Inconsistency and Scientific Realism

Juha Saatsi
Department of Philosophy
University of Leeds

April 18, 2011

Abstract

I erect a framework within the semantic view of theories for explaining the empirical success of internally inconsistent models and theories, with scientific realism in mind. The framework is an instance of the 'content-driven' approach to inconsistency, advocated by both Norton (1987) and Smith (1988), whose ideas my analysis aims to clarify and substantiate.

1 Introduction

Science is not as clean and untainted affair as one might think, logically speaking. In reality science harbours at least four kinds of inconsistencies. i) The historical record shows that older theories are often inconsistent (and sometimes radically so) with the later ones. ii) Our best and most fundamental theories of physics—quantum theory and general theory of relativity—are mutually inconsistent. iii) A typical field of physics, such as fluid dynamics, contains a number of mutually incompatible models, due to incompatible idealization schemes. iv) Some theories and models are *internally inconsistent* by virtue of being based on an inconsistent set of assumptions. All of these different kinds of inconsistencies have been used as ammunition in arguments against scientific realism.[1] It is not difficult to start thinking of an intuitive realist response to these arguments. Firstly, by resorting to a notion of 'approximate' truth (or 'partial' truth, or some such qualification) we can hope to locate the assumptions that give rise to an inconsistency in the 'false bits' of our theories and models. Secondly, in this way we can hope to hold onto the realist idea that their empirical success is explicable in terms of their 'truth content'.

Easy though it is to start thinking of this intuitive response, the challenge remains, of course: the realist must spell out in a principled way what 'approximate truth' (or 'partial truth', or whatever cognate notion she prefers) amounts to. One may further hope that the realist can give an account of this notion that provides

[1] See, for example, Laudan (1981); Barrett (2003); Cartwright (1983) and Morrison (2000); Frisch (2004).

a *unified response* to different kinds of inconsistencies. I believe that a unified account of this kind can be given.[2] This paper sketches such an account that focuses on *internal inconsistency*, but the perspective and the conceptual framework adopted are motivated also by diachronic inconsistencies in the history of science, as well as by idealisations.[3]

To anticipate what's to come, the basic idea is the following. I will argue that at least some internal inconsistencies—I will drop the word 'internal' from now on—can be analysed in terms of an inconsistent model being an *'inferentially veridical representation'*, a key notion to be clarified in §3. This notion can be used by the realist to capture a sense in which a scientific model may latch onto reality only partially, but critically in ways that are responsible for the model's empirical success. This notion serves to spell out the intuition that the empirical success of an inconsistent model may be explicable in terms of the model being closely related to a fully consistent model with comparable empirical success. Demonstrating that there is an appropriate relation between the two models—one inconsistent and one consistent—is business that requires close attention to the inconsistent model's *content*. For this reason my approach is an elaboration of the so-called *'content-driven'* perspective on inconsistency, to be surveyed in a context-setting way in §2. Some illustrations are provided in §4, after the framework gets first set up in the abstract in §3. Finally, section §5 reflects on the heuristic value of inconsistencies.

Although this paper is explicitly motivated by and written within my broader realist outlook, it is not in and of itself meant to provide a positive argument for realism. Here I only advance the claim that at least some internally inconsistent models can be viewed as inferentially veridical representations, and this is but a step in a full realist response to inconsistency. An anti-realist could also adopt the conceptual framework erected here to analyse and explicate the empirical success of an inconsistent model. Nevertheless, the framework has a natural place in a broader realist gambit, which would in addition incorporate two (not necessarily independent) further steps: (i) an argument to the effect that the predictive success of (some) inconsistent models can actually be *explained*, in a realist sense, in terms of those models' inferential veridicality, and (ii) an argument to the effect that (i) *justifies* realism regarding these kinds of inconsistent models.

2 Content-driven perspective to inconsistency

Inconsistency in science gives rise to many philosophical questions, and one can approach these questions from very different perspectives. A useful distinction has been made between LOGIC-DRIVEN and CONTENT-DRIVEN perspectives to inconsistency.[4] Take, for example, the following question:

[2] It would be foolishly optimistic, however, to think that it successfully deals with *each and every* instance of inconsistency in science.

[3] Saatsi (2005, 2008) uses some of the same conceptual resources to respond to 'pessimistic induction', and Saatsi (2011) applies the framework to idealization.

[4] See Smith (1988a,b), Norton (1987), Norton (2002).

Q_s**:** How should one explain the empirical success of inconsistent theories / models?

If we look at things from the point of view of classical logic, we know that from an inconsistent set of assumptions everything logically follows, including the right predictions, of course. But this logical point is hopelessly unilluminating, for we want to explain how scientists managed to derive exactly the predictions that amount to the empirical success, and not others. The *logic-driven* approach suggests that in response to the 'logical explosion' that results from inconsistency in classical logic, we should turn to non-classical, paraconsistent logics in order to capture (or to represent, or model, perhaps) the constraints that there must have been on scientists' inferences from an inconsistent set of assumptions. This arguably avoids the 'logical anarchy', and it may provide (part of) an answer to Q_s.

Alternatively, and in stark contrast to this formal approach, the *content-driven* perspective recommends that we explain the empirical success of an inconsistent theory (or model) in terms of an informal assessment of the particular inferences licensed by the theory (or model), *based on what it (explicitly) says about the world*.[5] From this perspective the key to answering Q_s above has to do with appreciating the content of the theoretical assumptions, and (arguably) with due attention to this content we can explain all the relevant inferences—the ones actually made, as well as the ones omitted—without leaving classical logic behind.

Thus characterised, the content-driven approach remains somewhat nebulous and open-ended. Exactly how can the specific content of some theoretical assumptions be used to answer Q_s? We can substantiate the content-driven approach by proposing, for example, that the 'benignancy' of inconsistency in an empirically successful theory can be understood in terms of the relation of that theory to some equally successful consistent theory. If a consistent alternative exists that makes more or less the same assumptions about the world and allows the relevant empirical results to be derived, then surely this is pertinent for understanding the empirical success of the inconsistent theory?[6] As Norton intimates:

[5]We need to look at the *explicit* content of the theoretical assumptions, of course, since these assumptions imply everything by virtue of being inconsistent.

[6]Making 'more or less the same assumptions' about the world is a contextual notion: the content of the respective assumptions must be compared in the appropriate theoretical context, taking into account the empirical success in question, for example. Here's an artificial toy example. From assumptions regarding the speed of light and the length of a stick one can calculate the time (T) it takes for light to travel back and forth the stick's length. Let's assume that both the speed of light ($c = 1$) and the stick's length ($l = 1$) are constant. It follows that $T = 2$. It would be inconsistent to assume that (i) the stick's length is constant; (ii) the stick's length is 0.99999 units when light travels one way; and (iii) 1.00001 units when light travels the other way. From these three assumptions one can calculate the time for light to travel back and forth. Although everything follows from an inconsistent set of assumptions, clearly most natural calculations give as the answer either 1.99998, or 2, or 2.00002 units, depending on which assumptions are actually employed in the calculation. All of these are very close to the actual value of T. This can be explained by pointing to the fact that each assumption in the inconsistent set {(i), (ii), (iii)} is (in a natural, intuitive sense) approximately true, given the aim of the calculation.

> If we have an empirically successful theory that turns out to be logically inconsistent, then it is not an unreasonable assumption that the theory is a close approximation of a logically consistent theory that would enjoy similar empirical success. The best way to deal with the inconsistency would be to recover this corrected, consistent theory and dispense with the inconsistent theory. (Norton, 2002: 193)

A natural realist spin on this is to further assert that the consistent theory that is 'approximated' by the inconsistent theory is also a *true* theory (or at least a theory that is closer to being true). In that way we can begin to give a *realist* response to the question Q_s above. But it is incumbent on the realist to say something a bit more systematic and principled regarding the relationship between the two theories. This is because, for one thing, we need to know more about one's realist commitments with respect to an empirically successful inconsistent theory for which we don't (yet) have a consistent alternative. If the realist can't say *anything* about how the two theories are related, apart from saying that they are related in *some way* that is explanatory of the empirical success of the inconsistent theory, then one's realist commitments seem rather deflated. Secondly, mere allusion to 'close approximation' just invites the usual criticisms of a wholly unexplicated notion of approximate truth.

The existing literature on the content-driven approach to inconsistency contains very little discussion of this issue. How should an advocate of the content-driven approach advance here? Partly this depends on how one chooses to represent theories, models, and their relationship to the world in the first place. Secondly, one may wish to develop a realist response to inconsistent models as part of a bigger realist picture that ties internal inconsistencies with other types of inconsistencies that the realist must also accommodate. In the next section I will proceed thus by taking a clue from a particular realist response to diachronic inconsistency ('pessimistic induction'), and adopting a version of the semantic view of theories to inform my conception of 'approximate' theory-world relation. But before we get to this I will critically review some other content-driven responses to inconsistency in order to further motivate and set the context for my positive view.

Let's begin with Smith (1988a,b), who frames his account of inconsistency in terms of *statements* and an epistemic attitude that he dubs *'entertaining'*.

> [P]araconsistency of deductive inference is not a necessary condition for obtaining information from an inconsistent set of statements about a given subject. There are many modes of inference other than deduction, just as there are many epistemic attitudes that we can have toward statements besides assent.
>
> ...
>
> To have a term with a weaker designatum than 'assent' to refer to the various epistemic attitudes we adopt toward statements employed for any of their pragmatic virtues, I will say that we 'entertain' them.

> ...
>
> All members of a set of mutually inconsistent statements cannot be rationally assented to, but each member of such a set can acquire confirming evidence. When we have evidence for the truth of each of two incompatible claims, it is quite rational to entertain both. However, the fact that they are inconsistent means that we must mentally flag them to guard against indiscriminate future use. One or the other is false. At best, both can be 'approximately true,' or 'partially true,' or 'true under some disambiguation,' etc. (Smith, 1988b: 243–244)

Implicit in Smith's view is the idea that theories (inconsistent, or otherwise) are naturally construed as *sets of statements*. The empirical success of an inconsistent theory is then explicable, if at all, by demonstrating how the relevant predictions are also deducible from a consistent alternative theory, which can be formed by replacing at least one of the original statements S by some true statement S' that S 'approximates.'

Arguably the semantic view of theories presents a better alternative to the 'received' statement view that Smith essentially adheres to. Smith's position can also be criticised for not really saying what 'entertaining a statement' amounts to, and for leaving the notion of 'approximate truth' (or 'partial truth') of a statement wholly unexplicated. Although Smith doesn't embark on explicating the sense of approximation or partiality at play, he does have a fruitful intuition in this regard. Namely, he thinks that a statement S can approximate another statement S', in a relevant way, by virtue of S' being a *'weakening'* of S in the sense that S says something *unnecessarily specific* about the system, something that S' leaves out.[7] For example, according to Smith the statement "light is a longitudinal wave" has an 'element of truth' in it since it correctly attributes to light the property of being a wave, even if it states a falsehood about the *specific kind* of wave it is (viz. *longitudinal* wave). I'll discuss this example from Smith more fully in §4, as an illustration of my position, which taps directly into this intuition and presents it not in terms of statements, but in terms properties that our models attribute to their target systems.

John Norton is another advocate of the content-driven approach. Norton (1987) looks at the inconsistent old quantum theory of blackbody radiation comprising assumptions that draw on 1. Thermodynamics; 2. Statistical mechanics; 3. Classical electrodynamics; 4. Quantum postulate. This theory had considerable empirical success, e.g. the derivation of the Planck distribution law for blackbody radiation. Norton suggest that one can extract a consistent *subtheory* which is in itself enough for the derivation of the Planck distribution law. According to Norton this explains the empirical success of the inconsistent theory.

[7] Unfortunately Smith doesn't say much about this kind of weakening in general, apart from the suggestive idea that such weakening is appropriate if "the evidence bears more directly on a *consequence* of the original hypothesis (and certain auxiliary claims)" (Smith 1988b: 247, my italics).

> In this paper I will show the manifest inconsistency produced by conjoining 4. to 1., 2. and 3. was inessential to the old quantum theory's recovery of the Planck distribution law and the results leading up to it. To do this I will extract a subtheory from 3. which no longer posits continuity of the relevant energies and show that the Planck distribution law can still be recovered from the conjunction of it with 1., 2. and 4. The resulting subtheory of the old quantum theory of black-body radiation will be free of manifest inconsistency and I conjecture its consistency. (1987: 328)

But what exactly is a theory, and what is a subtheory, for Norton? Although he's not very explicit about it, Norton (like Smith) seems sympathetic with the 'received' view of theories (which he calls the 'classical' view), with the important addition that there are also *restrictions* on the results that can be carried over from the classical to the quantum domain.[8] Norton doesn't say in general how such meta-level extra assumptions could be represented as being part of our theories or models. Although we have an indication of there being factors at play in the case of old quantum theory that explain the success of that theory, we are left wanting for a more perspicuous and general account of this content-driven approach to inconsistency. Can the realist do better?[9]

In the next section I will explore a way of looking at models as 'inferentially veridical' representations that not only yields a more general philosophical account of inconsistency but also captures the gist of Norton's analysis of the old quantum theory case. I take the gist of Norton's analysis to be the idea that in the case of old quantum theory scientists "extravagantly overcommitted themselves" (p. 331) with respect to classical electrodynamics by attributing properties to quantum systems on the basis of that theory. Norton's explanation of the empirical success of the old quantum theory then hinges on there being a "minimal characterisation" (op.cit.) of electromagnetic radiation that (a) can be extracted from classical electrodynamics; (b) is consistent with the other theoretical assumptions used in the derivation, including the quantum postulate; and (c) is enough for the actual derivation of the Planck distribution law. In other words, Norton effectively argues that only some of the properties attributed to radiation according to the old quantum theory are actually doing any work in the derivation. At the same time, none of the conjuncts 1.–4. are dispensable in its entirety, since each conjunct contains something of relevance for the derivation: each conjunct contributes to the derivation by entailing that light instantiates some particular property that is part of the minimal subtheory characterisation of light that is required the derivation. In the next section I

[8] These restrictions may not be articulated and made explicit.

[9] In a different context—Newtonian cosmology—Norton (2002) frames the issue in slightly more general terms, but now the content-driven approach seems to have a slightly different twist: the new idea is that "meta-level arguments" (concerning symmetry, say) can pick out some features that an unknown consistent replacement theory would have, and those features are then used to isolate some consequences of the inconsistent theory as worthy of scientific interest. I personally can't see how to appropriate this recipe in the case of the old quantum theory.

will give an account of 'overcommitment' and 'minimal characterisation' in more general terms that provide a way of analysing the underlying 'truth-content' that is explanatory of the empirical success of an inconsistent theory or model. I'll revisit Norton's case-study in §4 to illustrate.

3 Inconsistent models as inferentially veridical representations

The empirical success of an inconsistent piece of theorising can be understood in terms of that theorising latching onto reality in some relevant respects. This is what (a realist reading of) the content-driven perspective to inconsistency suggests in the first approximation. How this suggestion gets sharpened depends on how one understands theories, models, and their relation to the world.

I adopt the semantic view of theories according to which theories are families of models, and models relate to the world by a representation relation that can be, and often is, non-linguistic.[10] Although models typically do not relate to the world in the way propositions do, they still "say" various things about their targets by virtue of being *interpreted* in terms of some real world systems. Models essentially function by allowing modellers to attribute properties to their target systems. A *model description* specifies all the properties of a *model system*, and also the representation relationship between the model system and the target. Although language is typically involved in specifying a model system and also the relation between a model system and its target, the latter relation itself—the relation that is critical for analysing how models can partially latch onto reality—is non-linguistic.

For Giere and Teller the relationship between a model and the world is a symmetric one of 'similarity'. I do not subscribe to this specific assumption, nor do I wish to make any universal assumptions about the nature of the model system. For me it only matters that a model system is used to attribute properties to its target in such a way that we can start thinking about it partially latching onto reality by virtue of correctly representing the target with respect to some properties, but not others. This focus on properties is profitable (as I have argued in more detail elsewhere; cf. Saatsi (2011)) because in analysing the way in which models approximate, idealize, or abstract, we can mobilize conceptual resources associated with properties and especially the way in which properties are related in various ways. I will next use these conceptual resources to provide a framework, in general terms, that also allows us to analyse some instances of inconsistent theorising.

Recall Norton's suggestion that the empirical success of the old quantum theory of blackbody radiation could be explained by extracting a consistent *subtheory* which allows the derivation of the Planck distribution law. In the framework of the semantic view it is natural to talk instead of a consistent *sub-model*, since 'model' is the basic unit of analysis. We can approach the notion of sub-model by looking

[10]This version of the semantic view is most closely aligned with the view developed by Giere (1988). See also Teller (2001).

for a natural way of weakening the representational content of the original (inconsistent) model, so as to render it consistent. We can think of the representational content of the original model in terms of the properties that it explicitly attributes to its target. The following way of weakening this representational content then suggests itself. Take some property P_i that the model explicitly attributes to the system. Consider different *less specific* properties $p_{i_1}, p_{i_2}, ..., p_{i_k}, ...$ such that their instantiation would be guaranteed by the instantiation of P_i (together with the laws of nature). You can think of these 'less specific' properties simply as different *ways of exemplifying* P_i.[11] Now, if the instantiation of one of these less specific properties is consistent with all the other properties that the inconsistent model explicitly attributes to the target, then you've found *a* consistent weakening of the original model. If it is furthermore the case that this consistent weakening can yield the same empirical success as the original model—by following the same logico-mathematical derivations—then you have found a consistent model that is related to the inconsistent one in a way that is explanatory of the latter's empirical success. This is what it means for there to be a (fruitful) 'minimal characterisation' of a system that is consistent with the rest of our beliefs about it.

The consistent sub-model is explanatory of the empirical success of the inconsistent model by virtue of demonstrating that the inconsistent model in a sense 'contains' a consistent representation of all the relevant features of the system. This sub-model suffices for the derivation of the relevant results and predictive successes. Scientists working with the inconsistent model have 'overcommitted themselves' by using the model to attribute *unnecessarily specific* properties to the target system.[12] The inconsistency of the model resides in the over-committed, unnecessarily specific properties, and not in the properties that are relevant for and explanatory of the model's empirical success. Explaining the empirical success of a model in these terms requires an account of the particular properties that the two models trade on, an account that brings out the relationship between the relevant less specific properties, on the one hand, and the corresponding unnecessarily specific properties, on the other. That is, it requires attention to the *content*, in the spirit of the content-driven account.

It is important to emphasize that the model need not *explicitly* represent the target as instantiating the success-fuelling properties. It is enough that the explicit property attributions in the model *imply*—when taken separately, not jointly[13]—that the target system has those less-specific features that 'fuel' the actual derivation. In that case we can say that the model is an *inferentially veridical* representa-

[11] Examples of more specific and less specific properties are afforded by determinate–determinable pairs of properties: being red is one way of being coloured; weighing 10kg is one way of weighing between 9 and 11kg; being an equilateral triangle is one way of being a triangle; being neuro-toxic is one way of being poisonous; etc. See, Funkhouser (2006), for example, on the determination relation and how it differs from other kinds of the specification relations, such as multiple realization. My use of 'more specific' and 'less specific' is meant to cover all of these determination relations.

[12] Or, at least they have given a description of the model system that is unnecessarily specific, in case they are not committed to taking those aspects of the model as representing the target.

[13] Inconsistent property attributions imply everything if taken jointly, of course.

tion. Inferentially veridicality of a model is *contextual*: it is in part determined by the purpose to which the model is put. This contextuality is unsurprising and unproblematic: it follows from the fact that the explanandum at hand—the empirical success of a model—is itself contextual.

For the realist, furthermore, the consistent sub-model can provide a potential *realist* explanation of the empirical success of the inconsistent model by virtue of demonstrating that the inconsistent model in a sense gets all the relevant features of the system exactly right. The realist can thus employ the notion of inferential veridicality (as a part of the realist programme) to capture a sense in which a model that is less-than-fully-veridical can nevertheless 'latch onto reality', or be 'partially true', so as to explain its empirical success.[14] Since (non-linguistic) models in the semantic view are not candidates for truth or falsity—being non-linguistic—it is better to capture that sense of 'partially latching onto reality' not in terms of truth or falsehood, but directly in terms of how successful the model is in attributing properties to the target (vis-à-vis the purpose to which the model is put).

The realist can put the notion of inferential veridicality to work in connection with various kinds of inconsistencies: it is also applicable to some famous cases of radical theory-shifts in the history of science, and it can deal with at least some idealized models.[15] Hence it provides a unified framework for the realist, turning on the following two basic ideas: (1) a model that falsely represents a system instantiating some property P can latch onto reality by virtue of the fact that the attribution of P implies the attribution of some less specific property p; and (2) that it can be explanatory of the model's empirical success that the worldly system in question actually instantiates the property p. This was the key idea above also concerning (internally) inconsistent models: the inconsistency arises from some false property-attributions, but the inconsistency is "benign" since only some corresponding less specific properties really matter for the specific empirical success to be explained. In this sense the inconsistency is fully contained in the unnecessarily specific properties attributed to the system, with respect to which the model is over-committed.

The framework sketched above goes some way towards answering question Q_s in general terms, assuming of course that it is applicable to some actual inconsistent models in science. I will claim in the section below that it is indeed thus applicable. There are further aspects to Q_s that remain to be discussed as well. So far we have focused on the benignancy of inconsistency, but we can also ask about its positive epistemic worth: what good is it? The short answer is that overcommitment, inconsistent or otherwise, can be of heuristic, pragmatic value. I will give a slightly longer answer along these lines in §5 below.

[14]The realist requires a further positive argument to take any given consistent sub-model to give an actual realist explanation of the empirical success of some related inconsistent model.

[15]Cf. Saatsi (2005) for a treatment of the Fresnel-Maxwell case that can be parlayed into snugly fitting the present framework. Cf. Saatsi (2011) for a related discussion of idealization.

4 Illustrations

In order to illustrate the conceptual framework outlined above, I will now revisit two examples from the history of science that have been taken to support the content-driven perspective to inconsistency.

Early wave theory of light. Smith (1988b) suggests that the early 19th century wave theorists engaged in inconsistent reasoning when they tried to accommodate polarisation phenomena by appealing to an asymmetry of light with respect to its direction of propagation, counter to the basic assumption that light is a longitudinal wave. (Light was assumed to be a longitudinal wave since it was difficult to conceptualise an elastic ether that could carry exclusively transverse waves.) Arguably there was evidence for both hypotheses regarding the nature of light, and empirical success flowed from 'entertaining' each assumption in the inconsistent set.

> The statement that 'light is a longitudinal wave,' and the claim that 'light exhibits asymmetry with respect to its direction of propagation' are two [...] incompatible but individually confirmable hypotheses. The ultimate judgement of 19th century physics was that there is an element of truth in both. (1988b: 244))

We can explain this success, Smith argues, by pointing out that there is a natural way of 'weakening' the hypothesis that light is a *longitudinal* wave so that consistency is restored, without sacrificing the essence of the successful theoretical derivations that rely on that hypothesis. For what is really needed for those derivations is just the weaker assumption that light has a wave nature (of some kind) that obeys the Huygens–Fresnel principle, for example, which can be equally valid for both transverse and longitudinal waves.

> Once it was recognized that the 'longitudinal' aspect of the wave theorists' hypothesis was excess baggage as far as the confirming evidence was concerned, that hypothesis could be weakened so as to restore consistency to the entire proposal—including the claim that light exhibits asymmetries with respect to the direction of propagation. (Ibid.)

I share the intuition here, but I don't think Smith makes much headway in spelling out the intuition in terms of 'sets of inconsistent statements', where some statements contain 'excess baggage' and an 'element of truth' in them. (For example, how should we understand in Smith's terms the 'longitudinal aspect' of the wave hypothesis, and how should we pry that apart from the 'element of truth' of that hypothesis?) Following the non-linguistic, modelling-view of science, I maintain that a more perspicuous way of conceptualising the situation is to say that the early wave theorists latched onto reality by virtue of attributing such properties to light that:

(a) were *unnecessarily specific*, in the sense that there are less specific properties the instantiation of which is guaranteed by the instantiation of these more specific ones. And:

(b) the *relevant less specific properties* are sufficient for the actual logico-mathematical derivations that yield the empirical successes.

The early wave theorists were exactly right in representing light as *a* form of wave. Given that *being a wave*—viz. propagating in space with some magnitude oscillating sinusoidally—is a property that can be realized in many ways,[16] one can latch onto the wave nature of light and yet get the precise realization of the wave wrong. But it need not matter, as long as the theoretical inferences made from the wave nature of light didn't really rely on the more specific assumption regarding the precise way that light waves. It is in this sense that the early wave theorists' inconsistent reasoning can be regarded as *inferentially veridical*.[17]

Old Quantum Theory of Blackbody Radiation. Blackbody radiation is electromagnetic radiation in thermal equilibrium with a perfectly non-reflecting 'blackbody' that remains at some constant temperature. Blackbody radiation has a characteristic spectrum that only depends on the temperature T. This is given by Planck's law:

$$u(f,T) = \frac{8\pi h f^3}{c^3}\frac{1}{e^{\frac{hf}{kT}} - 1} \tag{1}$$

which gives the spectral energy density for temperature T as a function of frequency f (c, h, k are constants). John Norton (1987) considers actual theoretical derivations of Planck's law from a set of assumptions that is prima facie inconsistent. The inconsistency arises from the fact that in deriving (1) theorists on the one hand appealed to a quantum postulate, according to which the energy levels of the 'resonators', or 'radiation oscillators' that give rise to the blackbody radiation, are quantized, but on the other hand they appealed to results drawn from classical electrodynamics, according to which these energy levels can vary continuously.

In the face of this manifest inconsistency Norton suggests that there were implicit restrictions in place on the results drawn from the classical domain that could be employed in connection with the quantum postulate: only those results were employed that were, in a sense, *independent* from the classical assumption regarding continuous energy levels. That is, according to Norton we can extract

[16]This is of course manifested in the fact that mathematically both transverse and longitudinal waves can be described by the same mathematical equations.

[17]This example provides a useful illustration of the distinction between more specific and less specific properties. I don't take it to be a good example of an internally inconsistent model, however, given that the mutually inconsistent hypotheses did not feature in any particular derivation of, or argument for, a particular theoretical result. The example that follows is more pertinent in this regard, but it gives a less intuitive handle on the kinds of properties at play.

from classical electrodynamics a ‘subtheory’—providing a ‘minimal characterisation’ of electromagnetic radiation consistent with the quantum postulate—that is enough to get the results that were used in the derivation of Planck’s law. (Namely, the Stefan-Boltzmann law and the Wien displacement law.)

Norton’s analysis fits my framework very well. Basically, he shows that only some of the *properties* attributed to radiation by classical electrodynamics are actually fuelling the derivation of (1). From classical electrodynamics it follows that these critical properties are instantiated by electromagnetic radiation and blackbody resonators, but these critical properties are actually much less specific than those properties of classical electromagnetic radiation (and the corresponding classical blackbody resonators) that lead to the inconsistency. Here are examples of the requisite kinds of less specific properties:[18]

Rest mass property Radiation has zero rest mass.

Frequency property There is a family of types of radiation, parameterized by the positive real-valued index “frequency” f.

Spatial superposition property The mixing of radiation of different frequencies and of different directions occurs reversibly, without requiring or releasing energy.

According to Norton such properties characterise ‘radiation [that] comprises a more general class of zero rest mass matter than the electromagnetic radiation of classical theory.’ (1987: 331). This class of matter is characterised by many properties we associate with classical electromagnetic radiation, but ‘very few of the characteristically wavelike properties associated with frequency are posited.’ (1987: 332) These more abstract, *less specific properties*—like the ‘spatial superposition property’—are realised in classical electromagnetic radiation, but also in radiation that is not continuous but exhibits quantum discontinuity. In other words, at a certain level of description classical and quantum systems *share interesting properties*. The empirical success of the old quantum theory of blackbody radiation is explained in terms of it latching onto these properties. Certain crucial theoretical results (the Wien displacement law, the Planck resonator formula) that involve Planck’s constant h, for example, turn out to be independent of the value of h. Thus, having derived these result for the classical domain ($h = 0$), the results carry unproblematically over to the quantum domain ($h = 6.63 \times 10^{-34}$ J sec). We can employ the classical theory to attribute properties to quantum systems, as long as we stick to the results that are applicable to both kinds of systems.

From the perspective of my framework, we can construe Norton’s notion of consistent ‘subtheory’ as a consistent representation of the system in terms of the less specific properties that were correctly attributed to the system in the first place. The relevant theorists ‘extravagantly overcommitted themselves’—as Norton puts

[18] The three properties given here are enough to recover the Stefan-Boltzmann law, for example.

it—by virtue of making these property attributions *via* unnecessarily specific classical assumptions. Such overcommitment led to the manifest inconsistency, but we can explain its empirical success by pointing out that the model of blackbody radiation is inferentially veridical, as it latches onto reality at the level of the critical less specific properties (corresponding to Norton's 'minimal characterisation' of radiation).

This perspective complements Norton's exposition by (a) clarifying and providing a more general account of the notion of consistent 'subtheory' that is contained in an inconsistent theory; (b) showing how a realist can talk about an 'inconsistent theory latching onto reality' in a way that is explanatory of its empirical success; (c) clarifying why attributing inconsistent properties to a system in the first place can be heuristically invaluable and natural. I'll expand on (c) below.

5 What good is inconsistency?

Consider our initial question again:

> **Q_s:** How should one explain the empirical success of inconsistent theories / models?

So far, we have mainly focused on a way in which an inconsistent representation can latch onto reality. It's been part of the story that the inconsistency-generating aspects of a representation can be 'harmless' by virtue of being unnecessarily specific (in the contextual sense that they weren't actually used to gain some particular empirical success that Q_s refers to). But perhaps there is also a positive story to be told about the inconsistency-generating aspects of a representation; perhaps the particular ways in which scientists 'overcommit' themselves can play a pragmatic, heuristic role in theorising?

Given the distinction that I have drawn between dispensable vs. indispensable property attributions (vis-à-vis some particular empirical success), it is not difficult to sketch possible ways in which a dispensable, unnecessarily specific property attribution can nevertheless play a vital heuristic role.

To give an intuitive toy example, imagine a situation where a scientist is theorising about a system that actually has an unobservable wave nature of a longitudinal sort. If the system gives rise to wave-like phenomena, then a sensible starting point is to model the system as a wave of some sort. If the scientist is only acquainted with transverse waves, however, she might not be initially able to conceive of a model of the system in any other way. Or perhaps she holds some misleading background beliefs—of a broadly metaphysical sort, say—that suggest that the system simply cannot realize longitudinal wave motion. Again, she might be compelled to model the system in terms of transverse waves. Assume that the scientist from such a starting point manages to produce, for some theoretical result R, a logico-mathematical derivation that doesn't "care" whether the wave is actually longitudinal or transverse. Then any empirical success thereby generated has been heuristically driven, successfully, by approaching unfamiliar and abstract via an assumption that is more familiar and more concrete (albeit ultimately false and unnecessarily specific). It may furthermore happen that en route to the empirical success the result R gets combined with an assumption that is logically incompatible with the heuristically useful but false belief regarding the transverse wave nature of the system.

Smith (1988) discusses a toy example of a somewhat similar kind. Here a car mechanic tries to make an inductive inference about the cause of a worrying periodic clicking sound coming from an engine. One piece of background information is that similar sound is caused by faulty journal bearings hitting on the crankshaft. Another piece of information is that the oil pressure is normal. The latter piece of information "contradicts" the first, given the auxiliary assumption that if the journal bearing really was faulty, the oil pressure should drop below normal.[19]

Smith suggests that the mechanic can have a theory of the cause of the sound, which includes both of these pieces of information combined with the auxiliary assumption. Taken at face value, this theory is inconsistent, if we furthermore generalise by straight induction from the earlier cases of such clicking sounds being due to faulty journal bearings.

Smith tries to spell out how this inconsistent set of assumptions could be a heuristically useful starting point in search of a consistent alternative. His answer is that the original assumption captured by statement (A) 'similar sounds are caused by faulty journal bearing hitting on the crankshaft' has a *natural weakening* (given certain obvious auxiliary assumptions): (A*) 'similar sounds are caused by [some kind of] periodic impact on the crankshaft.' A* is a weakening of A because A entails A* (given certain obvious auxiliary assumptions), but not the other way around. Faulty journal bearings are only one possible source of a periodic impact on the crankshaft. By essentially moving to existential generalisation over all the possible sources of periodic impact we weaken the assumption. Given that some possible sources of periodic impact do not contradict the other piece of information regarding the oil pressure (e.g. loosened harmonic balances connected to crankshaft are not related to oil pressure), we are rationally lead to search through such alternative sources, forming consistent models that we can test.

I take myself to be following Smith's intuition here, but I wish to steer clear from his adherence to the statement view of theories. It is better to focus directly on the properties that we assume the crankshaft and various other parts of the engine to instantiate when they function in concert. By assumption A we correctly manage to latch onto the properties *periodic motion of the crankshaft*, and *something hitting the crankshaft* as causes of the clicking sound. We overcommit ourselves by focusing on the one possible realisation of these properties by the faulty journal bearing. Indeed, we might recognise that our evidence only justifies A*, if we simply haven't experienced other faulty parts periodically hitting the crankshaft (caus-

[19]There's no logical inconsistency yet, of course, given that the first piece of information strictly speaking concerns only the past.

ing a sound that could be compared to the case at hand). The challenge for Smith is to give a general account of what he means by 'weakening of a statement.'[20] I have effectively answered this question by focusing on the relevant properties (in a specific scientific context) and how they are linked together as more and less specific properties.

The relevant properties involved in actual scientific cases are typically bound to be much more intricate and less intuitive than those involved in the above toy examples. But the same conceptual resources are still applicable to real life cases. Norton's reconstruction of the case of old quantum theory of blackbody radiation indicates this kind of heuristic role for the classical theory of electrodynamics. Somehow scientists needed to grasp at some unfamiliar and unintuitive quantum properties, and they did that by abstracting from a relevant class of classical properties they were more familiar with. Once the scientists realized that radiation in a blackbody cavity seemed to be have some basic properties quite different from what one would expect on classical grounds, they successfully tried to see if some more abstract concepts associated with classical radiation were still applicable to it. For example, they considered a generalised form of radiation which shares with the classical domain some properties that are invariant under variation in the "parameter" h, say.[21]

6 Conclusion

A scientific realist—of the kind that wishes to defend the idea that empirical success is linked to theoretical success—must give an account of scientific theorising 'latching onto reality' in ways that are explanatory of its empirical success. Such an account must be compatible with the various kinds of inconsistencies that science exhibits. Here I've shown how at least some internal inconsistencies can be analysed and understood through a framework that gives a unified account of representational models latching onto reality in an appropriate realist sense.

References

Barrett, J. A. (2003). Are Out Best Physical Theories (Probably and/or Approximately) True? *Philosophy of Science 70*(5), 1206–1218.

Cartwright, N. (1983). *How the laws of physics lie*. Oxford: Oxford University Press.

Frisch, M. (2004). Inconsistency in Classical Electrodynamics. *Philosophy of Science 71*, 525–549.

[20]What Smith says is suggestive, but begs for a general account: "the evidence bears more directly on a *consequence* of the original hypothesis (and certain auxiliary claims)" (1988: 247))

[21]Cf. also Smith (1988a) on the old quantum theory case.)

Funkhouser, E. (2006). The Determinable-Determinate Relation. *Noûs 40*, 548–569.

Giere, R. (1988). *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press.

Laudan, L. (1981). A Confutation Of Convergent Realism. *Philosophy of Science 48*, 19–49.

Morrison, M. (2000). *Unifying Scientific Theories*. Cambridge: Cambridge University Press.

Norton, J. (1987). The Logical Inconsistency of the Old Quantum Theory of Black Body Radiation. *Philosophy of Science 54*, 327–350.

Norton, J. D. (2002). A Paradox in Newtonian Gravitation Theory II. In J. Meheus (Ed.), *Inconsistency in Science*, pp. 185–195. Kluwer Academic Publishers.

Saatsi, J. (2008). Eclectic realism—the proof of the pudding: a reply to Busch. *Studies in History and Philosophy of Science 39*, 273–276.

Saatsi, J. (2011). Idealized models as inferentially veridical representations: a conceptual framework. In P. Humphreys and C. Imbert (Eds.), *Models, Representations and Simulations*. Routledge.

Saatsi, J. T. (2005). Reconsidering the Fresnel-Maxwell case study. *Studies in History and Philosophy of Science 36*, 509–538.

Smith, J. M. (1988a). Inconsistency and Scientific Reasoning. *Studies in History and Philosophy of Science 19*, 429–445.

Smith, J. M. (1988b). Scientific reasoning or damage control: Alternative proposals for reasoning with inconsistent representations of the world. In A. Fine and J. Leplin (Eds.), *PSA1988*, Volume 1, pp. 241–248. East Lansing: Philosophy of Science Association.

Teller, P. (2001). Twilight of the Perfect Model Model. *Erkenntnis 55*(3), 393–415.