systems have to be *congruent* somehow, but *congruence* is a word with many meanings. In its colloquial form, it means "agreement," "compatibility," or "harmony," which are all rather vague. Even in mathematics, its technical definitions are highly divergent between different fields. In geometry, congruent figures share the same size and shape. But this is not really helping us either. In linear algebra, two matrices are congruent when they represent the same bilinear form with respect to different bases, and in topology, different manifolds can be congruent. None of this, however, pertains in any meaningful way to the modelling relation.

Rosen's own explanation doesn't quite work either, as we will see. But it highlights an important point: *congruence* very much hinges, not only on how we *encode* a formal model, but also (and this is crucial) on how we *decode* it back into the natural system in the ambience (as indicated by *arrow* δ in the diagram above). Rosen's criterion for a good model is, in fact, twofold. First, you must arrive at the *same outcome* whether you follow arrow c directly (*causal entailment* in the *natural system*), or if you follow first ε, then i, and then δ, *i.e.*, if you *infer* and *predict* the behavior of the natural system through your formal model.

But this is not enough. In many situations, you can get a predictive match through an arrow i that has nothing to do with arrow c. This is (in)famous among scientists as the *problem of model (over)fitting*: it was John von Neumann (of computer architecture fame) who said that "with four parameters, I can fit an elephant, and with five I can make him wiggle his trunk." Give a model enough *degrees of freedom*, and it will fit *anything* at all. Or it fits the noise in your data, instead of the system dynamics you actually want to understand. Philosophers deal with a related kind of problem, which they call *underdetermination*: usually, as we have seen in our discussion of observables above, the evidence we actually have does not uniquely or completely determine what the states of the system *have to* be. In general, there are many ways to get the same result, especially if our situation requires a model with lots of degrees of freedom.

Rosen calls a formal system that does *not* get the causal structure of its natural counterpart right a *simulation*. On the one hand, this is confusing because, nowadays, we use the word "simulation" for a computational process that *does* represent a physical one. On the other hand, we *do* find Rosen's original term useful, and in line with how Jean Baudrillard defined it in his "*Simulacra and Simulation*," the philosophy book behind the original "*Matrix*" movie. Baudrillard's "simulation" emphasises that we are dealing with a mere *imitation* of the operation of a natural system. This is what we called (*algorithmic*) *mimicry* earlier on. It may quack like a duck, it may walk like a duck, but it is emphatically *not* a duck. Not only is it *not* the real thing, but it does not even tell us how the real thing (or process, rather) works.

Which brings us to an important point: to be able to predict something reliably in a given situation does *not* mean you understand it! In fact, it is a common mistake to think that the main point of modelling in science is to *predict*, when it really is to *understand*. A *simulation* (in Rosen's sense) predicts *without* understanding. Think about the kind of large-language models (LLMs) that are popular in machine learning these days. They are a perfect example of a Rosennean simulation in that they predict patterns in human language use (to an astonishing degree) without either the algorithm or the user/programmer having any kind of understanding of what is going on. A rather unfortunate consequence of this is that LLMs work as long as they do, but sometimes they don't (they "*hallucinate*") without anyone having a clue as to why and when they do that. This, by the way, is a fundamental design feature of Rosennean simulations (and hence LLMs), not a bug that can eventually be smoothed out or eliminated.

If the point of modelling in science is *understanding*, then we need to do better than simulation! And this leads us to Rosen's second criterion: we need to encode the arrow *i* itself such that it allows us to infer something useful about arrow *c*, that is, the *flow of causation* in the natural system being modelled. More specifically, *i* must be encoded in a way that allows us to *decode* it back into arrow *c*. Only then, in Rosen's view, can the resulting formal system be counted as a proper *model*.

But Rosen slips, at this point, and delivers a pretty shallow and representationalist explanation of what he means by *congruence*, invoking some kind of encoding and decoding *dictionaries* to translate back and forth between natural and formal systems. The problem is, as should be evident from our previous discussions, that modelling is not just some straightforward translation from the physical to the symbolic (and back). We don't really have any reliable or general method or recipe that enables us to do that. In fact, we have seen that the *art of modelling* is not an entirely formalisable process at all.

What, then, is a better way to interpret *congruence* between natural and formal systems? What, exactly, *is* this *modelling relation*? It seems hard to pin down without drifting back into representationalism. Louie, in his book "*Intangible Life*," gives us a few intriguing hints.

The trick is to look at *congruence* explicitly in terms of the two *adjoint relations* of *encoding* and *decoding* a model. Look at the diagram above again to see how they are complementary to each other. Now think about *how* they ought to complement each other to be *useful*. Mathematically speaking, these arrows are *functors* which connect two different *categories*. We'll explain what that means in the next chapter. For now, let's just say that they can be interpreted in three increasingly relaxed ways, that take us away from strict representationalism towards a much more flexible and pragmatic view of models as useful tools.

The first and narrowest interpretation is in terms of the *invertibility* of the encoding and decoding arrows. This makes intuitive sense. Ideally, you'd like to go through arrow ε then back through δ and always end up at the same place (no matter where you started from). In this perfect case, we say that the natural and the formal system are *isomorphic* (or *equal* "*up to isomorphism*"). Note that being isomorphic isn't the same as being *exactly* equal, in *every* respect. Instead, it just means that there is a two-way one-to-one correspondence of the components and interactions between the systems. A model is never exactly the same as the system it models. Rather, the isomorphism captures what Rosen means when he talks about encoding and decoding *dictionaries* for the modelling relation: there is a comprehensive and detailed translation possible between the internal workings of the natural and the formal system. In some sense, we can say that the model captures *completely* what the natural system is.

But here we run into the same problem we encountered with states and observables before. First of all, invertible mathematical operations are very unlikely to apply in a situation where our picture of the natural system is *incomplete*. We cannot have a true one-to-one correspondence if individual details are missing. On top of that, encoding (by definition) is an act of *abstraction*, which necessarily *loses information* about the natural system. An abstraction is *never* entirely reversible, since we are explicitly and deliberately *forgetting* aspects of the system we deem to be irrelevant or redundant. This is exactly the reason why naïve representationalism fails: as in Borges' story, we'd need a map that is just as large and detailed as the territory for it to work. And this is neither practical nor desirable. It is precisely the point of a model to pick out only those aspects of a system that are *relevant* to the situation at hand.

This is where a second interpretation comes in handy. Mathematicians have a number of concepts that relax the strict requirements of complete *equality*. One of these is *equivalence*. It explicitly captures the idea that equality can be restricted to specific properties, and we can consider objects and systems as equivalent that are merely *similar* (instead of equal) with regard to these properties. Think of a number of people in a room, which you want to sort into age groups. In this case, it obviously doesn't matter how tall they are, how much they weigh, where they come from, what colour their eyes or skin. And you're not even looking for people who are *exactly* the same age, but only for those who fall into a specific range. This will include very different people as long as they are in the same age group. Such a group is what mathematicians call an *equivalence class* (which is really just another kind of set).

It may seem like we are belabouring a rather trivial point here, but it's crucial! Applied to the modelling relation, it greatly simplifies our task: each component of the formal model now simply needs to *do* something that is in some relevant way *equivalent* to what its natural counterpart does. It does *not* have to *be* the same in all its glorious details. In fact, it can be of a totally different nature. And we no longer lose any information through abstraction (as long as the aspects we've abstracted away really aren't relevant to our particular situation). The model and the natural system just have to be "similar enough" for our purposes. They must behave (and respond to perturbations) in *equivalent* ways, within some delimited range of conditions. Natural and formal systems can be radically different. What's important is only that we can understand the workings of the former from examining the latter.

We think that we have definitely taken a step in the right direction — in several ways, actually. By considering a formal model *equivalent* (rather than *isomorphic*), we openly acknowledge that we only ever have a partial picture of all the possible aspects and interactions of a natural system. And this is a good thing, since we *want* to focus on only those aspects of the system that matter in our given situation. Plus, we now have more specific criteria for evaluating *congruence* in this situation: the components of our formal system behave in a way that matches the behavior of their natural equivalents in some observable and relevant way. The better and more robust the match, the better and more solid our understanding and predictions derived from the model.

This is all great. But, unfortunately, there is still a lack of clarity about what it means to correctly *reproduce* an observed natural behaviour in a formal system. And we are still stuck with a view of modelling that treats it suspiciously as a passive mirroring of a given reality, rather than an (inter)active engagement and transjective "bringing forth." Therefore, one more step is required, and this step is reflected in a third mathematical concept, a further relaxation of the idea of equivalence. This is called *adjunction*. Remember: in the beginning of this section we talked about encoding and decoding as being adjoint arrows, and this is no accident. *Adjunction* is a quite technical and advanced concept, mathematically speaking. But it is possible to convey the gist of it here, and to interpret it in a way that perfectly fits our perspectival and pragmatist approach to models as tools for scientific research.

Considered philosophically, two adjoint arrows connecting a formal model and its natural system capture the situation where we don't have any straightforward *correspondence* between the two, but where we can still *draw inferences* from one to the other in a systematic and rigorous way. So, even if natural and formal systems are different, and do not even partially mirror each other's behaviour, they can still be highly informative about each other. In fact, adjunctions are *asymmetric* connections, which in our case reflects the obvious fact that the inferences we use when encoding the formal model from the natural system are not the same as the inferences we draw from the formal model when trying to understand or predict the behaviour of the natural system. Put simply: arrows ε and δ are no longer directly invertible, but we still know how to end up at the same spot when going back and forth along them, even if it is through a more indirect path — an inferential detour, if you want.

The mathematical trick at the heart of this is that there must be some kind of regularity in the translation process between natural and formal systems, but it does not have to be any kind of obvious one-to-one correspondence or representation. Mathematicians have a name for this: they say that the required *transformation* between the back-and-forth arrows needs to be *natural*. This kind of *naturality* is just a mathematically rigorous way to state the fact that, as long as the relation between encoding and decoding is *regular*, it can be as complicated and indirect as we want. This allows for a much greater flexibility when considering how models should be *congruent* to the natural systems they relate to. In particular, it allows us to drop the idea that this relation must be some kind of direct *representation*.

This view is closely paralleled in Michela Massimi's account of "*Perspectival Realism*." Massimi, like us, sees science fundamentally as some kind of skillful modelling activity, and suggests that models play the role of *inferential blueprints*: they enable researchers to draw robust and accurate inferences about the natural phenomena they are studying. Like traditional architectural blueprints, models are collaborative tools that offer various perspectives on a natural system — a useful set of maps to find our way around an area of interest. These maps are almost always idealised, one way or another, depending on the task to which they are being put. That is why we do not usually judge their quality by how well they represent what is actually happening, in detail, at this moment, but rather by how well they enable us to understand and predict certain relevant behaviours of the system under study, now and in the future.

And this is exactly what is captured by the mathematical concept of adjunction: rather than literal equality (or, at least, equivalence) between formal model and natural system, what counts is that the two are connected (however indirectly) in a *useful* and *reproducible* manner, which allows us to continue building robust scientific knowledge through our explorations of the world. This is what we take *congruence* to mean: a formal model that gives us *a better grip on reality* by generating solid conjectures and predictions, which we can then use to come up with further experiments and models. And so on. We've said it several times before: this adaptive process never ends. But it also never ceases to work for us. The aim is not to gain a complete understanding of everything, whatever that may mean. The aim is to build the best maps possible. So let's examine next what kind of tools we have available to achieve this.