**Paying Attention to the Difference Between Bugs and Features in Theories**
Dear Professor Aaronson,

We met in the fall of 2015 at the Integrated Information Theory conference hosted by David Chalmers. I appreciate your blog post dedicated to my work. However, there are some misunderstandings I’d like to clear up, either in private or, if you’d be so kind, as a link or post on your blog. I ask this because you once gave Giulio Tononi a chance to respond to your comments about Integrated Information Theory in a way that opened up an informative dialogue. Since your post is so specifically about my recent work, I’d love the chance to address your major claims to your audience. The claims I address are below:

“*In their new work, Hoel and others claim to make the amazing discovery that scientific reductionism is false…*”

We only say that reductionism is not *universally* true. We’re very clear that it depends on the properties of the system under consideration, and we give explicit examples where reductionism is the best strategy. As one of the main goals of causal emergence is generate insight into why science is structured in the hierarchical way that it is (economics above psychology, psychology above biology, biology above chemistry, and so on) a universal statement that reductionism is always false wouldn’t make sense. It depends on the properties of the system. So this is a far more reasonable claim than Scott makes it out to be.
 *Rather, they disagree with the meta-level claim that there’s anything shocking about such*[higher-level] *causation, anything that poses a special difficulty for the reductionist worldview that physics has held for centuries.  I.e., they consider it true both that*

*1.) my nose is made of subatomic particles, and its behavior is in principle fully determined (at least probabilistically) by the quantum state of those particles together with the laws governing them, and*

*2.) my nose itched.*

*…there seems to be no reason to imagine a contradiction between 1 and 2 that needs to be resolved, but “only” a vast network of intervening mechanisms to be elucidated.*

The above isn’t directly referring to the theory of causal emergence, but I want to address this reiterated point that there is no problem to begin with. What Scott describes is an event at two different levels, and then declares that there’s no conflict because there’s a “vast network of intervening mechanisms." But how can there be “intervening mechanisms” between different supervening levels of description? A supervening re-description can’t intervene on or cause itself. Huge numbers of people have published papers about exactly these issues in analytic philosophy. I’d refer interested readers to the best one, which is Jaegwon Kim’s “exclusion argument” (Kim 2000), which tanks simple examples like these. The exclusion argument says that 2) is epiphenomenal because 1) is causally closed, and therefore 2) is unnecessary except as shorthand for 1). So there is a problem to be solved.

However, rather than philosophical hand waving, investigation into emergence should lead to actionable information, such as how to engineer the causal relationships of systems, or how to tell at what scale of a system scientific experiments will be the most informative. So these are real scientific issues, not just things that disappear if we don’t feel the urge to resolve them.

*“…EI, like Φ, has aspects that could be criticized as arbitrary, and as not obviously connected with what we’re trying to understand*.”

This is an inapt comparison. Φ is a very complex measure with lots of moving parts. It has what might be called a large “attack surface” for a theory.*EI*was specifically used to demonstrate causal emergence because it is simple, non-arbitrary, and well-supported. *EI* is the mutual information between a set of interventions and their effects. Mutual information is basically the opposite of an arbitrary measure – it’s well formalized and widely used. *EI*has also been proven to be decomposable into causal properties like the determinism (sufficiency) and degeneracy (necessity) of causal relationships (Hoel et al. 2013). So I don’t see how Scott can possibly claim it’s not obviously connected with what we are trying to understand: causal structure.

“…*it’s clear that the entire thing boils down to, how shall I put this, a normalization issue…*.*we insist on the uniform distribution over microstates when calculating microscopic EI, and we also insist on the uniform distribution over macrostates when calculating macroscopic EI, and we ignore the fact that the uniform distribution over microstates gives rise to a non-uniform distribution over macrostates, because some macrostates can be formed in more ways than others.* ***If we fixed this****, demanding that the two distributions be compatible with each other, we’d immediately find that, surprise, knowing the complete initial microstate of a system always gives you at least as much power to predict the system’s future as knowing a macroscopic approximation to that state.*

This is Scott’s fundamental objection. Unfortunately, it is fundamentally wrong. It is wrong both at the conceptual level but also wrong in the technical details. Let’s go through why it’s wrong technically first, as that will lead us into why it’s wrong conceptually.
 There are numerous technical reasons given in the paper why *EI* needs to use the uniform distribution at different scales to accurately capture causation. None of which Scott mentions in his blog post. At a technical level, the uniform distribution screens off the marginal probabilities of states, which makes *EI*sensitive only to the conditional probabilities of state-transitions, that is, the causal structure of the system. The uniform distribution is what gives *EI*its proven connection to sufficiency and necessity, and makes it a measure of causal influence.
           Let me put it in even more concrete terms using a simple example (mentioned in the paper but, again, not in Scott’s post). Consider the deterministic causal relationship between a light switch and a light bulb. According to *EI*, this causal relationship should carry 1 bit. If the light switch is on, then the light bulb will turn on, and if the light switch is off, then the bulb will turn off. Importantly, this is true no matter what microstates the macrostates of the light switch are composed of. A small light switch or a big light switch have the same *EI.* Perhaps many microstates make up the “on” macrostate. It doesn’t matter. The light switch generates 1 bit of *EI*.

Let’s now consider Scott’s newly proposed “fixed” version of *EI*to see how wrongheaded it is. In this “fixed” version of *EI*, a normal everyday higher-level causal relationship like that of light switch and a light bulb is totally unbounded and undefined, no matter the obvious properties of that relationship. Even worse, by throwing out the uniform distribution *EI* loses its proven connection to causal properties like determinism (sufficiency) and degeneracy (necessity). It also loses its sensitivity to the conditional probabilities of state transitions. Not using the uniform distribution would mean that *EI* would be dependent on the frequency of interventions, not their effects. For instance, the causal relationship between a light switch and a light bulb would change drastically depending on how often someone left it on. Even simple logic gates would have *EI* values that wouldn’t correspond to what their mechanisms are (binary COPY gates wouldn’t have 1 bit, etc). It is like Scott is saying “I don’t like the result because it’s merely a consequence of enforcing *x*!” and then when *x* is not enforced it breaks literally everything, not just the result he doesn’t like. So Scott’s objection is wrong at the technical level about how *EI* should be calculated, based off what it’s meant to capture (causal structure).

At a conceptual level the objection makes even less sense. Conceptually, Scott’s argument is that one could get rid of causal emergence by simply demanding “*that the two distributions*[at different scales]*be compatible with each other.”* Of course, they already are compatible, as there’s a well-defined mapping between them. What Scott is proposing is to willfully change *EI* to “fix” it by enforcing the same underlying distribution of interventions at all scales. I think some of this is based on Scott’s confusion in thinking we are failing to do something that is obviously correct, as he also claims that it “*…seems reasonable to me to demand that in the comparison, the macro-distribution arise by coarse-graining the micro one.*”This is precisely what the theory claims: that macro-interventions are a coarse-grain of their underlying interventions. This leads to causal emergence. What Scott proposes is precisely the opposite: that when one coarse-grains the states or mechanisms of a system then the associated intervention distribution over those states or mechanisms *shouldn’t* really be coarse-grained. Rather it should be a sort of weighted average that tracks how many micro-interventions go into each macro-intervention. This like saying that temperature should always be weighted by how many particles there are.

The problem with Scott’s proposal, besides it being a technical mess and making the measure give nonsensical answers at all but the lowest possible scale, is that he’s trying to get around a property called multiple-realizability that is definitional to macro-states. Therefore, multiple-realizability is also definitional to macro-interventions on those states. What Scott is missing here is that causal emergence is about how causal structure objectively changes across scales, precisely due to this multiple-realizability. As a measure, *EI* is sensitive to these changes in the causal structure. There is no secret trick buried within *EI* itself, there is just the mechanism by which it is sensitive, and he’s confusing that mechanism with the thing it is sensitive to.

For instance, the conditional probabilities of state-transitions *objectively* change across scales, e.g., by becoming more deterministic. These changes in conditional probabilities at different scales lead to *objective* changes in the determinism and the degeneracy of the causal structure of the system. This is not a matter of some arbitrary normalization, as Scott makes it out to be in his post. *EI*is sensitive to such *objective* changes in causal structure across scales. Similar measures would be similarly sensitive. This whole thing is about noise minimization, not normalization. That’s a significant misunderstanding.

The universal-reductionism-saving kluge that Scott proposes only works if the intervention distributions used for the macroscale and microscale are *always* identical. As long as there is *any* change in the intervention distribution at the macroscale there can be causal emergence. Why is this? As Shannon discovered, for an information channel different input distributions use more or less of a channel’s capacity to transmit information. In a mathematically analogous manner, different sets of interventions, such as at different scales, use more or less of a system’s causal structure (or causal capacity or work). Neither is a “normalization trick.” They are features, not bugs.

I do think Scott was correct to point out there is a brief section in the discussion that at first read seems to support his point. I was trying to answer a different matter though, about how observers and experimenters are useful for formalization purposes when talking about causation. Just to make clear, let me emphasize the important part of what I wrote here: “*For formalization purposes, it is the experimenter who is the source of the intervention distribution,* ***which reveals a causal structure that already exists****.*” This is precisely what I’ve said here: the causal structure really does have different properties at different scales. This is objectively true and *EI* is a measure of this. But I agree this is buried in the paragraph. I go on to claim that experimenters aren’t necessary to pick out macroscales, as some systems may be obviously engaged in macroscale behavior over their input/outputs.

Ultimately, multiple realizability means that information and causal influence can increase at higher-scales precisely because multiply-realizable causal relationships can correct for errors. But Scott doesn’t mention the connection to the well-proven insight of Shannon that block-coding and other forms of coarse-graining or grouping reduce noise and allow for the transmission of extra information, which is by no means some “normalization trick,” despite being very similar mathematically (Hoel 2017). His only real objection consists of a proposal that that we could, if we wanted, break our measure of causation so that it no longer tracks the objective changes to the causal structure across scales. His reason for doing this is explicitly just to counterbalance multiple-realizability so that causal emergence cannot happen. It’s not much of an objection.
 To note, I think there are reasonable debates to be had about many aspects of the theory. None of Scott’s specific objections actually work out, and I’m not sure why he was so fervent about making them. But I do greatly thank him for allowing me to clarify these points better. The theory is still developing so please feel free to contact me if you’re interested in these topics, especially if you have a background in related fields.

Thank you for taking the time to read - all the best,

Erik Hoel
[www.erikphoel.com](http://www.erikphoel.com)
Columbia University
Department of Biological Sciences
New York, N, 10027

|  |
| --- |
|  |

**Citations:**

Hoel, E. P., Albantakis, L., & Tononi, G. (2013). Quantifying causal emergence shows that macro can beat micro*. Proceedings of the National Academy of Sciences*, *110*(49), 19790-19795.

Hoel, E. P. (2017). When the map is better than the territory. *Entropy*, *19*(5), 188.

|  |
| --- |
|  |

Kim, J. (2000). *Mind in a physical world: An essay on the mind-body problem and mental causation*. MIT press.