Is Entropy a Degenerative Concept?

Eugene Chua

"As a young man I tried to read thermodynamics, but I always came up against entropy as a brick wall that stopped my further progress. I found the ordinary mathematical explanation, of course, but no sort of physical idea underlying it. No author seemed even to try to give any physical idea. Having in those days great respect for textbooks, I concluded that the physical meaning must be so obvious that it needs no explanation, and that I was especially stupid on the particular subject."
(Swinburne 1904, 3)

"My greatest concern was what to call it. I thought of calling it 'information', but the word was overly used, so I decided to call it 'uncertainty'. When I discussed it with John von Neumann, he had a better idea. Von Neumann told me, 'You should call it entropy, for two reasons. In the first place your uncertainty function has been used in statistical mechanics under that name. In the second place, and more importantly, no one knows what entropy really is, so in a debate you will always have the advantage."
(Shannon according to Tribus & McIrvine 1971, 180)

Abstract

Lakatos's analysis of growth and degeneration in the Methodology of Scientific Research Programmes (1978) is well-known. Less known, however, are his thoughts on degeneration in Proofs and Refutations (1976). I propose and motivate two new criteria for degeneration based on the discussion in Proofs and Refutations – *superfluity* and *authoritarianism*. I then show how these criteria can fruitfully augment the criteria provided in the Methodology of Scientific Research Programmes, providing a generalized Lakatosian account of progress and degeneration. As a proof of concept, I employ this generalized account in evaluating a key transition point in the history of entropy – the transition to an information-theoretic interpretation of entropy – by assessing Jaynes's 1957 paper on information theory and statistical mechanics.

Word Count (Main Text): 11651

I. Introduction

Lakatos's (1976/2015) goal in *Proofs and Refutations* **(P&R)** was to argue that the comprehension of mathematical concepts must be accompanied with a clear understanding of how and why the concept came into existence. We must understand the concept's *problem-situation*, the problems or questions which led to the concept's genesis and evolution, *alongside* an understanding of the concept itself. A concept is not an *atom*. Instead, we should understand a concept as a temporally extended *process* through which the initial, primitive, concept is continually refined through past iterations. To rip a concept apart from its context of discovery is to miss a complete understanding of the concept.

All of the above – concerning how to *comprehend* a concept – has been much discussed over the last few decades.[1] What has been less discussed, to my knowledge, is how to *evaluate* a concept according to the heuristic approach presented in **P&R**: how do we know whether a concept is problematic, needs rehabilitation, or, worse still, must be abandoned, given the heuristic approach? In short, how do we know whether a concept is *degenerating*?

That not much has been said about this is curious, since Lakatos clearly had some such standard in mind. In this paper, I explore Lakatos' views on degeneration in **P&R**, which has often been neglected for the sort of degeneration discussed in *Methodology of Scientific Research Programmes* (**MSRP**). It seems to me that **P&R** offers new criteria for degeneration that sheds light on Lakatos's approach.

The primary goal here is to provide and motivate an account of degeneration based on **P&R**, and to show how this account of degeneration differs from the account found in Lakatos's later work in **MSRP**. In my view, the notions of degeneration suggested in **P&R** depart from the two conditions (to be discussed) found in **MSRP**, because the former do not relate to the *content* of a concept or theory, but to the *methodological* aspects of a theory (and the *depth* of research) instead. Based on my reading of **P&R**, I propose two criteria for degeneration: (i) *superfluity,* or generalization for generalization's sake, which involves the introduction of trivial extensions or terminology to a theory or concept, and (ii) *authoritarianism*, the introduction and employment of a concept into the discussion without justification while ignoring the problem-situation of said concept. This leads to my proposal for an extended account of Lakatosian degeneration, which takes into account the degeneration of both content and depth.

With this extended account in hand, the secondary goal of the paper is to apply this account of degeneration to the historical trajectory of entropy. The concept of entropy (or rather, the family of concepts called 'entropy') has a tumultuous past, with many interpretations,[2] complications,[3] and disagreements colouring its rich history. This is coupled, however, with its extensive usage (contentiously, I must add) in countless sciences, be it in black hole thermodynamics,[4] quantum theory,[5] AdS/CFT research in theoretical physics,[6] and even neuroscience.[7] This makes entropy an interesting case study for degeneration. I evaluate a key transition point for the concept of entropy – the transition from thermodynamic to information-theoretic interpretations of entropy. In Lakatosian style, I identify a key piece of writing and evaluate it with the twin criteria I have developed. By critiquing Jaynes's (1957) landmark paper on thermodynamics and information, I argue that this transition suffered from superfluity and authoritarianism. As such, given the extended Lakatosian account, entropy is a degenerative concept.

II. Degeneration in P&R

I begin by recapitulating Lakatos's (1976) approach in **P&R**. His main target was a 'deductivist' approach to mathematics:

> This style starts with a painstakingly stated list of axioms, lemmas and/or definitions. The axioms and definitions frequently look artificial and mystifyingly

[1] See e.g. Corfield (1997), Leng (2002) or Werndl (2009).
[2] See Uffink (2001) for how interpreting the original thermodynamic concept of entropy is itself a challenge.
[3] See e.g. Callender (1999), Goldstein *et al* (2020) or Chua (forthcoming) for some of these complications concerning how different entropies relate to others.
[4] See e.g. Dougherty and Callender (forthcoming) or Wallace (preprint).
[5] See e.g. Bub (2005).
[6] See e.g. Natsuume (2015).
[7] See Friston and Stephan (2007).

complicated. One is never told how these complications arose. The list of axioms and definitions is followed by the carefully worded theorems. These are loaded with heavy-going conditions; it seems impossible that anyone should ever have guessed them. The theorem is followed by the proof. (Lakatos 1976/2015, 151)

In Lakatos's view, however, this approach to mathematics is misguided. By rationally reconstructing the historical development of the Euler characteristic:[8]

$$V - E + F = 2$$

he showed that these definitions, axioms, theorems and so on developed only as the result of a long history of proofs and refutations: they are *proof-generated concepts*.

For Lakatos, actual mathematics is not deductivist. Nevertheless, it is a *rational* affair representative of what he called the *heuristic approach*:

> …deductivist style tears the proof-generated definitions off their 'proof-ancestors', presents them out of the blue, in an artificial and authoritarian way. It hides the global counterexamples which led to their discovery. Heuristic style on the contrary highlights these factors. It emphasises the *problem-situation*: it emphasises the 'logic' which gave birth to the new concept. (1976/2015, 153)

More generally, a concept is only appropriately understood when we understand its historical trajectory. In Lakatos's view, "there is a simple pattern of mathematical discovery– or of the growth of informal mathematical theories" (1976/2015, 135–136), which is given by the following seven stages, each stage involving what Lakatos called informal reasoning, or 'heuristics':

(1) *Primitive conjecture*.
(2) *Proof* (a rough thought-experiment or argument, decomposing the primitive conjecture into subconjectures or lemmas).
(3) *'Global' counterexamples* (counterexamples to the primitive conjecture) emerge.
(4) *Proof re-examined*: the 'guilty lemma' to which the global counterexample is a 'local' counterexample is spotted. This guilty lemma may have previously remained 'hidden' or may have been misidentified. Now it is made explicit, and built into the primitive conjecture as a condition. The theorem – the improved conjecture – supersedes the primitive conjecture with the new proof-generated concept as its paramount new feature.
(5) *Proofs of other theorems are examined to see if the newly found lemma or the new proof-generated concept occurs in them*: this concept may be found lying at cross-roads of different proofs, and thus emerge as of basic importance.
(6) *The hitherto accepted consequences of the original and now refuted conjecture are checked*.
(7) *Counterexamples are turned into new examples – new fields of inquiry open up*.

Call this the *heuristic process*. This tells us how a mathematical theory or concept ought to *grow*, from a rough primitive conjecture to a proof-generated concept (and beyond), through the use of heuristics like employing counterexamples, discovering hidden lemmas, and so on. We only comprehend a concept appropriately when we start with the primitive conjecture – the genesis of the concept – *and* grasp the ensuing adjustments and responses to the concept through which it is precisified and stretched.

[8] The Euler characteristic was initially postulated to universally describe polyhedra in terms of the number of faces (F), vertices (V), and edges (E) they have.

But as Gamma[9] points out, *growth* is opposed to *degeneration*. (1976/2015, 103) However, while Lakatos paints a clear picture as to how mathematical theories grow, he is much less clear in **P&R** about what it means for degeneration to occur.

This is curious because the term 'degeneration' is used on several occasions in **P&R** and seems to bear significant normative weight in assessing the methodology of research. In what follows, I will first pinpoint two distinct criteria for degeneration, before examining their normative import.

A. Superfluity

The first criterion for degeneration appears when Alpha charts the development of the dialogue in **P&R** thus far, in response to ever-exotic counterexamples: (1976/2015, 86)

> **(1)** one vertex is one vertex.
> **(2)** $V = E$ for all perfect polygons.
> **(3)** $V - E + F = 1$ for all normal open polygonal systems.
> **(4)** $V - E + F = 2$ for all normal closed polygonal systems, i.e. polyhedra.
> **(5)** $V - E + F = 2 - 2(\mathrm{n} - 1)$ for normal n-spheroid polyhedra.
> **(6)** $V - E + F = 2 - 2(\mathrm{n} - 1) + \sum_{k=1}^{F} e_k$ for normal-n-spheroid polyhedra with multiply-connected faces.
> **(7)** $V - E + F = \sum_{j=1}^{K}\{2 - 2(\mathrm{n}_j - 1) + \sum_{k=1}^{F} e_{kj}\}$ for normal n-spheroid polyhedra with multiply-connected faces and with cavities.

Alpha proclaims:

> Isn't this a miraculous unfolding of the hidden riches of the trivial starting point? And since (1) is indubitably true, so is the rest.

Rho retorts: "Hidden 'riches'? The last two only show how cheap generalisations may become!" Gamma concurs:

> (6) and (7) are not growth, but *degeneration*! Instead of going on to (6) and (7), I would rather find and explain some exciting new counterexample [to $V - E + F = 2$]! (1976/2015, 103, emphasis mine)

So, degeneration is tied in part to 'cheap generalisation' in the process of developing a concept or theory. It plays an evaluative role insofar as it tells us when, in something like the chain of generalizations from (1) to (7), we should stop in response to a new counterexample. Gamma describes (7) as a pointless generalization: who cares about polyhedrons with cavities or multiply-connected faces? To Gamma, "it serves only for making up complicated, pretentious formulas for nothing." (1976/2015, 102 – 103)

A cheap generalization, in Gamma's view, is a *trivial extension* of a concept. (1976/2015, 102) Why was counting cavities and multiply-connected faces important to begin with? It was not clear why researchers should care, and these new additions are not well-motivated. Nothing *deep* was gleaned from this generalization, even though our formula does become more general. It is, to put it bluntly, *generalization for the sake of generalization*. When a concept is extended in order to encompass new

[9] **P&R** takes the form of a dialogue between a fictional teacher and his students which includes Gamma (and later Alpha, Rho and Zeta, among others).

cases, but those cases were not relevant to the original problem-situation (or the heuristics that followed), Gamma would consider them trivial extensions.

Lakatos, channelling Pólya, assents to this in his description of what cheap generalization amounts to:

> Pólya points out that shallow, cheap, generalisation is 'more fashionable nowadays than it was formerly. *It dilutes a little idea with a big terminology. The author usually prefers to take even that little idea from somebody else, refrains from adding any original observation, and avoids solving any problem except a few problems arising from the difficulties of his own terminology*.
>
> Another of the greatest mathematicians of our century, John von Neumann, also warned against this 'danger of degeneration', but thought it would not be so bad 'if the discipline is under the influence of men with an exceptionally well-developed taste'. (1976/2015, 104, fn. 160, emphasis mine)

For our purposes, it certainly helps that Lakatos directly connects 'shallow, cheap, generalisation' to the 'danger of degeneration'. Alternatively stated, the generalization of a concept to new domains without justification is *superfluous* and only adds unnecessary terminology ("pretentious formulas"). This, in Lakatosian terms, is concept-stretching 'gone wrong' – a concept is stretched too far without justification, resulting in trivial generalizations. This leads to the concept's degeneration. Call this first criteria of degeneration *superfluity*.

What counts as justification and can avoid superfluity? Of course, the justification must be specific to the problem-situation at hand, but what else? We can distil further insights from Lakatos's discussion of the Euler characteristic: in that case, the problem-situation is one where the key concern is about classifying (what we intuitively might call) *polyhedra.* However, while cavities were helpful in categorizing polyhedra based on *Definition 1,* i.e. the *naïve* notion of a polyhedron as a solid, a single 'polyhedron with a cavity' corresponds to an entire class of polyhedra in the succeeding proof-generated concept of polyhedra as connected surfaces (Lakatos calls this *Definition 2*). (1976/2015, 16) On this definition, a polyhedron is not a *solid*. Given this background, the number of cavities does not actually pick out a unique type of polyhedra,[10] and hence any extension of the Euler characteristic which takes into account the number of cavities simply plays no real role in advancing the research of polyhedra.[11] Instead, we have merely added superfluous terminology which do not concern objects of interest, instead adding unnecessary generalizations.

From this we might infer that the sort of justification which vindicates any particular generalization or introduction of terminology is one that can motivate why this terminology or generalization was introduced – particularly in relation to the sort of objects or concepts we care about – and how it can possibly lead to the growth of the theory or concept. It cannot be just a *trivial pun,* which is essentially what a 'polyhedron with a cavity' is, since it is not a polyhedron *per se* at all given the problem-situation where Definition 2 is accepted as a proof-generated concept – it must serve some *use* and justify its own existence, so to speak.

[10] See Lakatos (1976/2015, 97, fn. 150(a)).

[11] Likewise for multiply-connected faces: it was simply not interesting for research because our means of classifying polyhedra simply did not require multiply-connected faces. As Lakatos notes, as the theory of polyhedra evolved, "the problem of how multiply-connected faces influence the Euler-characteristic of a polyhedron lost all its interest" (1976/2015, 97, fn. 150(c)).

As Lakatos notes, this requirement is not a clear-cut one, for it is not always clear what is trivial. Alpha questions: "You may be right after all. But who decides where to stop? Depth is only a matter of taste." (1976/2015, 103) In response, Gamma proposes:

> Why not have mathematical critics just as you have literary critics, to develop mathematical taste by public criticism? We may even stem the tide of pretentious trivialities in mathematical literature. (1976/2015, 104)

Lakatos does not say much more about what taste amounts to, though one guess would be the sort of 'scent' that physicists often talk about in relation to the fruitfulness of a theory – the scent that reveals whether something is worth pursuing, whether it generates interesting research, and so on, usually based on their experience in the field and the sort of people they know.

Hindsight helps, as in the case of cavities failing to pick out the relevant sort of properties for polyhedra: once we know what we count as relevant as a matter of practice, adoption, and actual contribution to the problem-situation, e.g. *Definition 2* as an improvement over *Definition 1,* then it is of course much easier to see what was irrelevant. As Kiss puts it succinctly (though in the context of the **MSRP**): "One step in a research program can be treated as progressive or degenerating only in hindsight, when we see future developments. Appraisal of research programs is as fallible as the theories themselves." (Kiss 2006, 316)

We can do better in other cases. In line with the heuristic approach, being consciously aware of the problem-situation – and the heuristics associated with it – allows one to avoid superfluity too. Lakatos discusses the example of the mathematician Becker, who aimed to provide a conclusion to the classification problem by providing a new generalization to the Euler characteristic:

$$V - E + F = 4 - 2n + q$$

where n is the number of cuts that is needed to divide the polyhedral surface into simply-connected surfaces for which $V - E + F = 1$, and q is the number of diagonals that one has to add to reduce all faces to simply-connected ones. (1976/2015, 103, fn. 158) Unfortunately for Becker, the mathematicians Lhuilier and Jordan had already written about this over half a century ago and provided far more general formulations in different terminologies. So while Becker's work does count as a valid generalization of Euler's original formulation, it was ultimately trivial – adding only new terminology – and did not contribute to the development of the concept. Here it is clear that a cognizance of the problem-situation would have helped: being aware of the concept's past iterations, problems and errors, one learns what *not* to do.

The lesson is to be aware of the problem-situation for the concept, and to avoid trivially generalizing it. Beyond that, we can only hope to pinpoint degeneration in retrospect unless we are blessed with the great gift of 'exceptional taste' in the subject matter at hand (in which case, a certain sort of clairvoyance – of what *will* work out in research – is possible).[12] However, given Lakatos's distaste for formalism and inclination towards informal heuristics – which are themselves not always precise – this might be exactly to Lakatos's liking.

[12] Interestingly, Corfield (2002, 117) accuses Lakatos precisely of failing to appreciate the role of taste, instead focusing too much on the axiomatization process: "what have typically been ignored by philosophers of mathematics are the judgement skills of mathematicians, their *bons sens* or *finesse* […] In effect, what I have been indicating above is the fact that Lakatos' appreciation of this *bon sens* was somewhat restricted." Yet, my discussion here suggests there was consideration of this notion, though Lakatos is skeptical that *bon sens* is enough to halt degeneration.

B. Authoritarianism

In his explication of the heuristic approach, Lakatos notes a common response to byzantine definitions (in this case, Carath*é*odory's definition of a measurable set) with disapproval:

> Of course there is always the easy way out: mathematicians define their concepts just as they like. But *serious teachers do not take this easy refuge*. Nor can they just say that this is the correct, true definition … and that mature mathematical insight should see it as such. (Lakatos 1976/2015, 162)

There is something problematic with the deductivist style of simply introducing concepts out of thin air in order to facilitate a proof without appropriately situating them within the proof's problem-background. There is something *authoritarian* about this approach:

> One can easily give more examples, where stating the primitive conjecture, showing the proof, the counterexamples, and following the heuristic order up to the theorem and to the proof-generated definition would dispel the *authoritarian mysticism of abstract mathematics*, and *would act as a brake on degeneration*. A couple of case-studies in this degeneration would do much good for mathematics. Unfortunately the deductivist style and the atomisation of mathematical knowledge protect '*degenerate*' papers to a very considerable degree. (Lakatos 1976/2015, 163, emphasis mine)

This provides a new criterion for degeneration, which occurs when a concept (or terminology, or definition, or perhaps even a theory) is introduced without justification into some line of inquiry. These new concepts are instead used with the attitude "that this is the correct, true definition" without qualification. This adds a 'mystical' and 'authoritarian' element to these new concepts which ignores the background problem-situation, i.e. the heuristics and the process of proofs and refutations, that led to that line of inquiry to begin with. Call this criterion for degeneration *authoritarianism.*

Lakatos raises the example of Rudin's discussion of bounded variation. while introducing the Riemann-Stieltjes integral, Rudin introduces the notion of bounded variation.[13] He then proves a theorem to the effect that a function of bounded variation, satisfying other criteria, is also a member of the class of Riemann-Stieljes integrable functions. However, Lakatos accuses Rudin of failing to explain why the Riemann-Stieljes integral was important or relevant to begin with, and why we should care about it and the notion of bounded variation:

> So now we have got a theorem in which two mystical concepts, bounded variation and Riemann-integrability, occur. But two mysteries do not add up to understanding. Or perhaps they do for those who have the 'ability and inclination to pursue an abstract train of thought'? (1976/2015, 156)

The *non-degenerative* way of presenting the concept would have shown that the two concepts arose as proof-generated concepts out of the same problem-situation:

> A heuristic presentation would show that both concepts – Riemann-Stieltjes integrability and bounded variation – are proof-generated concepts, originating in one and the same proof: Dirichlet's proof of the Fourier conjecture. This proof gives the problem-background of both concepts. (1976/2015, 156)

[13] For Lakatos's discussion, see Lakatos (1976/2015, 155 – 162).

Lakatos further notes that Rudin *does* mention this, but in a way that was disconnected from the two aforementioned concepts: it was hidden in some exercise in a different chapter. Again he declares that the two concepts, introduced this way, were introduced in "an authoritarian way".[14] (156, fn. 12)

Thus, both superfluity and authoritarianism arise from a failure to grapple with the problem-situation. While superfluity arises from trivial generalizations of concepts arising from this lack of awareness, authoritarianism arises from the unjustified introduction and application of concepts to new domains.

Instead of proceeding with this sort of degeneration, Lakatos thinks we should have proceeded from the original primitive conjecture of the problem-situation, which Lakatos situates in Fourier's conjecture that any arbitrary function is Fourier-expandable, i.e. 'expandable into a trigonometrical series with the Fourier-coefficients'. (157, fn. 14) From there we can find the heuristic approach at play, with Dirichlet proving Fourier's conjecture and further improved Fourier's conjecture by adding the lemmas he used and discovered in his proof into Fourier's original conjecture as conditions – the well-known Dirichlet conditions.[15]

Lakatos then comments:

> Dirichlet's proof-analysis was faulty only as regards the third condition: the proof in fact hinges only on the bounded variation of the function. Dirichlet's proof-analysis was criticised and his mistake corrected by C. Jordan in 1881, who thus became the discoverer of the concept of bounded variation. But he did not invent the concept, he did not 'introduce' it – he rather discovered it in Dirichlet's proof in the course of a critical re-examination.

Another weakness in Dirichlet's proof was the use of the Cauchy integral, under which discontinuous functions were not integrable at all, and hence not Fourier-expandable in general (Dirichlet had to come up with functions that had at most countably many discontinuities as a fix). This is why Riemann found the Cauchy integral inadequate and invented the Riemann-Stieltjes integral.

For Lakatos, to 'introduce' a term out of the blue into a discussion is a "magical operation which is resorted to very often in history written in deductivist style!" To treat bounded variation or the Riemann-Stieltjes integral as an "introduced" concept rather than a proof-generated one – as he has shown – is to miss out on the understanding of the concept:

> So the two mysterious definitions of bounded variation and of the Riemann-integral are *entzaubert*, deprived of their authoritarian magic; their origin can be traced to some clear-cut problem situation and to the criticism of previous attempted solutions of these problems. (1976/2015, 158)

[14] Some might worry that this might be too demanding. However, Lakatos's reply is "let us try." (1976/2015, 153) He thinks of this stringency as a positive point: while this makes academic work demanding and long, "… it has to be admitted that they would be much fewer too, as the statement of the problem-situation would too obviously display the pointlessness of quite a few of them." (153, fn. 5) Thanks to Nuhu Osman Attah for raising the point.

[15] "All functions are Fourier-expandable (1) the value of which at a point of jump is $\frac{1}{2}[f(x+0)+f(x-0)]$, (2) which have only a finite number of discontinuities, and (3) which have only a finite number of maxima and minima." (1976/2015, 157)

In sum, I understand authoritarianism – the second criterion of degeneration – as introducing new concepts into some line of inquiry without justification, while ignoring the problem-situation and the heuristics that led us to that discourse to begin with.

C. Their Normative Import

Recall that the heuristic approach places emphasis on understanding the background and historical trajectory of a concept, over and above the concept in its current form. It seems to me that *both* superfluity and authoritarianism run afoul of this emphasis on the refutations and errors in the history of a concept. The original problem and the errors that followed (i.e. the problem-situation) are just as important as the end-product and the proof-generated concept, because they tell us what has not worked (and will not worked), why the concept is the way it is now, and hopefully the available routes for development based on those errors (at least, it rules out the *un*available routes for development). By failing to grasp this problem-situation, superfluity and authoritarianism miss out on a complete understanding of the concept as *part of a long historical trajectory,* instead atomizing it as a stand-alone concept – methodologies with superfluity and authoritarian tendencies thus hinder complete understanding of the concept. As Zeta puts it in the dialogue, "A problem never comes out of the blue. It is always related to our background knowledge". (1976/2015, 74) By adopting a methodology which elides this background knowledge, a concept and the associated problem is rendered incomplete.

On one hand, authoritarianism fails to account for past problems and errors by tearing apart present discussions from past problems – the discussion is presented mystically 'as is', without context; the reader is told to accept it on faith, or that they need 'mathematical maturity' to understand why it is what it is. This obscures the errors and problems that were crucial in generating the proof-generated concept, and hence we lose sight of the direction the concept was developing in the long chain of proofs and refutations: we are "rewriting history to purge it from error" (1976/2015, 49) and hence "the zig-zag of discovery cannot be discerned in the end-product". (1976/2015, 44) The degenerate concept becomes atomized as a result, which hinders growth in Lakatos's view.

On the other hand, superfluity reflects a lack of concern for a concept's problem-situation. Terminology is produced because simply one *can,* not because it presents any particularly insightful development for the concept and its trajectory. Sometimes, as in Becker's case, one assumes themselves to be presenting fruitful development for a concept, even though their work adds no actual value. Again, this approach treats the concept at hand as one that is divorced from its problem-situation – instead of considering what problems the terminology is meant to resolve, we are instead pursuing 'cheap, shallow generalisations' because it would be *nice* even if the result is ultimately trivial. We have seen this in the case of generalizing the Euler characteristic to (6) and (7): obviously, we *can* generalize the Euler characteristic to the case of intuitive polyhedra with cavities and multiply-connected faces, but the naïve terminology – of cavities and multiply-connected faces – and the accompanying generalizations were simply no longer fruitful to the discussion of polyhedra at that point of the trajectory of the concept in the dialogue. A superfluous extension of a concept involves ever more esoteric 'generalizations' to cover cases no one cares about (in the context of the line of inquiry surrounding that concept). In Lakatos's words:

> Quite a few mathematicians cannot distinguish the trivial from the non-trivial. This is especially awkward when a lack of feeling for relevance is coupled with the illusion that one can construct a perfectly complete formula that covers all conceivable cases. Such mathematicians may work for years on the 'ultimate'

> generalisation of a formula, and end up by extending it with a few trivial corrections. (1976/2015, 103, fn. 158)

By failing to grasp what is trivial (which can be aided by hindsight or a grasp of the problem-situation), research degenerates by either treading trodden grounds (as with the case of Becker) or extending a concept to domains which are simply unfruitful (as in the case of cavities and multiply-connected faces).

In short, both superfluity and authoritarianism have clear normative import: if our goal is to pursue growth for the concept or theory by having a clearer understanding of the concept, we ought to avoid both forms of degeneration. Degeneration in these two senses thus play an *evaluative* role for the growth of a concept.

III. Degeneration in P&R vs. Degeneration in MSRP

I have focused on degeneration in the context of **P&R**, but how does this connect with Lakatos's much more famous classification of scientific research programmes as degenerative or progressive? While not much is explicitly said about degeneration in **P&R**, Lakatos *does* classify the trajectory of a scientific theory (rather, a family of theories) in a similar fashion. For Lakatos in **MSRP**, science should be understood analogous to the above approach to mathematical theories and concepts: not as isolated atoms, but as a *sequence* of theories or concepts – what he later call a *scientific research programme* – grouped together by various criteria (such as their positive and negative heuristics[16]). This "shifts the problem of how to appraise *theories* to the problem of how to appraise *series of theories.* Not an isolated *theory,* but only a series of theories can be said to be scientific or unscientific: to apply the term 'scientific' to one *single* theory is a category mistake." (1978, 34)

We can classify the trajectory of this sequence of theories over time (i.e. its 'problemshift' from one theory to another) as progressive or degenerative, according to two criteria: whether it is (**i**) *theoretically* progressive and (**ii**) *empirically* progressive (1978, 33–34). Being theoretically progressive refers to a succeeding theory containing 'excess empirical content' by predicting novel facts compared to its predecessor, and would distinguish, according to Lakatos, a 'scientific' problemshift from a non-'scientific' one. Being empirically progressive refers to the excess empirical content of this succeeding theory leading to the discovery of new facts, thereby corroborating the new theory's novel predictions. A problemshift is deemed *overall progressive* if it is *both* theoretically and empirically progressive, and *overall degenerating* if it is not.

Much ink has been spilled over Lakatos's account on this point.[17] What I am interested in is how the account of degeneration presented in **MSRP** can be augmented by the account of degeneration I have presented based on **P&R**, and how they can be collectively marshalled for the philosopher.[18] In my view, the two accounts of degeneration (and growth/progress) complement each other: while the account of degeneration in **MSRP** focused on *content,* the account of degeneration in **P&R** focused on *depth* – how deep or trivial the research is, how it connects with its predecessors, the potential or actual fruitfulness of the research, and so on (discussed above in **II**) – which in turn hinges on *methodology.*

---

[16] See Lakatos (1978, 47–52).

[17] For an excellent collection of essays on Lakatos's methodology, see *Appraising Lakatos: Mathematics, Methodology, and the Man*, eds. Kampis, Kvasz and Stöltzner (2002).

[18] Interestingly, Stöltzner (2002, 157 – 187) already does something to similar effect but goes in the opposite direction of what I am doing here. He does the reverse, by proposing that we can apply the application of the conditions of progress and degeneration described in **MSRP** *to* mathematics (and mathematical physics) profitably. What I am doing here can be seen as a continuation of that sort of project – to bring Lakatos's insights from his discussion of informal mathematics into physics.

This former account of degeneration with respect to *content* takes a central role in Lakatos's project for scientific research programmes. Furthermore, it can be straightforwardly cashed out in terms of the *number* of theoretical predictions a theory makes and actually corroborated predictions relative to its rivals and competitors. These properties make them an easy target for analysis: for starters, we can just count the number of propositions (or sentences, or whatever your favourite truth-bearers are) non-vacuously entailed by the theory (and whether they are corroborated)![19] This is not to say that considerations about content is somehow unimportant. We need yardsticks for discussing, comparing, and evaluating the content produced by scientific research programmes. If these yardsticks are clear and easy to apply, all the better. However, the focus on content – in terms of the predictions of a theory – does not really consider the *methodological* issues that might also be considered progressive or degenerating. I think this is important. In Rho's words in **P&R**: "not every increase in content is also an increase in depth." (1976/2015, 103) A theory might have superior content while simultaneously experiencing a degeneration in methodology and *depth* of research.

This is where the latter account of degeneration comes into play. The latter focuses on the *depth* of research at each problemshift – and this of course depends much more on the style and methodology of research, which may also be far more diverse than a generally accepted theory and its contents. Nevertheless, just as there are authoritative interpretations of theories even if there are generally numerous interpretations of any single theory, there are also authoritative figures, presentations, rhetoric, and methodologies, which may yet be open to analysis of depth. (An attempt to analyze one such authoritative presentation is made later in **IV**.) This, in turn, requires analysis of notions like taste, triviality, fruitfulness, awareness of the problem-situation and so on, as we have discussed in **II**, which are not obviously amenable to logical analysis like the content-oriented notions of degeneration are.

But both *are* inseparable aspects of scientific research – we need to be concerned about both the depth of research, in terms of whether authoritarianism and superfluity are occurring, and the content being produced by the research, in terms of whether there is theoretical and empirical progress.

If I am right, we can extend Lakatos's classification of scientific research programmes quite straightforwardly: a problemshift is *overall progressive* if it is overall progressive with respect to content (content-degeneration) *and* avoids degeneration with respect to depth (depth-degeneration), that is, by being theoretically and empirically progressive *while* avoiding authoritarianism and superfluity. It is *degenerating otherwise*. As with all things, depth-degeneration will come in degrees – obviously not all research at any one time will typically contain authoritarianism and superfluity, but *how much* research falls afoul of authoritarianism and superfluity will determine the degree of degeneration with respect to depth. Hence, it seems to me that my account of degeneration, based on the **P&R**, augments the account of degeneration found in **MSRP** and provides a new dimension of analysis for the application of Lakatos's overall framework.

IV. The Case of Entropy – Degeneration from Physics to Information

I now show how my account can be put to work by using it to analyze a key problemshift in the historical development of entropy, which plays a role of ever-increasing importance in our best sciences. Despite its ubiquity, the concept of entropy is not easily grasped. Some like Swinburne (1904, 3) complained that "there is no sort of physical idea underlying [entropy]"; the concept of

[19] Of course, what counts as 'non-vacuous' may yet be a matter for contention.

entropy does not come equipped with an obvious physical idea for us to latch onto, despite being defined in terms of physical quantities. This hints at degeneration – how can a concept that is so ubiquitous in physics be so imprecisely understood? In Lakatosian fashion, entropy, like other concepts, comes with its own historical trajectory and must be understood in those terms in order for us to have a firm grasp of it.

As a proof of concept, I begin this project by analyzing in detail one important shift for the concept of entropy – the incorporation of information-theoretic notions into entropy by Jaynes (1957).[20] In doing so, I will follow Lakatos's method in the Appendix of **P&R:** highlight a prominent piece of work that was influential in dictating a concept's trajectory and point out its various degenerative traits, while suggesting what could have been done otherwise.

A. Jaynes's "Information Theory and Statistical Mechanics"

Jaynes was not the first to propose a marriage between information theory and statistical mechanics – that honour goes to Leon Brillouin.[21] However, Jaynes's paper was certainly one of the (if not *the*) most influential. As Seidenfeld (1986, 468) notes, "I doubt there is a more staunch defender of the generality of entropy as a basis for quantifying (probabilistic) uncertainty than the physicist E. T. Jaynes." In a footnote to his famous 1973 paper on black hole entropy, Bekenstein observed that "the derivation of statistical mechanics from information theory was first carried out by E. T. Jaynes". (1973, fn. 17) In that same paper, he notes how, by 1973, "the connection between entropy and information is well known", (2335) and later states, in a matter-of-fact way, that entropy "is the uncertainty in one's knowledge of the internal configuration of the body." (2339) Clearly, Jaynes's information-theoretic subjectivist interpretation of statistical mechanics had won out by the 1970s – entropy has transformed from a quantity that keeps track of the reversibility or irreversibility of thermodynamic processes to a quantity that keeps track of the amount of information we have (or have lost) about said processes.

This makes Jaynes's paper the perfect candidate for evaluating the degeneration or lack thereof in the transition from entropy as a thermodynamic, physical, concept about physical systems to an information-theoretic, subjective, concept about our ignorance or partial knowledge about physical systems.

Prior to Jaynes, the Gibbsian approach to statistical mechanics was by far the dominant paradigm in physics. Under the Gibbsian approach, the Gibbs entropy S_G of a physical system is defined by:

$$S_G = -k \int_{\Gamma} \rho(x,t) \log \rho(x,t)\, dx$$

Here, Γ is the 6N-dimensional phase space of the physical system in question, x is a point in Γ, dx is the volume element of Γ, and, importantly, $\rho(x,t)$ is some probability distribution defined over Γ which may or may not change over time. The Gibbs entropy is thus a function of these probability distributions. To define $\rho(x)$, we are invited to consider a fictitious infinite ensemble of systems having (generally differing) microstates (position and momentum) consistent with the known macrostates (e.g. temperature, volume, pressure) of the actual system. In short, the macrostate(s) determines the choice of $\rho(x,t)$. S_G, in turn, is supposed to match the

[20] I focus solely on the first of two papers he published on the topic in 1957, as the second part focuses largely on applying the account developed in part I, rather than presenting any novel arguments for the account itself.

[21] See e.g. Brillouin (1956), Ch. 12 and beyond.

thermodynamic entropy at the thermodynamic limit,[22] which justifies its definition and also provides a physical basis for using it. Just as the thermodynamic entropy tracked the reversibility or irreversibility of thermodynamic processes, being equal to zero for reversible processes and greater than zero for irreversible ones (for closed systems over time), so does S_G at the appropriate limit.

Setting aside the debate over the nature of these fictitious ensembles (among other conceptual issues with the Gibbsian approach) for present purposes,[23] the important thing is that the approach so far is physical and world-oriented. In particular, the probability distributions $\rho(x)$ depend on the physical state of the system and are empirically determined – for instance, a system in equilibrium with an arbitrarily large heat bath is described with the canonical ensemble distribution, an isolated system with constant energy is described with the microcanonical ensemble distribution, and so on. These distributions, in turn, tell us the probability of some set of microstates obtaining given said constraints. We are here concerned simply with whether (and how likely) certain microstates of the actual physical system occur. Nothing in the Gibbsian theory *forces* us to employ notions of ignorance or knowledge so far, i.e. notions that would typically be described as 'subjective' do not need to be employed in Gibbsian statistical mechanics.[24]

In contrast, Jaynes (1957) explicitly introduces the notion of "subjective statistical mechanics" – where the usual rules of statistical mechanics can be "justified independently of any physical argument, and in particular independently of experimental verification". (620) To elaborate, Jaynes believed that statistical mechanics should not be interpreted as a physical theory in itself, with its equations, choice of distributions, and rules of computation justified by physical reasoning. Rather, it should be interpreted as a system of statistical inference, having to do first and foremost with our partial knowledge about physical systems. This system is then underpinned by the maximum entropy principle, which prescribes maximizing entropy as a formal means of representing maximal ignorance about that which we do not know. This principle is intended by Jaynes as an *a priori* principle of reasoning. The physics provides only the means of enumerating the possible states of the system and their properties (1957, 260). We use these as constraints on our knowledge (or lack thereof) and infer from these a set of equations via an appeal to information theory and subjectivist interpretations of probability and entropy.

Jaynes makes two related but distinct claims about his proposed account of statistical mechanics. First, Jaynes' first overarching claim is that statistical mechanics should be interpreted in a *subjectivist* fashion. This is opposed to an *objectivist* approach, which treats the probabilities produced by statistical mechanics as objective chances about events in the world (independent of what we think about those events). In his words:

> The "subjective" school of thought regards probabilities as expressions of human ignorance; the probability of an event is merely a formal expression of our expectation that the event will or did occur, based on whatever information is available. (1957, 622)

For the subjectivist interpretation of statistical mechanics, the probability distributions, such as the canonical ensemble, the grand canonical ensemble, the microcanonical ensemble etc. are used to represent our partial knowledge of the system given certain constraints. The probabilities given by

[22] i.e. when the number of particles in the system and the volume of the system itself approach infinity, with the ratio between particle number and volume held constant.

[23] See, again, Callender (1999) and Goldstein *et al* (2020) for a nice overview of the issues with the Gibbsian approach.

[24] There is a debate about whether the use of coarse-graining in Gibbsian statistical mechanics (which is necessary to recover the second law) might be anthropocentric and hence 'subjective' – see Robertson (2020) for a discussion of why that is not necessarily the case (and hence why the Gibbsian approach still need not appeal to subjectivist notions).

these ensembles are not really about the objective chances of the microstates occurring *per se*. Rather, these probabilities are interpreted to represent the degrees of belief we ought to have about these microstates, given suitable constraints.

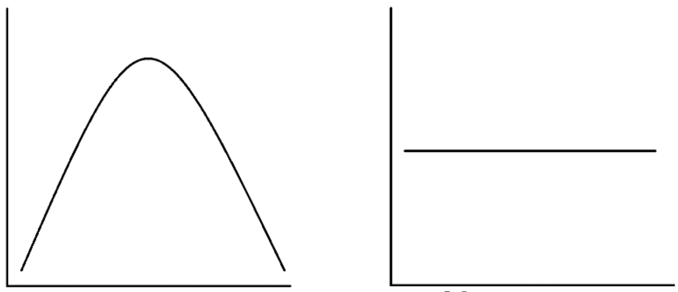
What is important, however, is that the suitable constraints are not merely physical ones given by how the system is set up or behaving. This is his second major claim.

In addition to the subjectivist interpretation of statistical mechanics, Jaynes famously proposed an additional constraint to the inferential process: the maximum entropy principle (or MAXENT). In short, the principle calls for the maximization of the entropy of the system, in addition to the other relevant physical constraints. However, the entropy of the system is interpreted in an information-theoretic way, *over and above* the subjectivist interpretation. Recall the Gibbs entropy:

$$S_G = -k \int_{\Gamma} \rho(x,t) \log \rho(x,t)\, dx$$

Under the subjectivist interpretation, $\rho(x,t)$ now represents the degrees of belief we ought to have in some (set of) microstates x obtaining at time t. Over and above that, S_G is to be interpreted (with the Boltzmann constant k set to unity via a choice of units) as the Shannon entropy instead, representing 'uncertainty' contained in $\rho(x,t)$, i.e. the 'uncertainty' contained within our degrees of belief about said system. The intuition is that a peaked distribution contains less uncertainty than a flat distribution, and it turns out that S_G for a peaked distribution is indeed lower than a flat distribution. Furthermore, just as collecting information is additive, so too is S_G additive (as a simple result of its logarithmic form).

In sum, MAXENT is treated as a rationality constraint: we assume maximal ignorance about a system except what we know about it, where 'maximal ignorance' is equated to adopting a maximum information-entropy constraint on the system. This is what Jaynes meant by his account of statistical mechanics being a general account of statistical inference. (621, 629–630) Jaynes then showed that the adoption of MAXENT can recover all the usual equations and expressions of statistical mechanics.



Lower entropy *Maximum entropy*

As with Lakatos's account in **MSRP**, much has already been said about the content of Jaynes's claims, and I will not add more to the mix.[25] I focus on the methodological *depth* of Jaynes's paper instead by applying the extended account of growth and degeneration.

To begin, let us consider what the problem-situation is, so that we may better evaluate Jaynes's paper. The founding motivations of statistical physics, found in the works of Boltzmann and Gibbs,[26] are quite clear: to understand the molecular foundations of thermodynamics, and to

[25] See e.g. Seidenfeld (1986) and the references therein.

[26] The transition from Boltzmann to Gibbs is itself an interesting chapter in the history of statistical physics, but will be left, hopefully, to future work.

interpret thermodynamics in terms of molecular mechanics. As Boltzmann states in the introduction to his *Lectures on Gas Theory*:

> I hope to prove in the following that the mechanical analogy between the facts on which the second law of thermodynamics is based, and the statistical laws of motion of gas molecules, is also more than a mere superficial resemblance. (1896/1995, 5)

Gibbs, while differing in his approach to statistical mechanics, held a similar view about the goal of statistical mechanics:

> We may […] confidently believe that nothing will more conduce to the clear apprehension of the relation of thermodynamics to rational mechanics, and to the interpretation of observed phenomena with reference to their evidence respecting the molecular constitution of bodies, than the study of the fundamental notions and principles of that department of mechanics to which thermodynamics is especially related. (1902, ix)

The search for an appropriate *interpretation* of statistical mechanics which connects thermodynamics to statistical mechanics was a prime focus of both Gibbs and Boltzmann, even though their methods differed significantly.

As is well-known, the Gibbsian approach – and hence its associated entropy – face several conceptual troubles in attaining this goal. Recall the unclear nature of the fictitious ensembles and how to interpret the probabilities provided by the approach, as well as issues such as the requirement to assume various physical hypotheses. For instance, ergodicity or metric transitivity[27] is typically introduced as a founding assumption[28] in statistical mechanics textbooks to connect measured (i.e. time) averages to the expectation values over phase space. To his credit, Jaynes does mention (in passing) some of these worries, such as the appropriate interpretation of statistical mechanical probabilities and the requirement for Gibbsian statistical mechanics to include 'physical hypotheses' such as ergodicity or an *a priori* principle of indifference (1957, 621). However, the question is then whether his paper contributes in a significant way to this problem-situation. In particular, does his proposed interpretation of entropy-as-ignorance solve the issues faced by this problem-situation? Does it provides a clearer understanding of the issues above, or does it obfuscate the issues at hand?

The stated motivations for his paper (1957, 261) suggest preliminary grounds for concern about superfluity. Jaynes claims that his primary motivations are (i) bringing in new mathematical machinery to statistical mechanics, and (ii) the notion that information theory is "felt by many people to be of great significance for statistical mechanics", although "the exact way in which it should be applied has remained obscure." But these motivations do not seem to help with respect to the problem-situation above. He also does not specify any concrete problem-situation relevant to statistical mechanics, only mentioning the above issues in passing. These concerns are validated when we analyze the paper in closer detail.

---

27 Roughly, ergodicity (or what Jaynes referred to as 'metric transitivity', and also known elsewhere as 'metric indecomposibility' (see e.g. Sklar 1993, 165)) is the idea that the phase space for a system is such that a phase point's trajectory is not confined only to one region – another way to think about it is that the phase point's trajectory traverses the entire volume of phase space.

28 See Frigg and Werndl (2020) for an overview of the status of this principle and related notions.

B. Assessing the Depth of Jaynes's Claims: Superfluity

The distinction between subjective and objective interpretations of probabilities was presumably an attempt by Jaynes to address the issue with interpreting the probabilities prescribed by Gibbsian statistical mechanics. This is, of course, a real interpretative issue with statistical mechanics. As mentioned, we can distinguish between interpreting probabilities as worldly objective chances, or subjective degrees of belief about events (which may or may not be subject to further rational constraints); the usual debate ensues as to which interpretation is appropriate.[29] However, regardless of the result of *that* debate, it seems to me that Jaynes's *actual* proposal with regards to MAXENT simply does not rely on a choice between them.

In other words, the discussion of this distinction is simply *superfluous* in the context of Jaynes's paper. For one to see the irrelevance and superfluity of that distinction, it is enough to consider that concepts like *information* and *uncertainty*, and what is sometimes confusingly called 'knowledge' or 'our knowledge' about something, in information theory, are in fact neutral between the two interpretations of probability (contrary to folk usage of these terms). The Shannon entropy is simply a formal quantity that tracks the flatness of *any* probability distribution (be it a distribution for objective chances or degrees of belief): the more peaked it is, the more information (and less entropy) it contains. Indeed, looking at its use in communications, the relevant distributions involved are typically distributions of objective *frequencies* (e.g. of letters, words and so on), not *degrees of belief*. Unless one is understandably tricked by the occurrence of subjective-sounding words like 'surprise', 'information' (in the sense that it informs *someone*), 'uncertainty' and 'knowledge', the notion of information is, I claim, neutral between the objective and subjective interpretations of probability.

And why should it? Information theory is an extremely useful and practical tool for our everyday communications, but it is ultimately a mathematical tool, without explicit metaphysical import. The distinction between objective and subjective interpretations of probability, on the contrary, is clearly a metaphysical one. as Jaynes notes: "the theories of subjective and objective probability are mathematically identical", though they differ conceptually. (1957, 622) So it is for the information-theoretic (Shannon) entropy. Even though common introductions gloss it over as a measure of 'uncertainty', it must be emphasized that this does not force a subjective interpretation of probability onto the Shannon entropy.

If so, then the MAXENT proposal which simply requires including the constraint that the Shannon entropy of any probability distribution is maximized, over and above other physical constraints, is likewise *neutral* between interpretations. We can certainly *choose* to interpret MAXENT in a subjectivist way as Jaynes did. Since we are considering only probability distributions as degrees of belief, maximizing entropy is akin to adopting the 'flattest' distribution of degrees of belief regarding a certain class of events given the available constraints. But we can also consider probability distributions as objective chances, in which case the MAXENT proposal becomes one in which we postulate that the probabilistic behavior of systems simply act in a way that maximizes entropy given the constraints.

Jaynes seems to think that the latter is unpalatable, while the former is acceptable. At several points Jaynes mentioned how his subjectivist proposal avoids 'arbitrary assumptions' (630) or 'physical hypotheses' (621). But it is confusing to see why physical hypotheses should be avoided or labelled arbitrary in the field we call *physics*. Ergodicity might have its own conceptual issues,[30] but it is surely a valid hypothesis to be considered and debated about, rather than dismissed seemingly *a*

[29] For a more detailed discussion of the general positions one might take, see Sklar (1993, ch. 3).
[30] See, again, Frigg and Werndl (2020).

priori. This is especially since ergodicity allows us to reproduce much of the physics we care about at the macroscopic level. And, at the very least, we are making a claim about the system's actual behavior and why it fits the predictions we make about it in our theory. Compare this to the subjectivist proposal, in which the theory of statistical mechanics no longer describe the dynamics of the chances of events occurring on phase space, but merely our degrees of belief about those events occurring. As Albert famously quipped,

> Can anybody seriously think that our merely being *ignorant* of the exact microconditions of thermodynamic systems plays some part in *bringing it about,* in *making it the case,* that (say) *milk dissolves in coffee?* How could that *be?* What can all those guys have been *up* to? (2000, 64)

Is the subjectivist proposal really any less arbitrary when it comes to connecting our physical theories to the world? Jaynes does not elaborate. I do not want to adjudicate such a debate here, though it suffices to say that the interpretation of probabilities is simply superfluous to the actual MAXENT proposal – the proposal itself, as a piece of mathematics, is independent of interpretation. Since both interpretations will inevitably reproduce the same mathematics (and hence the same equations), both interpretations inevitably rise and fall together.

There is little reason to think that Jaynes meant the MAXENT proposal to be much more than just a useful piece of mathematics that can help us compute and make predictions in a more tractable fashion. However, if that is the case, the question of interpreting probabilities in statistical mechanics *does not even arise,* since the actual goal of the paper isn't even about interpretation or the metaphysics of statistical mechanics. In turn, the question of interpreting the thermodynamic entropy, defined over these probabilities *about the system,* does not arise. Rather, MAXENT is a proposal concerning convenient prediction and computation. Jaynes writes:

> Although the principle of maximum-entropy inference appears capable of handling most of the prediction problems of statistical mechanics, it is to be noted that prediction is only one of the functions of statistical mechanics. Equally important is the problem of interpretation; given certain observed behavior of a system, what conclusions can we draw as to the microscopic causes of that behavior? To treat this problem and others like it, a different theory, which we may call objective statistical mechanics, is needed. (627)

This shows that the MAXENT proposal is merely a convenient proposal for arriving at the computations required for prediction problems. Hence Jaynes claimed that adopting the 'subjective point of view' and MAXENT for predictions serves a "great practical convenience". But if we were to press Jaynes on the interpretation and metaphysics of statistical mechanics, we would still need 'objective statistical mechanics':

> In the problem of interpretation, one will, of course, consider the probabilities of different states in the objective sense; i.e., the probability of state *n* is the fraction of the time that the system spends in state *n.* (627)

Jaynes's take on the interpretation of probabilities *about the actual physical system* remains an objectivist one – and one seemingly adopting some version of the ergodic hypothesis![31] This goes to show that the probabilities prescribed by statistical mechanics about actual systems – and their

[31] Roughly, an ergodic principle states that the time-averages of some parameter is equal (in some limit) to the expectation value of that parameter (and hence the probability of that parameter having some quantity is equal to that amount of time the system spent in the region of phase space with that quantity, as Jaynes says here).

interpretations – are not even in question here in Jaynes's paper, since his proposal is supposed to be one concerning 'subjective statistical mechanics', rather than 'objective statistical mechanics'. However, if we are not concerned about the issue of interpreting the probabilities given by the theory about actual physical systems, this renders discussing the interpretation of probabilities entirely superfluous! If we are only interested in prediction and computation, all we need is the *mathematical* MAXENT proposal, and a formal proof that it does in fact recover the equations we want. The ability for the MAXENT proposal to shorten and speed up the derivations of certain equations (as Jaynes shows in the paper) is, by and large, not in question here. Yet there is also no need to provide an interpretation for MAXENT and the notion of entropy involved in that case. There is no more need to interpret the mathematical shortcuts that one takes, any more than one needs to justify and interpret the algorithms behind WolframAlpha when one takes a shortcut with their integrals. All this renders Jaynes's insistence on packaging the MAXENT proposal with a choice of interpretation for both probabilities and entropy confusing.

Furthermore, since we are not even tackling the *actual* issue, of how to interpret the probabilities assigned to the states of the actual system, the subjectivist MAXENT package is not even *relevant* to the original problem-situation. In other words, the insistence on providing an information-theoretic interpretation of entropy – replacing the previous thermodynamic and physical interpretation via the second law of thermodynamics and notions of reversibility/irreversibility of actual processes – is simply unjustified because the MAXENT proposal has no real need for interpretation.

Additionally, no other genuine argument for the information-theoretic interpretation can be found in his paper. He starts off with a *proviso*:

> The mere fact that the same mathematical expression $-\sum p_i \log p_i$ occurs both in statistical mechanics and in information theory does not in itself establish any connection between these fields. This can be done only by finding new viewpoints from which thermodynamic entropy and information-theory entropy appear as the same *concept*. In this paper we suggest a reinterpretation of statistical mechanics which accomplishes this, so that information theory can be applied to the problem of justification of statistical mechanics. (621)

As I have shown, information theory is not applied to the *justification of statistical mechanics,* since that would require interpreting statistical mechanics and the metaphysics within it (e.g. about whether swarms of particles can actually recover the macroscopic description). Despite Jaynes claiming here that he is proposing a reinterpretation of statistical mechanics, he does not succeed in doing so – that is all left in the 'objective statistical mechanics' side of things, which he has chosen to downplay. Rather, Jaynes's proposal is the adoption of new mathematical tools for computing predictions in statistical mechanics, which does not force any interpretation at all. In any case, such interpretations have no real import for the actual issue of the problem-situation, that of interpreting the probabilities attached to events themselves. Jaynes has not yet shown or even begun to show that thermodynamic entropy and information-theoretic entropy are the same concepts! Later he writes:

> Since $[-\sum p_i \log p_i]$ is just the expression for entropy as found in statistical mechanics, it will be called the entropy of the probability distribution p_i; henceforth we will consider the terms "entropy" and "uncertainty" as synonymous. (622)

He has not shown that the thermodynamic entropy, i.e. "entropy", and the information-theoretic entropy, i.e. "uncertainty", are the same as a matter of interpretation, because the paper is not at all concerned with interpretation and 'objective statistical mechanics', only prediction and 'subjective statistical mechanics'.

In sum, Jaynes's discussion of interpretative issues is *superfluous*. Jaynes has added unnecessary terminology from information theory and confused these new concepts with old questions without actually addressing any of the old questions from the original problem-situation. The insistence on interpreting entropy as information-theoretic ignorance in a subjectivist sense, defined over distributions interpreted as degrees of belief, is likewise superfluous.

C. Assessing the Depth of Jaynes's Claims: Authoritarianism

Jaynes also displays authoritarianism when insisting on treating statistical mechanics as a general means of prediction, apparently viewed through subjectivist lens.

Recall that authoritarians introduce new concepts into a line of inquiry without justification, while ignoring the problem-situation and heuristics which led us to those concepts. This is an important issue in Jaynes's paper, since the problem-situation at hand is barely specified. No details about the issues facing 'objective statistical mechanics' or the Gibbsian approach are presented. Instead, Jaynes simply presents the information-theoretic interpretation, the subjectivist interpretation, and the maximum entropy principle, as though they must be taken altogether.

Jaynes proclaims that, in

> freeing [statistical mechanics] from its apparent dependence on physical hypotheses of the above type, we make it possible to see statistical mechanics in a much more general light. (621)

Throughout the paper, Jaynes insists that the subjectivist approach is necessary for approaching the prediction issue. However, two questions arise: first, why the downplaying of 'physical hypotheses' used by 'objective statistical mechanics' and why do we need to 'free' statistical mechanics from them? Second, why the focus on prediction and information theory, and the downplaying of the importance of interpretation? Both questions are unfortunately unanswered.

To the first question, it is unexplained why Jaynes constantly denounces the use of physical hypotheses, in deriving certain predictions about statistical mechanical systems. It is clear that he believes that these hypotheses are undesirable – he spends some time claiming that metric transitivity is not needed if we adopt the MAXENT principle. (624) However, he never says *why* we should not adopt *any* physical hypotheses about the systems we are studying.

He focuses instead on how MAXENT can help us do away with these hypotheses. But it is not clear that it does – it merely shifts our attention away from whether those hypotheses hold. Jaynes notes that adopting MAXENT is *in fact* akin to adopting ergodicity – except about our own degrees of belief about the system's behavior, rather than about the actual system's behavior:

> Even if we had a clear proof that a system is not metrically transitive, we would still have no rational basis for excluding any region of phase space that is allowed by the information available to us. In its effect on our ultimate predictions, this fact [i.e. MAXENT] is equivalent to an ergodic hypothesis, quite independently of whether physical systems are in fact ergodic. (624)

Is the system *really* ergodic after all? And do we really need ergodicity to derive the equations concerning those systems' behavior? Jaynes's proposal has two options: one is to say nothing at all – clearly an unsatisfactory answer. Another option is to reply: the MAXENT proposal says that you should have degrees of belief matching the situation where the system is ergodic (as the quote above suggests). But that just means I ought to believe that the system is ergodic after all. And that means believing the *physical hypothesis of ergodicity*. Yet that was the original issue to begin with, in our problem-situation: we want to know whether ergodicity is necessary for the statistical mechanical system to behave in accordance with our observations. Either MAXENT is irrelevant to our problem-situation, or it simply adds nothing new. Old questions remain.

The original problem-situation has been neglected. Yet, we are made to believe that these questions are to be ignored in favour of the new proposal – MAXENT, information theory, subjectivism – without justification for why that should be so. This is a case of authoritarianism.

Turning to the second question now: as I have discussed above, the founding fathers of statistical mechanics were concerned first and foremost with the interpretative issues – how do we connect the particles or systems of statistical mechanics to the bulk macroscopic behavior we find in thermodynamics? Of course, that is not to say that prediction has no role to play in statistical mechanics. However, it is strange to ignore a core tenet of statistical mechanics, which seems like what Jaynes has done here. Reading the paper, one gets the impression that prediction holds supreme place in statistical mechanics. Interpretation seems to be an after-thought. But prediction goes hand in hand with interpretation – to predict the behavior of the system we must understand what the system *is*, and that is a matter of interpretation. As I have argued in **IV.B**, Jaynes's paper is completely divorced from interpretative issues. In this respect his paper is authoritarian: it ignores the problem-situation of statistical mechanics, such as the importance of interpretative issues.

The introduction of information theory, and the shift in focus on statistical mechanics as a general tool of statistical inference, is likewise authoritarian. Jaynes offers no reason for adopting information theory – we are told that the Gibbs entropy *can* be interpreted as the Shannon entropy, and that "the development of information theory has been felt by many people to be of great significance for statistical mechanics". (621) Likewise, we are not told why statistical mechanics should be a general tool of statistical inference, freed from physics, where "the usual rules are thus justified independently of any physical argument, and in particular independently of experimental verification." (620) These are all core tenets of the MAXENT proposal, but they remain unjustified.

In conclusion, Jaynes's paper falls afoul of both superfluity and authoritarianism. With respect to methodological depth, then, it was a degenerative piece of work. Since the key transition of entropy from a concept concerned with thermodynamics and actual physical systems to a concept concerned with ignorance and our knowledge of said systems occurred here, this shift is a degenerative one as well.

Yet, Jaynes' paper changed the trajectory of the entropy concept. For instance, by appearing as though the paper presented an interpretative package, despite the actual proposal not needing one, the paper introduced confusion to the actual interpretative issues. The interpretative package of information-theoretic entropy and subjectivism became adopted as an answer for the interpretative issues associated with 'objective statistical mechanics' instead. Recall the Bekenstein quote: a mere twenty years later the information-theoretic interpretation has escaped from 'subjective statistical mechanics' into 'objective statistical mechanics', with *thermodynamic* entropy (defined over probability distributions of the microstates of the *actual* systems) being interpreted as ignorance or uncertainty. Three years later, Hawking would simply assert: "an intimate connection between holes (black or white) and thermodynamics […] arises because information is lost down the hole." (1976, 13) Degeneration has occurred.

D. Content-Oriented Degeneration

Recall that my depth-oriented tools for analysis were meant as an extension of Lakatos's original tools found in **MSRP**. So it is worthwhile to briefly consider the theoretically and empirically progressive/degenerative nature of the MAXENT proposal, which seems to me theoretically and empirically degenerative as well.

Recall the definitions: being theoretically progressive refers to a succeeding theory predicting more novel facts compared to its predecessor. Being empirically progressive refers to the excess empirical content of this succeeding theory actually leading to the discovery of new facts, thereby corroborating the new theory's novel predictions.

As Jaynes notes, however, nothing new is added in terms of theoretical progress, because 'subjective statistical mechanics' will recover *exactly the same predictions* as 'objective statistical mechanics':

> Conventional arguments, which exploit all that is known about the laws of physics, in particular the constants of the motion, lead to exactly the same predictions that one obtains directly from maximizing the entropy. (624)

And that:

> the subjective theory leads to exactly the same predictions that one has attempted to justify in the objective sense. (625)

This shows that no new predictions are provided by this proposal. The 'new' proposals attached in Jaynes's papers are typically just new ways of doing the same calculations. For instance: Jaynes's treatment of Siegert's 'pressure ensemble' is also merely a reworking of Siegert's own derivations, published just a year prior (Lewis and Siegert 1956). In short, the MAXENT proposal is theoretically degenerative. Furthermore, since there are no new predictions, there are no new predictions to corroborate. The proposal is empirically degenerative. This, of course, also adds to that sense of superfluity one gets when analyzing Jaynes's proposal in detail.

Overall, then, Jaynes's paper is degenerative *tout court.* Given that it had such a huge influence on the current understanding of entropy as ignorance and uncertainty, especially in the field of contemporary black hole thermodynamics (Bekenstein and Hawking are typically known as the founding fathers of black hole thermodynamics), this current understanding must be re-assessed.[32]

V. Conclusion

I have provided and motivated an extension to Lakatos's analysis of growth and degeneration found in **MSRP**, by appealing to discussions found in **P&R** in terms of superfluity and authoritarianism. I have also shown that this extension makes possible a new dimension of analysis of a piece of mathematical or scientific work, independent of the analysis in terms of theoretical and empirical progress or degeneration found in **MSRP.**

[32] Some have already begun this work – see e.g. Prunkl and Timpson (preprint).

This extension can be fruitful for future philosophical work for two reasons.

Firstly, developing more tools to understand how scientific and mathematical concepts degenerate or grow has natural affinities with a growing understanding of philosophy as conceptual engineering.[33] As Chalmers (2020, 4) writes, conceptual engineering is "the project of designing, evaluating, and implementing concepts", where we consider not only what a concept *is,* but also what it *should be*. Developing new tools for identifying points of degeneration in a concept's historical trajectory helps us evaluate a concept and consider alternative ways of designing and developing said concept.

Secondly, there are benefits to understanding degeneration and evaluating concepts for the philosopher of science and physics. It seems to me that Lakatos's warning against the deductivist style in **P&R** can be instructive for physics where a mere skim of any textbook or lecture notes will reveal a familiar deductivist style. Despite the differences between physics and mathematics, Lakatos emphasized that "mathematical heuristic is very like scientific heuristic – not because both are inductive, but because both are characterised by conjectures, proofs, and refutations. The important difference lies in the nature of the respective conjectures, proofs (or, in science, explanations), and counterexamples". (1976/2015, 78) Lakatos, too, would think of the heuristic style as applying to both physics (and the sciences) and mathematics.

As proof of concept, I have evaluated Jaynes's proposal, a key transition point in the historical trajectory of the concept of entropy. I hope to have shown that my account does provide a novel means of assessing the degeneration or progress of this transition, by critically analyzing the aspects of his paper which exhibited superfluity and authoritarianism.

Regrettably, the history of entropy (and thermodynamics) is a rich one and cannot be adequately grappled with in its entirety within the confines of this paper. This paper leaves behind a variety of fruitful directions, ripe for the picking by the hopeful conceptual engineer. For the philosopher of physics: if we want to re-engineer and design a newer, better, conceptually clearer notion of entropy, we would do well to engage with – and dispel – the misunderstandings engendered by this paper, and other similarly degenerative transition points. Likewise for other concepts in physics, mathematics, or even philosophy itself. There remains much to be done.

References

Albert, D. Z. (2000). *Time and chance*. Harvard University Press.

Bekenstein, J. D. (1973). Black Holes and Entropy. *Physical Review D*, 7(8), 2333–2346. doi:10.1103/physrevd.7.2333

Boltzmann, L. E. (1995). *Lectures on gas theory*. New York: Dover.

Brillouin, L. (1956). *Science and information theory.* New York: Academic Press.

Bub, J. (2005). Quantum mechanics is about quantum information. *Foundations of Physics*, vol. 35, no. 4, pp. 541 - 560.

Callender, C. (1999). Reducing Thermodynamics to Statistical Mechanics: The Case of Entropy. *The Journal of Philosophy, 96*(7), 348. doi:10.2307/2564602

Chalmers, D. (2020). What is conceptual engineering and what should it be? *Inquiry*. doi: 10.1080/0020174X.2020.1817141

Chua, E. (forthcoming). Does Von Neumann Entropy Correspond to Thermodynamic Entropy? *Philosophy of Science*. doi:10.1086/710072

Corfield, D. (1997). Assaying Lakatos's philosophy of mathematics. *Studies in History and Philosophy of Science*, 28:99–121.

[33] See e.g. Haslanger (2000), Chalmers (2020).

Corfield D. (2002). "Argumentation and the Mathematical Process." in Kampis G., Kvasz L., Stöltzner M. (eds) *Appraising Lakatos: Mathematics, methodology, and the man*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-0769-5_8

Dougherty, J. and Callender, C. (preprint) Black Hole Thermodynamics: More Than an Analogy? URL: http://philsci-archive.pitt.edu/id/eprint/13195 (accessed 2021-01-03).

Frigg, R., & Werndl, C. (2020). Can Somebody Please Say What Gibbsian Statistical Mechanics Says? *The British Journal for the Philosophy of Science*. doi:10.1093/bjps/axy057

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159*(3), 417-458. doi:10.1007/s11229-007-9237-y

Gibbs, J. W. (1902). *Elementary principles of statistical mechanics: Developed with special reference to the rational foundation of thermodynamics*. Yale University Press.

Goldstein, S., Lebowitz, J., Tumulka, R., and Zanghi, N. (2020). "Gibbs and Boltzmann Entropy in Classical and Quantum Mechanics", in Allori, V. (ed) *Statistical Mechanics and Scientific Explanation*, pp. 519–581, Singapore: World Scientific.

Haslanger, S. (2000). Gender and Race: (What) Are They? (What) Do We Want Them To Be? *Noûs, 34*(1), 31-55. doi:10.1111/0029-4624.00201

Hawking, S. W. (1976). Black holes and thermodynamics. *Physical Review D, 13*(2), 191-197. doi:10.1103/physrevd.13.191

Jaynes, E. T. (1957). Information Theory and Statistical Mechanics. *Physical Review, 106*(4), 620-630. doi:10.1103/physrev.106.620

Kampis, G., Kvasz, L., & Stöltzner, M. (2002). *Appraising Lakatos: Mathematics, methodology, and the man*. Dordrecht: Kluwer.

Kiss, O. (2006). Heuristic, Methodology or Logic of Discovery? Lakatos on Patterns of Thinking. *Perspectives on Science, 14*(3), 302-317. doi:10.1162/posc.2006.14.3.302

Lakatos, I. (1978). *The methodology of scientific research programmes*, edited by J. Worrall & G. Currie. New York: Cambridge University Press.

Lakatos, I. (1976/2015). *Proofs and refutations: The logic of mathematical discovery* edited by J. Worrall & E. Zahar. Cambridge: Cambridge University Press.

Leng, M. (2002) Phenomenology and mathematical practice. *Philosophia Mathematica*, 10:3–25.

Lewis, M. B., & Siegert, A. J. F. (1956). *Extension of the Condensation Theory of Yang and Lee to the Pressure Ensemble. Physical Review, 101(4), 1227–1233.* doi:10.1103/physrev.101.1227

McIrvine, E. C., & Tribus, M. (1971, September). Energy and Information. *Scientific American*. Natsuume, M. (2015). AdS/CFT Duality User Guide in Lecture Notes in Physics 903. Tokyo: Springer, Tokyo. doi:10.1007/978-4-431-55441-7

Robertson, K. (2020). Asymmetry, Abstraction, and Autonomy: Justifying Coarse-Graining in Statistical Mechanics. *The British Journal for the Philosophy of Science, 71*(2), 547-579. doi:10.1093/bjps/axy020

Seidenfeld, T. (1986). Entropy and Uncertainty. *Philosophy of Science,* Vol. 53, No. 4 (Dec., 1986), pp. 467-491.

Sklar, L. (1993). *Physics and chance: Philosophical issues in the foundations of statistical mechanics*. Cambridge: Cambridge University Press.

Stöltzner M. (2002). "What Lakatos Could Teach the Mathematical Physicist." in Kampis G., Kvasz L., Stöltzner M. (eds) *Appraising Lakatos: Mathematics, methodology, and the man*. Dordrecht: Kluwer. https://doi.org/10.1007/978-94-017-0769-5_10

Swinburne, J. (1904). *Entropy, or, Thermodynamics from an engineer's standpoint: And the reversibility of thermodynamics*. Westminster: Constable.

Uffink, J. (2001). Bluff Your Way in the Second Law of Thermodynamics. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics, 32*(3), 305-394. doi:10.1016/s1355-2198(01)00016-8

Wallace, David (Preprint). The Necessity of Gibbsian Statistical Mechanics. URL: http://philsci-archive.pitt.edu/id/eprint/15290 (accessed 2021-01-03).

Werndl, Charlotte (2009) Justifying definitions in mathematics: going beyond Lakatos.

Philosophia Mathematica, 17 (3). pp. 313-340.