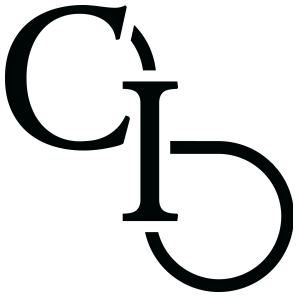
Constructor Theory

Module 2

Irreversibility

Logan Chipkin



CONJECTURE UNIVERSITY

Irreversibility

As we saw in Module One, all dynamical laws of motion are *time-reversal symmetric*: if the trajectory of a system from state A to state B is allowed by particular laws of motion, then the trajectory of that system from state B to state A is *also* allowed by those same laws.

Yet this fundamental reversibility sits in stark tension with the one-way processes we observe all around us. A glass effortlessly falls to the floor and shatters into hundreds of pieces, yet hundreds of shards of glass never spontaneously recombine into a pristine glass. Some stars die in violent explosions called supernovae, yet their spilt guts never spontaneously reassemble to reform the well-ordered fiery sphere from which they erupted. Living things undergo developmental lifecycles with a definite direction—mammals, for instance, always develop from a single cell into a multicellular bag of organs, and older elephants and lions never develop back into embryos. Engines become less efficient as they accumulate damage, but worn machines never spontaneously return to the efficiency of those fresh off the assembly line.

All of those processes, as well as their reversals, are permitted by the same reversible underlying laws. Yet in practice we only ever observe one direction.

The conflict between reversible laws of motion and high-level irreversibility had been noticed as early as 1870, by which time the laws of thermodynamics—those that explain the regularities of work, heat, temperature, and efficiency—had ripened following a few decades of investigation. Thermodynamics had placed the commonsense view that some processes were irreversible in stark relief: for example, one version of the second law of thermodynamics held that mechanical energy could be entirely converted into heat energy, but that the inverse was impossible. Since 1870, a number of attempted solutions to this conflict have been offered that can't quite solve the problem. But, as is often the case in science, understanding the reasons *why* they fail shed light on what attributes the *actual* solution must have.

Statistical Mechanics

The statistical mechanics formulation of thermodynamics states that, although the underlying laws of motion are symmetrical, some trajectories of the high-level physical system in question are simply more probable than others. For example, although the laws of motion governing the air molecules in a room allow for them all to end up in one corner, that is a far less probable arrangement than, say, a relatively homogeneous distribution of the air molecules across the room. Similarly, although the laws of motion underlying the molecules in a warm glass of water with ice cubes allow for the molecules to rearrange themselves into a cold, cylindrical block of ice, it is far more probable that the water molecules collectively reach equilibrium temperature with their environment.

But this probabilistic picture sits uneasily with something we established in Module 1: the laws of motion are deterministic. Once the state of an isolated system is fixed at a given time, its entire future and past are fixed as well. There are not multiple possible futures, each assigned some probability—there is only the one trajectory the system follows. Its trajectory is not described by a set of possible trajectories, each weighted by some probability. There is only the one trajectory that the (isolated) system must follow. Yet statistical mechanics says that, although each air molecule in a room does follow deterministic laws of motion, collectively they act in probabilistic fashion—i.e. they are *probably* going to spread themselves evenly across the room and are very *unlikely* to congregate into one corner. So statistical mechanics has replaced the conflict between reversible underlying dynamics and high-level irreversibility with the conflict between deterministic underlying dynamics with high-level probability. It is not obvious that this shift in conflict is an improvement.

Moreover, this statistical mechanical view cannot even be a fundamental law of Nature at all but rather reduces the second law of thermodynamics to an approximative tendency, not an exact law on par with those of, say, general relativity or quantum mechanics. After all, the air molecules will only 'tend to' equilibrate across the room, ice cubes in warm glasses of water will only 'tend to' melt and match the temperature of the surrounding water. But there is still a small probability that the air molecules *won't* equilibrate, that the ice cubes *won't* melt away. So the very irreversibility that had seemed like a real phenomenon has been waved away as an illusion, leaving thermodynamics as something that only applies to some systems (we don't know exactly which) under some

circumstances (we don't know exactly which). But that only leaves us with the problem of explaining when and why the 'apparent' irreversibility manifests, leaving us exactly where we started.

Coarse-Graining

Another proposed solution to the conflict between reversible dynamics and emergent irreversibility is *coarse-graining*. In the context of thermodynamics, this entails ignoring some microscopic details of the system under scrutiny. For example, it could be that your ability to measure the state of every water molecule in a glass of water is extremely limited (due to, say, technological constraints). You therefore model the water not as a collection of particles, each with its own deterministic trajectory, but as a collection of coarse-grained subsystems: the ice cubes, the water, the glass, and the environment. The evolution of *this* system's properties—such as temperature, state of matter, and whether or not it is in equilibrium—may indeed be irreversible, and you may have even deduced rigorous mathematics that describe such a system.

But this approach introduces arbitrariness. *Which* microscopic variables should be ignored? *How much* detail must be suppressed to recover irreversibility?

Then there is the fact that, under this approach, one person's reversibility is another's irreversibility. For example, one scientist may coarse-grain only very slightly such that he 'confirms' that the evolution of the particles comprising the glass of water is, indeed, reversible, and that the particles will return to their initial state over long enough timescales. His colleague, meanwhile, may coarse-grain enough so that he concludes that the ice cubes in the water will melt and never reappear. Apparently, coarse-graining itself leads to subjectivity with respect to irreversibility—which again, would imply that thermodynamics is not fundamental.

To give a more informal example: from a very coarse-grained vantage point, the position of an elevator seems entirely reversible, as it can go back and forth between any two floors. However, the technology governing an elevator's motion is not infinitely precise, and so its height from the Earth's surface when it reaches, say, the sixth floor will not in general be consistent. In fact, at a fine-grained enough level, it will reach a *different* height from the Earth's surface far more often than not—so much more often that the

evolution of the elevator would fairly be called irreversible. Once again, whether or not the elevator's trajectory is reversible is a matter of the level of detail we choose to focus on, not an objective property of the elevator nor of the laws of physics.

Axiomatic Thermodynamics

The approaches considered so far—statistical mechanics and coarse-graining—both try to derive irreversibility from properties of trajectories. A different strategy starts not with trajectories but with transformations. The so-called axiomatic approach to thermodynamics is grounded in the idea of an *adiabatically enclosed system*, a physical system that *cannot* exchange heat with its environment but *can* exchange energy through mechanical work. A fire piston, a bicycle pump, and a thermos flask are all examples of approximate adiabatically enclosed systems, as they are all effective at inhibiting heat flow while still allowing the transfer of energy through work.

Consider an idealized thermos flask filled with water insulated so that no heat can flow between it and the outside world. Let the state of the water at initial temperature T_0 be X. The temperature of the water will not rise or fall to equilibrate with the environment's temperature, as would happen if the flask allowed for heat transference. If we wanted to raise the water's temperature to T_1, we could attach a stirring machine to it that repeats the same stir over and over in a cycle, thereby transferring mechanical energy from the machine to the water. Let the state of the water at this new temperature be Y. We say that the water's state Y is *adiabatically accessible* from the water's state X. In layman's terms, we can heat up the water by performing work on it, thereby transferring energy to it, but we can't cool it down by performing the same method.

The notion of adiabatic accessibility generalizes: state Y is adiabatically accessible from state X if a device can bring about the transformation from state X to state Y with the only side-effect on the environment being the (equivalent of) displacement of a weight in a gravitational field. The device in our example was the stirrer, the side-effect is the corresponding movements of the machine's parts for any given stir, and state Y and state X are the water at temperature T_1 and T_0, respectively.

Now consider transforming the water from state *Y* to state *X*—that is, lowering the temperature. This is *not* possible by the same method we'd employed when raising the temperature—that is, state *X* is adiabatically *in*accessible from state *Y*.

We have found a transformation that is possible in one direction but impossible in another—an irreversible process. The axiomatic second law states that for any physical system there exist states *X* and *Y* such that *Y* is adiabatically accessible from *X*, but *X* is not adiabatically accessible from *Y*.

More informally, we can say that mechanical energy can be entirely converted into heat energy (without side-effects), but the inverse is not true.

Unlike the statistical mechanical formulation, this axiomatic approach does not reduce thermodynamics to something less than a fundamental theory, nor does it introduce a conflict between deterministic laws of motion and probabilistic high-level 'tendencies'. And unlike the coarse-graining approach, the axiomatic formulation does not turn thermodynamics into a subjective game that makes irreversibility itself in the eyes of the beholder.

The statistical mechanical and coarse-graining approaches try to explain irreversibility in terms of trajectories (see Module 1): the former says that some trajectories are more probable than others, while the latter says that if you neglect certain (arbitrary) details, then some trajectories are ruled out. Interestingly, the axiomatic approach is not about trajectories at all, but is instead about the possibility and impossibility of transformations.

While the axiomatic formulation does not suffer the aforementioned issues of the other two proposals, it cannot be the ultimate solution to the conflict between reversible laws of motion and irreversible emergent phenomena for the simple reason that it is *scale-dependent*.

It works beautifully for everyday systems—such as containers of liquid, steam engines (indeed, thermodynamics was born as an effort to better explain the physics of such machines), and pulley systems. Its core concepts—displaced weights in gravitational fields, adiabatic accessibility, mechanical work—are drawn from that macroscopic world. But these notions do not obviously apply to all physical systems. For example, what are the differences between work transfers and heat transfers in the context of a massless quantum particle, such as a photon? What is the equivalent of a displaced weight for such

a system? What about work and heat transformations involved in performing nanoscale *computations*—that is, how do work transfers and heat transfers interface with informational transformations? A fully general, scale-free version of the axiomatic second law of thermodynamics would express the relationship between work and heat not just for macroscopic systems, but for systems arbitrarily alien to those of our everyday experience.

Three Approaches to Irreversibility in Thermodynamics

	Statistical Mechanics	**Coarse-Graining**	**Axiomatic Approach**
Approach	Assign probabilities to trajectories such that some are more probable than others	Ignore some microscopic details such that some macroscopic processes are subsequently irreversible	Express irreversibility in terms of asymmetrical transformations—work can be entirely converted into heat but not vice versa
Shortcomings	Reduces thermodynamics to something less than fundamental; introduces new conflict between deterministic laws of motion and probabilistic 'tendencies'	Makes irreversibility subjective; depends on arbitrary choices about which details to neglect	Scale-dependent such that it does not generalize to all possible physical systems (i.e. massless quantum particles, nanoscale computation)

Attributes of the Solution

What sort of theory could allow for a generalization of the axiomatic approach that would, at long last, give us a scale-free way to reconcile reversible laws of motion and irreversible emergent phenomena?

It could not be tied down to the formalism of any particular dynamical theory, so its own formalism would have to transcend that of quantum mechanics and general relativity alike.

Second, it could not be expressed in terms of trajectories. It would instead have to be expressible in terms which transformations are physically possible and which are impossible.

Finally, the theory would have to be able to handle *substrate-independent* quantities such as work and heat, which retain their defining attributes even as they are transferred between arbitrarily distinct physical systems.

Scale-dependence not only blocks axiomatic thermodynamics from fully resolving the irreversibility problem, but it plagues several thermodynamical laws as well. In the next module, we will see how even our best formulation of the laws of thermodynamics fails at certain scales.

Thanks to Conjecture Institute Cofounder David Kedmey, Dirk Meulenbelt, and Edwin de Wit for valuable feedback.