Entropy and Experience

Adam Elga

February 12, 2002

1 Fred and Derf

In our world: Fred happily contemplates his tuna melt. His photoreceptors absorb light reflected from the melt. He takes a big bite. Tuna juice drips down his chin. He thinks, “Delicious!”

In the temporal inverse of our world: Derf (the temporally inverted counterpart of Fred) sits in front of a tuna melt with a bite-shaped piece missing. His neurons engage in the reverse of the activity that underlies Fred’s thought “Delicious!” Tuna juice climbs up Derf’s chin. His arms bring the melt to his mouth, which unchews a bite and seals it onto the melt. Meanwhile, his photoreceptors shoot light in the direction of the melt.[1]

Fred has a great time eating his tuna melt. Does Derf enjoy assembling his? How does it feel when Derf’s eyes shoot light at the melt?

I claim: Derf doesn’t enjoy assembling his melt. And it doesn’t feel like anything when light shoots out of his eyes. Derf has no experiences

---

[1] I owe the image of a temporally-reversed person shooting light from his eyes to Tim Maudlin.

whatsoever. And not just Derf. None of Derf's fellow temporally inverted creatures have any experiences, either.

And not just them. Take any world with fundamental dynamical laws rather like ours, but in which entropy doesn't increase. Take any system in that world that changes state by changing thermodynamically. That system doesn't have any experiences, either.

In slogan form:

> Conscious experience requires entropic increase.

That's my main claim. But I'll defend a warm-up claim first:

> Temperature regulation requires entropy increase.

It makes sense to start with the warm-up claim because creatures are more complicated than thermostats.

2 Temperature regulation

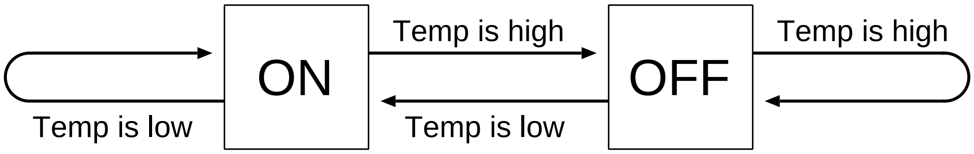
2.1 What it takes to regulate the temperature

There's a thermostat in my oven that regulates the oven temperature. My oven is now set to "Bake 375". While it is so set, the thermostat ensures that the oven's heating element is on whenever the oven temperature is low (less than 375 degrees), and off otherwise.

My oven's thermostat is now regulating the oven temperature. In order to do that, it satisfies (at least) these three conditions:

1. It has the right sorts of states. Idealizing somewhat, the thermostat has two states: ON and OFF. When the thermostat is in the ON state, the heating element is on. When it's in the OFF state, the heating element is off.

2. It has the right sorts of dispositions to change state. What the thermostat is disposed to do depends on its current state, as shown:



For example, when it's the ON state, the thermostat has the robust disposition to stay in the ON state when the temperature is low, and to move into the OFF state otherwise. And it has the reverse dispositions when it's in the OFF state.

3. It manifests its dispositions to change state. For example, when it is in the OFF state and the temperature is low, it moves into the ON state. When it is in the OFF state and the temperature is high, it stays in the OFF state.

A system must meet all of these conditions in order to regulate the temperature.[2] In particular, in order to regulate the temperature, it is not enough for a system to have the right sorts of states and the right sorts of dispositions to change state. If such a system doesn't also manifest its dispositions to change state, then that system doesn't succeed in regulating

---

[2] It has to meet some other conditions, too, but they aren't important for present purposes.

the temperature.

For example, consider Unlucky Thermostat. Unlucky Thermostat has the same construction as my thermostat. Unlucky is built into my friend's oven, which is set at 375 degrees. At this moment, Unlucky is in its OFF state and the oven is cold. Unlucky has the right sort of dispositions to change state. It is disposed to move into the ON state when the oven temperature is low, and to stay in the OFF state otherwise. But it doesn't display those dispositions. Instead, there is a giant microphysical coincidence in the exact configuration of air molecules in the oven and in the molecules that compose Unlucky. This coincidence prevents Unlucky from going into its ON state, even though the oven is cold. Eventually Unlucky does go into its ON state. But later, once the oven is very hot, another microphysical coincidence prevents Unlucky from going into its OFF state, even though it's disposed to do so.

Unlucky has the right sort of dispositions to change between its ON and OFF states. But much of the time, it doesn't change state in accord with those dispositions. Therefore, Unlucky doesn't regulate the temperature in my friend's oven.

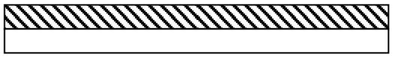
2.2 How my thermostat works

Here's how the thermostat in my oven works.[3] It contains a bimetal strip: a strip consisting of two different conductive metals that have been bonded

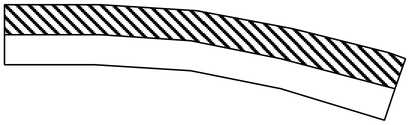
---

[3] Actually, this is how my 7th grade science teacher said a simple thermostat might work. But it will serve our purposes.

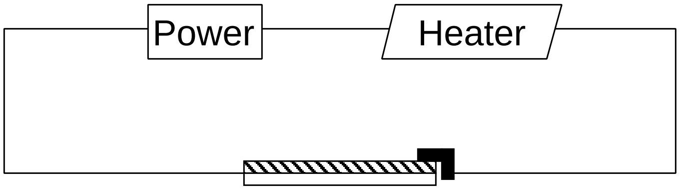
together:



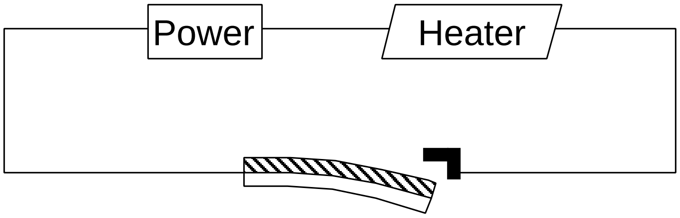
When it is heated, the top part of the strip expands more than the bottom part does. That difference makes the strip curl downwards:



The strip acts as a heat-sensitive switch. When the temperature is low, the strip lies flat, completing a circuit that powers the oven's heating element:



This is the thermostat's ON state. When it gets hot enough, the strip curls. That breaks the circuit and turns off the heating element:



This is the thermostat's OFF state.

Notice that when the thermostat changes state, it undergoes a thermodynamic change: the parts of the strip either expand or contract when the strip changes temperature. This fact will be important below.

2.3 There are no thermostats in entropy-decreasing worlds

In entropy-decreasing worlds, bimetal strips don't act as switches in temperature-regulating systems. Such strips *have* the right sorts of robust dispositions to act as temperature-sensitive switches. But they don't *manifest* those dispositions.

Here is a bimetal strip in an entropy-decreasing world. It is cold, flat, and surrounded by cold air. It has the robust disposition to stay flat when surrounded by cold air.[4] So far, so good. But though the air remains cold, does the strip do what it's disposed to do and stay flat? No! Instead, it spontaneously heats up and curls by sucking heat from the cold air. Later the air heats up.

Bimetal strips in entropy-decreasing worlds do this sort of thing all the time. (Doubt it? Note that spontaneous-curling process is the time-reverse of the following familiar one: A bent strip is surrounded by hot air. The air suddenly cools off, then the strip cools off and flattens.) When a bimetal strip in an entropy-decreasing world is disposed to undergo one sort of thermodynamic change, it actually undergoes another. (See [2] for details.)

Now consider an arbitrary system in an entropy-decreasing world. Does it regulate temperature? Suppose that the system has two states: ON and OFF. And suppose that the system has appropriate robust dispositions to change state. If the way the system changes state is by undergoing a

---

[4] Here I have applied the *time-slice* analysis of counterfactuals, described in [2].

thermodynamic change, then the system does *not* regulate temperature. That's because in entropy-decreasing worlds, systems don't manifest robust dispositions to undergo thermodynamic changes. (I have argued for this claim in [2].) So the system doesn't manifest its dispositions to change state. And manifesting its dispositions to change state is necessary in order for a system to regulate the temperature.

For example, suppose that the change from the system's ON state to its OFF state involves the heating and curling of a bimetal strip. Then the system doesn't manifest its robust disposition to stay ON when the air temperature is low. That's because (as we just saw) in entropy-decreasing worlds, bimetal strips quite often curl even when the temperature is low.

Moral: Restrict attention to those systems that change state by changing thermodynamically. In entropy-decreasing worlds, no such systems regulate the temperature.

3 Experience

This section argues that a similar moral holds in the case of conscious experience. The argument has two premises.

PREMISE 1 In order to have an experience, a system must manifest certain robust dispositions to change state.

I argue for this premise in §3.1, below.

The second premise was the main conclusion of the previous chapter:

PREMISE 2 In worlds with fundamental laws rather like ours, but in which entropy doesn't increase, no system manifests robust dispositions to undergo thermodynamic change.

Consider any system *S* that changes state by undergoing thermodynamic change. Suppose that the system inhabits a world with fundamental laws rather like ours. It follows from PREMISE 2 that this system doesn't manifests robust dispositions to change state. But according to PREMISE 1, the only way to have an experience is to manifest robust dispositions to change state. So it follows from PREMISE 1 that *S* doesn't have any experiences.

In other worlds, Premises 1 and 2 together entail:

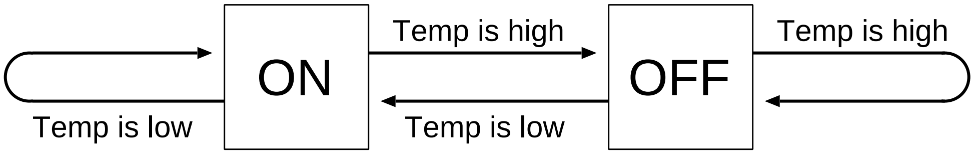
CONCLUSION In worlds with fundamental laws rather like ours, among systems that change state by changing thermodynamically, conscious experience requires the increase of entropy.

I have defended PREMISE 2 in the previous chapter. It remains only to defend PREMISE 1.

3.1 Defending Premise 1

Back to Fred. Fred has a great time eating his tuna melt. What makes it the case that Fred has such a great time is that he has an appropriate sequence of mental states. Vastly simplifying: he starts out in a mental state that represents that there is a melt in front of him, and that it will be delicious. In that state Fred's salivary glands are active, and his memories of previous tuna melts are primed. Provided certain other conditions are satisfied (for example, provided that he believes that no one has poisoned the melt), that state disposes Fred to move his hands toward the melt. And so on.

Fred manages to have this sequence of mental states in part by having a certain *functional organization*. The functional organization of a system is the structure of its states: what states it has, and how it is disposed to act and change state when it is in each one. For example, this diagram from the last section illustrates the functional organization of a simplified thermostat:



It's a little misleading to talk of "the" functional organization of a system, since any system instantiates many functional organizations, at different levels of description. For example, the above diagram represents the functional organization of the thermostat at a very coarse level of description. A finer-grained characterization might distinguish between a state in which the thermostat's bimetal strip is partly bent and one in which it is completely bent. Or such a characterization might distinguish between a state in which the bimetal strip is clean, and one in which a tiny speck of dust has settled on it. An extremely finely-grained characterization would distinguish between functional states of the thermostat on the basis of *any* physical difference between states, however minute. On such a characterization, the relevant dispositions to change state would have extremely detailed manifestation-conditions.

When evaluating whether a system counts as a thermostat, however, such a finely-grained characterization of it functional states is not relevant. What is relevant is a characterization of the system's functional organization at a coarser level of description. What's wanted are dispositions to change state such as "Move into the ON state when the temperature is low", and not ones such as "Move into state 493,302,302 (in which particle 1 is at position (0.1, 2.30, 2.43) . . ., particle 2 is at . . .) whenever the state of the entire rest of the universe is Φ", (where "Φ" abbreviates a detailed specification of the state of the entire rest of the universe).

Fred has the functional organization that he does in virtue of having a certain sort of brain and body. Fred changes state (for example, when he starts to reach for the melt) in virtue of his body and brain changing state. For instance, neurons fire in his brain, he salivates, and blood flows through his muscles. Likewise, Fred has certain robust dispositions to change state in virtue of his brain and body having robust dispositions to change state.

Note that Fred has fragile dispositions to change state, as well. For example, he has dispositions whose occurrence conditions specify the exact physical state of his environment. But these dispositions aren't relevant to Fred's functional organization at the level of description relevant for evaluating whether Fred has conscious experience.

In sum: Fred has experiences by manifesting certain robust dispositions to change state.

Fred's no fluke.

Every actual creature that has experiences does so by manifesting certain dispositions to change state. Of course, different creatures have different functional organizations. But every experiencing creature has *some* functional organization, and changes state in roughly the way it's disposed to do so.

That's no fluke, either. In order to have an experience, it is necessary to manifest certain dispositions to change state.

Let me illustrate this claim with some examples.

MOMENTARY MAN Momentary Man comes into existence in a giant microphysical coincidence. He forms spontaneously out of a pile of minerals. Once he is formed, Momentary Man has the same structure that Fred has when he is midway through eating the melt. For one microsecond, Momentary Man is a near-duplicate of Fred. In particular, his body and brain have all of the same dispositions that Fred's body and brain do. At the end of that microsecond, Momentary Man is destroyed.

While he exists, Momentary Man has all the dispositions to change state that are required in order to have an experience. But he doesn't exist long enough for those dispositions to be manifested. For example, no neural signals are transmitted during his existence, and the state of his body changes only negligibly. So Momentary Man never undergoes any experience.

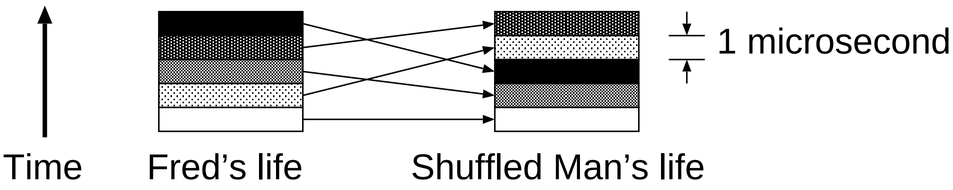
We often attribute phenomenal states to subjects at particular times. We say things such as, “Fred is enjoying his tuna melt at time *t*.” Such sentences make it superficially sound as though Fred’s “enjoyment at *t*” depends only on what’s going on with Fred *at t*. But this superficial appearance is misleading. Enjoying a tuna melt is an experience that involves changes that occur over an interval of time. Part of what is required in order for Fred to enjoy his tuna melt at *t* is for Fred to be in appropriate states at times surrounding *t*.

For example, it is widely thought that many of the brain’s computational states are realized by neurons firing at particular rates. If that’s right, then in order for Fred to be enjoying his melt at *t*, certain of his neurons must be firing at particular rates at *t*. But the proposition that some neuron is firing at rate *r* at *t* concerns more than just what’s going on at *t*. That proposition concerns what the neuron is doing at a time interval *surrounding t*—a time interval that is larger than the time it takes for a neuron to fire once. That time interval is much longer than a microsecond. So Momentary Man doesn’t exist long enough for any of his neurons to have well-defined firing rates.

The upshot is that Momentary Man fails to undergo any experience because he fails to be in appropriate functional states at times closely surrounding the short interval during which he exists.

Another example:

SHUFFLED MAN Chop up Fred's life into temporal stages, each of which is a microsecond long. Now shuffle the stages around in a random order and reassemble them. That's roughly what Shuffled Man's life is like.



In more detail: Shuffled Man's life starts the same way that Momentary Man's does. For a microsecond, he is a near duplicate of Fred. During that microsecond, his body and brain have the same dispositions that Fred's body and brain do. But at the end of the microsecond, a giant microphysical coincidence suddenly puts him into the state that Fred was in an hour ago. After another microsecond, another coincidence puts Shuffled Man into the state Fred will be in next year. Shuffled Man continues in this way: every microsecond, a coincidence gets him to change state in a way that is contrary to his dispositions.

At any moment of time, Shuffled Man has all the dispositions to change state that are required in order to have an experience. But he never manifests those dispositions. As a result, there's a mismatch between his functional organization and the sequence of states that he actually occupies. For this reason, Shuffled Man doesn't have any experiences.

These examples make PREMISE 1 plausible. They make it plausible that in order to have an experience, it's not enough to have the right sort of dispositions to change state. One also needs to actually change one's state in accord with those dispositions.

Note that PREMISE 1 is the claim that manifesting certain dispositions is *necessary* in order for a system to have an experience. This is much weaker than the claim that manifesting such dispositions is necessary *and sufficient* for having an experience. That stronger claim is a version of the controversial doctrine of *functionalism*. One need not accept functionalism in order to endorse PREMISE 1.

3.2 The scope of the conclusion

Having defended Premises 1 and 2, we're in a position to draw the main conclusion:

CONCLUSION In worlds with fundamental laws rather like ours, among systems that change state by changing thermodynamically, conscious experience requires the increase of entropy.

This conclusion is qualified in two ways. It concerns only worlds with fundamental laws like ours, and only systems that change state by changing thermodynamically. Why the qualifications? Why not just say:

> Conscious experience requires the increase of entropy.

The qualification about the laws is necessary because in worlds with strange laws, all hell can break loose. For example, I've relied throughout on the notion of entropy. But there are worlds in which there is no well-defined notion of entropy at all.

Another example: In the previous chapter, I argued that entropy-decreasing trajectories are fragile—that small modifications of an entropy-decreasing process rarely result in another entropy-decreasing process. That argument is plausible in worlds with fundamental dynamical laws rather like those of our world. But it is not plausible with respect to worlds with very strange dynamical laws.

The qualification about thermodynamic change is also necessary. Consider a system that has a complicated functional structure, but which *doesn't* change state by changing thermodynamically. For example:

ROCK COMPUTER Rock Computer inhabits a world in which entropy decreases. It consists of a gigantic arrangement of inert chunks of rock floating in space. Rock Computer's functional states are realized by different arrangements of the rocks. Its state changes when the rocks orbit into new arrangements.

Rock Computer doesn't change state by changing thermodynamically, since the mere orbiting of inert rocks involves only negligible thermodynamic change. I have no idea whether it is possible for a system like Rock Computer to instantiate the sort of functional organization that would allow it to have experiences. But CONCLUSION does nothing to rule out such a possibility. That's because Rock Computer manifests robust dispositions to change state. So (for all the argument shows), Rock Computer has experiences even though it inhabits an entropy-decreasing world.

So CONCLUSION has two significant qualifications, and both are required. Nevertheless, CONCLUSION applies to a very wide range of systems. Systems such as Rock Computer operate on utterly different principles than the ones that any known computer or creature do. For example, when Fred changes state, he certainly does so by changing thermodynamically. The firing of neurons, the pumping of blood, and the operation of salivary glands—these involve temperature and pressure changes, chemical reactions, or the diffusion of substances from higher to lower concentration. Every biological process we know of involves some sort of thermodynamic change. So Derf and his fellow time-reversed organisms change state by changing thermodynamically. Therefore, CONCLUSION entails that none of these creatures have experiences.

It is a controversial matter whether the increase of entropy has much to do with conscious experience. John Earman seems skeptical of such a connection when he writes:

> One rather strong and interesting claim which has been made on behalf of entropy is that the entropic behavior of branch systems dictates the choice of the future direction for time which human beings make; for, it is claimed, (a) human beings could

> not long survive in an environment in which entropy is on the decrease, and/or (b) the functioning of the biological mechanisms which underlie our perception of time order and our sense of the flow of time depend in an essential way upon the entropic behavior of branch systems. These claims are as crude as they are bold; for, so far as I now, there is little or no scientific evidence to support them, and the evidence we do have suggests that perception of time order is more subtle and complex than (b) would have it. [1]

I agree that claim (b) is crude. But as we've seen, there is reason to think it is true.

References

[1] John Earman. An attempt to add a little direction to "the problem of the direction of time". *Philosophy of Science*, 151:15–47, 1974.

[2] Adam Elga. Dispositions and entropy. Manuscript, 2002.